Routledge
Taylor & Francis Group

Scientific Studies of Reading

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hssr20

A Meta-Analysis on the Optimal Cumulative Dosage of Early Phonemic Awareness Instruction

Florina Erbeli, Marianne Rice, Ying Xu, Megan E. Bishop & J. Marc Goodrich

To cite this article: Florina Erbeli, Marianne Rice, Ying Xu, Megan E. Bishop & J. Marc Goodrich (25 Jan 2024): A Meta-Analysis on the Optimal Cumulative Dosage of Early Phonemic Awareness Instruction, Scientific Studies of Reading, DOI: 10.1080/10888438.2024.2309386

To link to this article: https://doi.org/10.1080/10888438.2024.2309386

SCIENTIFIC STUDIES OF READING
https://doi.org/10.1080/10888438.2024.2309386

A Meta-Analysis on the Optimal Cumulative Dosage of Early Phonemic Awareness Instruction

Florina Erbeli [a], Marianne Rice [a], Ying Xu[a], Megan E. Bishop [a], and J. Marc Goodrich [b]

[a]Department of Educational Psychology, Texas A&M University, College Station, USA; [b]Department of Teaching, Learning & Culture, Texas A&M University, College Station, USA

ABSTRACT

Purpose: Research on how much early phonemic awareness (PA) instruction is optimal has produced inconclusive answers. We conducted a nonlinear meta-analysis to estimate the optimal cumulative dosage of early PA instruction on PA outcomes with an associated maximum effect size in preschool through first-grade students.

Method: Sixteen experimental and quasi-experimental primary studies (35 effect sizes) on PA instruction effectiveness that reported cumulative dosage data were included. There were 613 students in treatment and 542 students in control conditions (Mage = 5.20 years; SDage = 0.87).

Results: The cumulative dosage response model took a concave parabolic form (an upside-down U shape). Specifically, PA instruction effects improved with increasing dosage up to 10.20 hours of instruction (dmax = 0.74), after which the effects declined. Moderator analyses revealed these results held for students at-risk for reading disabilities and basic PA skills instruction. Furthermore, moderator analyses showed that the dosage response curves exhibited a convex parabolic form (a U shape) in PA instruction with letters, with effects continually increasing after 16 hours of PA instruction.

Discussion: Overall, our findings highlight the importance of planning the optimal cumulative dosage of early PA instruction in preschool through first-grade settings so that students acquire the PA and phonological-orthographic associations taught and show progress in learning to read.

Phonemic awareness (PA) is the ability to recognize and manipulate sounds or phonemes in spoken words. It is a foundational skill needed for word reading, and it supports later reading and spelling development (Erbeli et al., 2018, 2023; Foorman, 2003; Hulme et al., 2012; Melby-Lervåg et al., 2012; Muter et al., 2004; National Reading Panel, 2000). The basics of phoneme perception are acquired as part of language development prior to formal schooling, but explicit PA instruction helps students refine PA skills and understand how sounds are represented in writing (Rayner et al., 2001; Rice et al., 2022, 2023). Typically, formal instruction in PA begins in kindergarten in the United States. Most students will acquire adequate PA skills by the end of kindergarten or the beginning of first grade. To optimally plan PA instruction, it is crucial to identify salient instruction characteristics and understand how to implement these characteristics to meet the needs of students most effectively. One such characteristic is the cumulative dosage of PA instruction, defined as the total amount of hours PA instruction is administered. Given that different amounts of PA instruction dosage might lead to different magnitude of effects, we ask how much time should optimally be spent on PA instruction to be the most effective for PA outcomes. The goal of this study was to use nonlinear meta-analysis to

CONTACT Florina Erbeli erbeli@tamu.edu Department of Educational Psychology, Texas A&M University, 4225 TAMU, Texas A&M University, College Station, TX 77843, USA
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10888438.2024.2309386

estimate the optimal cumulative dosage of early PA instruction and associated effect sizes on PA outcomes in preschool through first grade.

In this study, we define PA as the ability to focus on and manipulate phonemes in spoken words (National Reading Panel, 2000). Typically, PA is broken down into specific skills/tasks, which are used to improve students' PA through instruction (National Reading Panel, 2000). Basic PA skills include identification (recognizing words that start with identical phonemes), isolation (producing a single phoneme from a word), categorization (grouping words that do or do not share phonemes), blending (combining phonemes to pronounce a word), and segmenting (separating a word into its phonemes). More complex PA skills, sometimes referred to as advanced PA skills, encompass deletion (removing a phoneme from a word) and substitution (changing a phoneme to produce a new word).

We report primary studies that estimated the effects of PA instruction on PA outcomes. We focused on instruction at the phoneme level, regardless of whether phonological awareness skills (syllables and onset-rime) were taught in addition to PA skills or not (see Inclusionary and Exclusionary criteria). Note that we acknowledge phonological awareness's role as a predictor of early word reading skills in many alphabetic languages, including English, and recognize that English children develop phonological recoding at small (phonemes and corresponding graphemes) as well as large grain sizes (syllables, rimes; Ziegler & Goswami, 2005). Even though English's small reading units might be fairly inconsistent, its larger units are less inconsistent (Treiman et al., 1995), pushing English readers to use both (e.g. Brown & Deavers, 1999). The reason why we focused on PA instruction rather than only phonological awareness instruction was three-fold. First, PA is a unitary construct (Schatschneider et al. 1999). Schatschneider and colleagues (1999) used both terms, phonemic awareness and phonological awareness, interchangeably to refer to PA tasks, although PA has been shown to be dissociable from the phonological awareness construct in some studies (e.g. Lee-Webb et al., 2004). Second, PA tasks are the strongest correlate of variability in word reading, even after controlling for variations in phonological awareness tasks, such as rime awareness (Melby-Lervåg et al., 2012). Third, and relatedly, PA, along with letter naming fluency, has a causal influence on the development of children's early literacy skills (e.g. Bus & van Ijzendoorn, 1999; L. Ehri, 2005; Foorman et al., 1998; Hulme et al., 2005, 2012; Wagner et al., 1997). Altogether, research findings support the importance of studying instruction that includes PA skills.

Regarding the definition of the optimal cumulative dosage of PA instruction, we use S. F. Warren et al. (2007) framework. We define it as the product of dose (the number of minutes per instruction session), dose frequency (the number of instruction sessions per week), and duration (the total number of weeks instruction is implemented). Typically, the cumulative dosage is measured in minutes or hours. In this study, we measure the cumulative dosage of PA instruction in hours.

Theoretical Perspectives on Studying the Optimal Cumulative Dosage of PA Instruction

Emphasis on studying the optimal cumulative dosage of PA instruction can be, in part, predicated upon studies in the field of pharmacology. Specifically, the medical community has widely recognized the importance of treatment parameters like form (liquid/pill), dose frequency (daily/hourly), and cumulative dosage (1 mg every 12 hours for 7 days for a total of 14 mg). Each parameter contributes unique information within decision guides that physicians may use to maximize treatment response. In much the same way, we can think of PA instruction as a "treatment" or an approach to helping children acquire PA skills, such that variations in instruction parameters, like cumulative dosage, can differentially influence gains in PA outcomes.

There are several theoretical perspectives on why understanding the optimal cumulative dosage of PA instruction represents an important undertaking. The first perspective refers to the role of PA skills in reading instruction. Based on reading theories (Harm & Seidenberg, 1999; C. Perfetti, 2007; C. A. Perfetti, 1992), PA is not acquired for its own sake but instead helps children understand and use the alphabetic system to read and spell. When children begin with reading instruction, their phonemic and phonological representations do not need to be fully specified to support the use of spoken language. Children typically only have partial knowledge of these structures before reading. With increased exposure to orthography and explicit instruction in sublexical orthographic-phonological representations, children's phonemic and phonological representations are further refined. In other words, an optimal dosage in instruction will facilitate the dynamic, reciprocal relations between the child's phonemic representations, orthography-phonology mapping, and reading (Foorman & Francis, 1994; Foorman et al., 1991; Rayner et al., 2001; Wagner et al., 1994) in the most effective manner possible as well as enable children to progress in their reading through self-teaching (Share, 1995). Even though the rate at which children learn to amalgamate/bond (L. C. Ehri et al., 2001; C. A. Perfetti, 1992) orthographic and phonological representations is item- (i.e. child's acquisition of individual word representations; see also Harm & Seidenberg, 1999) and child-specific (i.e. how fully specified and redundant these representations are within each child; C. A. Perfetti, 1992), it is important to understand the dosage of PA instruction needed to most effectively assist the process of children's development of specified representations. This meta-analysis attempts to answer the dosage question.

How does this theoretical background lend itself to experimental design studies? For PA skills, the notion of an "optimal cumulative dosage of instruction" means that the effect sizes in a treatment group receiving PA instruction will increase only up to the point of relative PA mastery, after which any further increase in the cumulative dosage of PA instruction will probably yield diminishing returns relative to a control group. In such instances, the control group will likely still acquire PA skills but at a slower rate. The control group will eventually catch up to the treatment group because the control group will continue improving their skills while the treatment group has already mastered them. Hence, after further observation of both treatment and control groups, the between-group effect sizes will reduce and probably disappear when studying PA. This mechanism suggests that at some point during the instruction, the differences between the treatment and control groups will be the largest. In our study, that point is referred to as the optimum dosage.

A complementary reason why studying optimal dosage is important refers to student characteristics that may influence a differential response to the dosage of PA instruction (S. F. Warren et al., 2007). Deficits in PA among students at risk for reading disabilities lead to limited growth in building phonemic-orthographic associations. Therefore, explicit, intensive PA instruction is needed for these students to succeed in learning to read. In addition, previous syntheses and intervention research have suggested that lower pretest scores on measures of phonological processing, rapid naming ability, verbal ability, and attention are associated with diminished response to reading instruction and interventions (Al Otaiba & Fuchs, 2002; Fletcher et al., 2011; Nelson et al., 2003; Schatschneider et al., 2004). Hence, changes in the cumulative dosage of PA instruction may systematically produce different effects for different participants, including students at risk for reading and other disabilities.

Finally, the spacing effect theory (S. F. Warren et al., 2007) serves as another perspective that highlights the importance of studying the optimal cumulative dosage of PA instruction. This theory posits that scheduling higher treatment intensity in a shorter time (i.e. massed, concentrated practice) would lead to differential impacts than treatment intensity delivered over a longer time (i.e. spaced, distributed practice; S. F. Warren et al., 2007). To illustrate, the prediction from the spacing effect theory would be that a student who received 20 hours of PA instruction distributed across 20 1-hour sessions per week would not learn PA skills as efficiently as if she were exposed to 20 hours of PA instruction distributed across 60 20-minute sessions held three times per week. This would be the case even though the number of treatment minutes per week would be the same, and the cumulative dosage would be the same. This theory highlights the significance of measuring dose, dose frequency, duration, and cumulative dosage separately to assess the effect of each parameter on the treatment's efficacy and, in turn, optimize it accordingly. Current research on this theory is inconclusive for PA and other reading outcomes, with results suggesting no differences between concentrated and distributed practice (Erbeli & Rice, 2022; Ukrainetz et al., 2009).

Altogether, theoretical perspectives underlying the optimal cumulative dosage of PA instruction have at least one thing in common – the discourse that is directly relevant to practice (Fuchs & Fuchs, 2015). Without knowledge about an optimal cumulative dosage, it might mean that practitioners are administering PA instruction below or above the optimal level, leading to suboptimal allocation of time and financial resources for promoting reading achievement.

Empirical Support from Randomized Control Trials on the Optimal Cumulative Dosage of PA Instruction

In reading research, several designs can be used to answer the question of the optimal cumulative dosage of early PA instruction. Experimental designs such as randomized control trials (RCTs) are the gold standard. Students can be randomly assigned to two or more levels on a key variable, cumulative dosage. Although we could not locate RCTs that had carried out such experiments for PA outcomes, some studies have explored the cumulative dosage question for other reading outcomes.

One of the first studies of this kind was by Wanzek and Vaughn (2008). The researchers manipulated dosage (minutes per intervention session), which, in turn, impacted the cumulative dosage of the intervention. The authors reported on two studies examining the effects of a multi-component (i.e. phonics, fluency, and comprehension) reading intervention with first-grade students who were at-risk for reading difficulties. Study 1 randomly assigned students to 50 sessions at 30 minutes daily (25 total hours of instruction) or a control group. Study 2 randomly assigned students to 50 sessions at 60 minutes daily (50 total hours) or a control group. The intervention procedures were identical in both studies except for dosage. Outcomes from both studies found that neither treatment group significantly outperformed the control group on phonics, word recognition, reading fluency, and passage comprehension. In addition, both treatment groups (25 hours vs. 50 hours) had similar pretest to posttest reading effect sizes, alluding to no differences in outcomes as a function of intervention cumulative dosage. Hatcher and colleagues (Hatcher et al., 2006) reached a similar conclusion in their study. They found that Year 1 British students with reading difficulties who received supplemental intervention in PA, phonics, and word and text reading for two consecutive 10-week periods (33 hours of instruction) performed comparably on measures of PA, letter knowledge, and single-word reading to a group who received the same intervention only during the second 10-week period (16.5 hours of instruction). Such a finding was further demonstrated in another study. Denton and colleagues (Denton et al., 2011) compared the effects on reading outcomes of delivering a multi-component reading intervention (i.e. PA, phonics, fluency, vocabulary, and comprehension) to first-grade students at risk for reading difficulties. The authors assigned students to three treatment conditions: extended practice (4 sessions per week, 16 weeks; 29.5 hours of instruction), concentrated practice (4 sessions per week, 8 weeks; 14 hours of instruction), or distributed practice (2 sessions per week, 16 weeks; 15 hours of instruction). Denton et al. (2011) found no statistically significant differences in decoding, fluency, and reading comprehension, suggesting that longer, more spread-out interventions were not necessarily associated with higher gains than briefer, more concentrated interventions.

These findings, however, were in contrast with those of Al Otaiba and colleagues (Al Otaiba et al., 2005). Al Otaiba et al. (2005) found that at-risk kindergarten students who received a year-long multi-component reading intervention (i.e. PA, phonics, fluency, vocabulary, and comprehension) four times per week had more robust basic reading outcomes, word identification, and passage comprehension than those who received the same intervention only two times per week. Results from Al Otaiba et al. (2005) study suggest that higher cumulative dosage of instruction might lead to better outcomes. In sum, most existing experimental research reveals no differences in reading outcomes based on the cumulative dosage, even though at least one study (Al Otaiba et al., 2005) reported an opposite finding. While these studies represent critical steps aiming to pinpoint the ideal cumulative dosage of reading instruction, none specifically evaluated PA outcomes.

Empirical Support from Meta-Analyses on the Optimal Cumulative Dosage of PA Instruction

One method that can, in part (but not ideally since correlational methods are used) circumvent the issue of a dearth of experimental research and help address the question of the optimal cumulative dosage of PA instruction is a meta-analysis. In meta-analytic studies examining PA instruction effectiveness, the cumulative dosage parameter typically serves as a moderator in a meta-regression model. Hence, if the cumulative dosage were a significant moderator in these studies, we could conclude that the effects of PA instruction differ as a function of cumulative dosage of PA instruction.

Perhaps one of the best-known meta-analyses exploring the effectiveness of PA instruction on PA outcomes was that conducted by the National Reading Panel (National Reading Panel, 2000). The NRP meta-analyzed 52 studies on PA instruction effectiveness for preschoolers to sixth graders and reported that PA instruction of moderate cumulative dosage was more effective than the instruction of shorter or longer cumulative dosage. Specifically, the National Reading Panel (2000) reported that PA instruction of a total of 5–9.3 hours ($d = 1.37$) or 10–18 hours ($d = 1.14$) was more effective compared to 1–4.5 hours ($d = 0.61$) or 20–75 hours ($d = 0.65$). Some caution is warranted about these results, though. The populations of students (at-risk versus no risk for learning disabilities) were not clearly identified in the NRP primary studies. Further, given that the cumulative dosage is a continuously distributed parameter, caution about it being estimated as a categorical parameter, like in the NRP report (2000), was raised by Wanzek and colleagues (Wanzek et al., 2016). They suggested examining cumulative dosage as a continuous variable as a more robust analysis approach. More recently, Rice and colleagues (Rice et al., 2022) attempted to replicate National Reading Panel (2000) findings in their examination of PA instructional effectiveness on PA outcomes for the treatment parameter of duration. Using duration as a continuous moderator, there were no significant moderating effects for preschool through first-grade students on PA outcomes (Rice et al., 2022).

Overall, it appears that meta-analyses on PA instruction effectiveness have been unable to reach a consensus on the optimal cumulative dosage of PA instruction. This failure to identify the optimal cumulative dosage can arise for at least two reasons. First, there may truly be no significant differences in PA instruction effectiveness as a function of cumulative dosage. Second, it may be that the way we measure the relation between PA instruction effectiveness and cumulative dosage is not the most compelling. The use of a linear meta-regression to study a moderator's effect is typical in the reading field, including the abovementioned meta-analyses. However, the relation between PA instruction effectiveness and cumulative dosage might be of a nonlinear rather than linear nature, which means that the linear methods will not be able to describe the data adequately. In the current study, we use a nonlinear meta-analytic approach to determine whether nonlinear relations between PA instruction effectiveness on PA outcomes and cumulative dosage are at play.

How Can We Statistically Model the Optimal Cumulative Dosage of PA Instruction?

In our report, we expand upon prior educational research emphasizing the importance of considering non-linearity for educational outcomes, including in meta-analyses (e.g. Dumas & McNeish, 2017, 2018; Roberts et al., 2022). This type of meta-analysis approach is guided by established optimization methods in pharmacology, where the efficacy of a drug (treatment group), in comparison to a placebo (control group), decreases after reaching the optimal cumulative dosage (Ogungbenro et al., 2009). Theoretical perspectives suggest that the effectiveness of PA instruction may exhibit a concave, parabolic shape (an upside-down U), characterized by increasing benefits up to a certain point of cumulative dosage (i.e. the maximum or the optimum), after which benefits begin to decline as a function of increased cumulative dosage. To model this shape for early PA instruction, we utilize a model developed by Cudeck and du Toit (2002). This model was initially employed in a study of medication effects in kidney transplant patients and later by Roberts et al. (2022), who applied it in the reading research literature. As illustrated in Figure 1, the model enables us to identify the cumulative dosage apex at which the maximum effect of treatment is achieved (in this study, the treatment refers

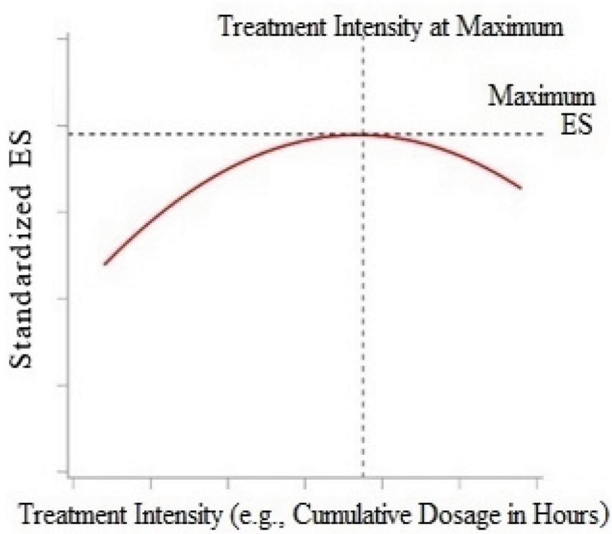

Figure 1. Conceptual diagram of optimal cumulative dosage - Model from Cudeck and du Toit (2002). *Note*. ES = effect size.

to early PA instruction). Note, however, that mathematically speaking, the model might not exhibit a concave, parabolic shape. Depending on the data, the model can also take a convex, parabolic shape (a U shape) characterized by a minimum. See Cudeck and du Toit (2002) for an example of a learning experiment where the function takes a parabolic shape with an associated minimum.

Nonlinear effects modeling within a meta-analysis has been scarcely used in reading research. In the only published study of its kind, Roberts and colleagues (Roberts et al., 2022) used a nonlinear meta-analytic approach to test the optimal cumulative dosage of reading interventions on various reading outcomes. Roberts et al. (2022) found that 186 effect sizes of reading interventions followed a concave parabolic shape with an associated maximum in cumulative dosage. The findings indicated the optimal cumulative dosage of reading interventions was 39.92 hours, with an associated overall maximum effect of d = 0.77. Moreover, the authors noted that the effect was not significantly different from the overall model for foundational skills (d = 0.90; 107 effect sizes). In addition, they found that the effects of one-on-one interventions increased indefinitely and did not have an associated cumulative dosage optimum. Roberts et al. (2022) study is the first step in outlining the optimal cumulative dosage of reading instruction using a nonlinear modeling approach in a meta-analysis.

Moderators

In examining the literature on the optimal cumulative dosage of PA instruction, it is essential to consider salient moderators that might impact observation of when the maximum effect occurs and its magnitude. In our study, we included PA instruction-level and participant-level moderators.

The first moderator was group size. Reducing group size is a commonly used strategy to intensify reading instruction and intervention. Theoretical explanations and empirical evidence suggest that as group size decreases, effect sizes do not necessarily improve (e.g. small group vs. individual instruction), alluding to a potential nonlinear association between group size and PA instructional effectiveness. Theoretically, one possible explanation is that there may be "diminishing returns" that come with the reduced size of instructional groups. For example, small group instruction allows the teacher to control the group's composition. Teachers may form a group of students with the same skill level on specific PA tasks and practice those skills. Alternatively, teachers could form groups of somewhat higher- and lower-performing students, with higherperforming students serving as models for lower-performing students. In both scenarios, opportunities for response are significantly increased. Students can see other student models and hear corrective feedback from teachers on their own and others' performance. In a one-on-one instructional setting, there might be more opportunities for exposure and feedback. Still, students lose the opportunity to view other students' responses and hear specific feedback for others. Simply increasing opportunities for exposure and feedback alone may not compensate for the additional opportunities lost when shifting from a small group to a one-on-one instructional format. Empirical evidence supports this theoretical explanation. For example, Vaughn et al. (2003) studied the effects of reading intervention on three group sizes (one teacher to one student, a small group of one teacher to three students, and a large group of one teacher to ten students). Results indicated that one-on-one grouping demonstrated more improvement than the large-group instruction for phoneme segmentation. However, the one-on-one grouping was not more effective than the small group on a PA outcome. The National Reading Panel (2000) also studied the effects of group-size on PA outcomes. The authors found that small-group instruction ($d = 1.38$) was more effective than whole-group instruction ($d = 0.67$) and individual instruction ($d = 0.60$) on PA outcomes. A similar result, though statistically non-significant, was reported in the Rice et al. (2022) meta-analysis. Small group PA instruction ($g = 0.70$) was descriptively more effective than the whole class ($g = 0.65$) or one-on-one ($g = 0.59$) instruction. However, note that a decrease in group size typically happens when reading instruction is intensified, alluding to a potential confound of student characteristics when discussing group size. For instance, students at risk for reading disabilities are typically taught in small groups, whereas students with reading disabilities could be trained in one-on-one settings. Thus, it is possible that a lack of benefit for one-on-one settings when compared to small-group instruction is due to one-on-one instruction typically being provided to the students with the most significant learning needs.

The second moderator was the use of letters with PA instruction. The National Reading Panel (2000) and a meta-analysis (Rice et al., 2022) reviewed studies where PA instruction was taught with and without letter representations. Within the studies reviewed, the authors of both reports found that using letters was not significantly associated with PA outcomes but did play a role in other reading outcomes (National Reading Panel, 2000). However, due to phonology playing an essential role in helping students establish word-specific orthographic representations (see also Share's (1995) self-teaching hypothesis), reading theorists emphasize the importance of explicitly incorporating letter knowledge instruction within PA and literacy instruction (e.g. L. C. Ehri & Roberts, 2006; Foorman & Francis, 1994; Hulme et al., 2012; Kim et al., 2010). Hence, it may be hypothesized that PA and alphabetic instruction in tandem might lead to more PA learning gains than PA instruction with no letters.

The third moderator was whether students were at risk for reading disabilities. The NRP report (2000) and the meta-analysis by Rice et al. (2022) reported no significant differences between typically developing readers and at-risk students on PA outcomes. In contrast, some studies found PA instruction to be more effective for students at-risk for reading disabilities than typically developing readers (Hatcher et al., 2004). Given a hypothesis that students at risk for reading disabilities might need an increased total PA instruction length to display a maximum PA instruction effect, further investigation is warranted for this moderator.

The fourth moderator was whether PA instruction targeted basic versus "advanced" PA skills. Previous research has suggested that basic skills like blending and segmentation are critical for reading development (L. C. Ehri et al., 2001). However, state education agencies and publishers have suggested "advanced PA training," including deletion and substitution, may be vital for reading development despite a lack of empirical evidence to support this claim (Clemens et al., 2021). Rice and colleagues (Rice et al., 2022) found no differences in the PA instruction effectiveness for this moderator. Nonetheless, since basic versus "advanced" PA skills progress from less to more complex (Schuele & Boudreau, 2008), we can hypothesize that a longer dosage of PA instruction on complex PA skills outcomes, such as deletion and substitution, may be required when "advanced" skills are evaluated as the outcome. However, given that previous research (Rice et al., 2022) has not shown any differences in effectiveness based on PA skills taught, it may be predicted that no difference in the maximum effect size magnitude will be observed.

The Current Study

In the current study, we used the same approach as Roberts et al. (2022) but focused on early PA instruction rather than reading interventions. The goal was to conduct a nonlinear meta-analysis and estimate the maximum effect size achieved by PA instruction and the optimal cumulative dosage at which the maximal effect occurred. Moreover, we aimed to determine what PA instruction and participant characteristics were salient in moderating the estimated optimal cumulative dosage of PA instruction and the estimated maximum effect. Our research questions were: (1) What is the optimal cumulative dosage of PA instruction and the maximum effect associated with that dosage on PA outcomes for pre-school through first-grade students? (2) When studying group size, the use of letters, risk for reading disabilities status, and types of PA skills as moderators, to what degree did the optimal cumulative dosage and maximum predicted effect vary from the overall optimal dosage and maximum predicted effect? Based on theory (C. A. Perfetti, 1992) and prior empirical reports (Roberts et al., 2022), we predicted that PA instruction effects would follow a concave parabolic shape (an upside-down U). We hypothesized that increasing cumulative dosage would improve PA instruction effectiveness to a point, after which the instruction would become less effective as the instruction's cumulative dosage increased. In addition, we hypothesized that our moderators would be significantly associated with the optimal cumulative dosage of PA instruction and the maximum effect at which this cumulative dosage occurred. Specifically, per National Reading Panel (2000) and Rice et al. (2022), we predicted the optimal cumulative dosage of PA instruction in small groups would be shorter and in one-on-one settings would be longer than the overall model. In accord with theory and empirical evidence (L. C. Ehri et al., 2001; Foorman & Francis, 1994), we expected the effect of PA instruction to be more accelerated if letters were added to the PA instruction. Based on theory (L. C. Ehri et al., 2001; C. A. Perfetti, 1992), we expected that the optimal cumulative dosage of PA instruction for students at risk for reading disabilities would be longer than the overall model. Finally, for "basic" PA skills, we predicted no differences from the overall model for the dosage and effect. To the extent that "advanced" PA skills were taught, we predicted that a longer optimal cumulative dosage of PA instruction might be required to observe a maximum PA instruction effect.

Method

Literature Search Procedure

We used four steps to search for and review studies relevant to our research questions (title and abstract screening/primary screening). We used one step to identify studies that adhered to our inclusionary and exclusionary criteria, which were included in the final meta-analysis (full-text screening/secondary screening). The search procedure resulted in 16 studies included in the final meta-analysis. Please see the flowchart in Figure SI1.

For the primary screening, we first conducted a comprehensive computerized search of PsycINFO, ERIC, and Academic Search Ultimate databases. We used Boolean search terms from four categories – PA, reading, instructor, and experimental design – linked with an AND. A full list of search terms is available in the Supplementary Materials. We conducted the search on January 25, 2022. We did not restrict our search to publication year. Peer-reviewed articles, book chapters, theses, and dissertations were queried. The electronic database search yielded a total of 8,961 publications (PsycINFO 2,942, ERIC 2,272, Academic Search 3,747). We removed 1,702 duplicates, so that the first step of the search yielded 7,259 publications for screening. We reviewed all studies by reading titles and abstracts on the platform Rayyan, locating any publications that were eligible for inclusion in the meta-analysis based on the inclusion and exclusion criteria. Through this screening process, we excluded 7,206 studies for different reasons (see Figure SI1 in Supplementary Materials). Two doctoral students completed the computerized screening with an agreement percentage of 99% on the overlapping titles/abstracts reviewed in Rayyan. Any disagreements were resolved through discussion. The first step of the primary screening resulted in 53 publications which were eligible for secondary screening. In the second step of primary screening (backward search), we examined the references of relevant articles and screened What Works Clearinghouse (WWC) Intervention Reports relevant to early PA instruction to search for additional potential studies to be included, which were not identified by the computerized search. Seven additional studies were identified through the second step. In the third step of primary screening, we performed a forward search of the relevant articles to find any studies that cited these works. We used Google Scholar with its "Cited By" link and the Web of Science Cited Reference Search. From this search method, four studies were identified. In the fourth step of primary screening, we aimed to identify gray literature. To this end, we emailed three authors who have published in the field and asked them to provide any unpublished data or literature relevant to the current study. No additional studies were found from this search method.

Altogether, the primary screening resulted in 64 studies that were eligible for secondary screening. Full articles from the resulting 64 studies were obtained and reviewed carefully for secondary screening for eligibility. Of the 64 studies, 48 studies were further excluded. See the Supplementary Materials for reasons for exclusion. This resulted in 16 studies that were included in the final meta-analysis. The first and second authors completed the secondary screening process with an agreement percentage of 98% for the included studies. The one disagreement was resolved through discussion and consulting the original article.

This study is a meta-analysis, which only involves analysis of aggregate data extracted from primary studies. All data provided in primary studies were de-identified. Hence, we did not request informed consent from parents or assent from included children prior to their inclusion in the meta-analysis. The meta-analysis conforms to recognized standards from U.S. Federal Policy for the Protection of Human Subjects.

Search Terms

Please see the Supplementary Materials for details.

Inclusion and Exclusion Criteria

We used the following inclusion and exclusion criteria. First, a study must have implemented a group experimental or quasi-experimental design for evaluating PA instruction with a control group for comparison. Studies not including a control group were excluded. Second, the instruction must have been focused on PA skills and included at least one measure of PA as an outcome measure (i.e., one or more of the PA skills listed in the Introduction). Studies may have also included instruction in other phonological skills (e.g., onset-rime or syllable clapping). However, they were excluded if they only focused on phonological awareness skills without including the phoneme-level skills. Studies with instruction progressing further than PA instruction, such as reading decodable texts, were also excluded. Third, a study must have provided an estimate for the total cumulative dosage of PA instruction (in minutes or hours) or data to compute it. Fourth, a study must have provided sufficient quantitative information (pre- and posttest score data) to permit effect size calculation. The same PA measure must have been administered at pre- and posttest to calculate the effect size. Fifth, participants in a study must have been in early childhood programs, preschool, kindergarten, or first grade. A study was also included if disaggregated data for one or more of these grade levels were reported. Sixth, a study must have implemented PA instruction delivered by a classroom teacher, parent, and/or computer. Seventh, a study must have been published in English, and the PA instruction must have been conducted in English.

Study Coding Procedures

Coding was completed using the full texts of the 16 studies eligible for our meta-analysis. A codebook was created with the following information listed by study: study information, instruction characteristics, participant characteristics, and outcome measures. For study information, studies were coded for publication type (i.e., journal article or dissertation/thesis) and study design (i.e., randomized controlled trial or quasi-experimental design). For instruction characteristics, studies were coded for group size, phonemic awareness (PA) skills taught, use of letters, and cumulative dosage. Group size was coded as whole class, small group (i.e., any number dividing the class into groups), or one-on-one. PA skills taught was coded as basic PA skills (i.e., identification, isolation, categorization, blending, or segmenting only) or advanced PA skills (i.e., had to teach deletion and/or substitution but could also include basic PA skills). Use of letters was coded as no (i.e., no letters were taught/used with the PA instruction) or yes (i.e., letters were taught/used along with the PA instruction). Cumulative dosage was coded as the average cumulative dosage reported in the primary studies in hours. When the average cumulative dosage was not available, we calculated the product of duration, dose, and dose frequency to compute the cumulative dosage intensity (Roberts et al., 2022; S. F. Warren et al., 2007). As to participant characteristics, studies were coded for grade level, and whether students were at-risk for reading disabilities. Grade level was coded as preschool, kindergarten, first grade, or mixed grade levels. Risk status was coded as at-risk (i.e., the authors of primary studies identified the participants as at-risk for reading disabilities) or low risk (i.e., the authors of primary studies did not describe the participants as being at-risk for reading disabilities). Students who were at risk were described by authors as having low PA skills, low socioeconomic status, teacher referral, or as receiving special education services. We also extracted quantitative information on the outcome measures (i.e., means and standard deviations for both pre- and posttest) for treatment and control groups along with the number of participants in each group to calculate effect sizes.

The study coding process was completed by doctoral students in educational psychology with prior research experience in meta-analysis in the reading field. The second author developed a detailed codebook and trained another doctoral student on the coding categories and definitions. Two practice studies were coded independently, and then the coders met to confirm agreement on coding. After confirming the coding, the raters completed coding the remaining studies independently and overlapped on 65% of the included studies. The agreement percentages ranged from 94% to 100% across all categories, and disagreements were resolved through discussion and consulting the original article.

Effect Size Calculations and Modeling the Optimal cumulative Dosage of PA Instruction

As proposed by Becker (1988), the standardized mean-change measure was used as the measure of effect size. We followed established practices (Roberts et al., 2022) and chose this effect size because it helped us estimate and model the optimal treatment cumulative dosage. To model the outcome in Equation 2, we needed an effect size representing the difference between treatment and control groups. The standardized mean-change measure effect size (Becker, 1988) compares the standardized pre- to post-PA instruction change in the treatment group to the standardized pre- to post-PA instruction change in the control group. Standardization is formulated so that it is relative to the pretest standard deviation. In terms of the metric, the standardized mean-change effect size is on the same metric as Cohen's d (however, see the Limitations section). In terms of conceptualization, this effect size is akin to a standardized effect for an entire two by two design, and it incorporates pre- and post-instruction change across treatment and control groups. In terms of interpretability, a standardized mean-change measure effect size estimate of 2.0 would mean that the standardized change in the PA instruction group is two pretest standard deviations larger than the standardized change in the control group (Becker, 1988). The standardized mean-change measure effect size is calculated as shown in Equation 1 (Becker, 1988).

$$d_{si} = \left(1 - \frac{3}{4(nI-1)-1}\right) * \left(\frac{MpostI - MpreI}{SDpreI}\right) - \left(1 - \frac{3}{4(nC-1)-1}\right) * \left(\frac{MpostC - MpreC}{SDpreC}\right) \quad (1)$$

where

- d_{si} is the standardized effect size nested within PA instruction *i* in a study *s*,
- *MpostI* is the post-PA instruction mean of the treatment group,
- *MpreI* is the pre-PA instruction mean of the treatment group,
- *SDpreI* is the standard deviation of pretest scores in the treatment group,
- *nI* is the sample size in the treatment group,
- *MpostC* is the posttreatment mean of the control group,
- *MpreC* is the pretreatment mean of the control group,
- *SDpreC* is the pretest standard deviation of the control group,
- *nC* is the sample size in the control group,
- and $\left(1 - \frac{3}{4(n-1)-1}\right)$ is a small-sample bias correction.

As previously mentioned, this meta-analysis aimed to identify the maximum effect size and optimal cumulative dosage associated with that effect size in early PA instruction. To accomplish this, we used an approach briefly described in the Introduction. The model is formulated as a nonlinear mixed effects model (Equation 2), such that

$$d_{si} = \beta_{1i} - (\beta_{1i} - \beta_{0i}) * \left(\frac{Dosage\ in\ hours_{si}}{\beta_{2i}} - 1\right)^2 + e_{si} \quad (2)$$

$\beta_{0i} = \gamma_{00} + u_{01}$
$\beta_{1i} = \gamma_{10}$
$\beta_{2i} = \gamma_{20}$
where

- d_{si} is the standardized effect size nested within PA instruction *i* in a study *s*,
- β_{0i} is the standardized effect size for a hypothetical PA instruction with a cumulative dosage of zero (i.e., the intercept) in PA instruction *i*,
- β_{1i} is the maximum standardized effect size achieved in PA instruction *i*,
- β_{2i} is the cumulative dosage factor at which the maximum value occurs in PA instruction *i*,
- and e_{si} is a normally distributed error term for study *s* in PA instruction *i* that captures the distance between the observed standardized effect size and the predicted standardized effect size.

To account for the nesting of effect sizes within PA samples (e.g., multiple effect sizes were derived from a single PA instructional program), the intercept β_{0i} is further decomposed such that $\beta_{0i} = \gamma_{00} + u_{01}$ where γ_{00} is a fixed effect that captures the average intercept across all articles and u_{01} is a normally distributed random effect with a mean of zero and variance of tau [$u_{01} \sim N(0, \tau)$], which captures the deviation of PA instruction i's intercept from the fixed effect. We fit the model as shown in Equation 2 in SAS Proc NLMIXED with maximum likelihood estimation via adaptive Gaussian quadrature with quasi-Newton optimization. We included random intercepts in the model to account for statistical dependencies of multiple effect sizes from the same sample (i.e., effect sizes nested within samples/studies). Because sample sizes vary across studies, we used normalized weights to prioritize the contribution of effect sizes from larger and presumably more reliable studies while retaining unbiased standard errors in our models (Roberts et al., 2022).

Moderator Analyses

We conducted separate moderator analyses for each potential moderator. We added moderating effects of effect size characteristics to the equations for $\beta_{0i,}$ $\beta_{1i,}$ and β_{2i}. We tested to what extent the intercept, maximum effect size, and optimal cumulative dosage changed as a function of the moderators. We fit the model much the same way as described in the previous section for the overall model. We compared values from the overall model with those from the moderator analyses using custom hypothesis testing as available in the ESTIMATE statement in SAS. The statistical significance level was set at .05.

Note that in our nonlinear meta-analysis, we did not base our decisions on conducting moderator analyses on factors like the number of effect sizes and/or other rules of thumb typically used in linear meta-regression models (e.g., requiring df > 4). Instead, we followed the recommendation by Timmons and Preacher (2015). Specifically, we examined the spacing and concentration of individual data points (i.e., effect sizes) around the area of the greatest curvature, which is mathematically determined by the absolute value of the function's second derivative with respect to time. If there were only few data points concentrated around the apex, we could not have complete confidence in the precision of our nonlinear curve estimation. Hence, we did not report the results for those moderators. Simulations have demonstrated that for the nonlinear curve to be fully estimated with precision, it is more important to consider the spacing and concentration of individual data points around the apex of the curve than the raw number of data points (Timmons & Preacher, 2015).

Results

Descriptive Statistics

This meta-analysis included 16 studies with 20 treatment conditions and 35 effect sizes. There were 613 students who had data at posttest in treatment conditions and 542 students with data in control conditions across all studies. Across the independent samples that reported age, the mean age was 5.20 years ($SD = 0.87$, range 4.17–6.67). Across the studies, the cumulative dosage ranged from 2 hours (Nutkins, 2004) to 35 hours (P. Warren, 2009). Six treatment conditions delivered instruction in a small group, and 14 delivered one-on-one instruction. Thirteen treatment conditions included letters in their PA instruction, whereas 7 treatments did not. Sixteen treatments were administered to students who might be at risk for reading disabilities; and 4 treatments were delivered to students with low risk for reading disabilities. Among the treatments provided, there were 14 treatments that taught basic PA skills, 5 treatments that taught advanced PA skills, and 1 treatment condition that did not report data on the type of PA skills taught (Kelly et al., 2019; the study was retained in this meta-analysis). As to the control groups, 14 groups received no treatment or business-as-usual, four engaged in math games, and three received instruction focused on another literacy skill (e.g., grammar, vocabulary). See Table 1 for more information about the included studies.

Effect Size Calculations and Modeling the Optimal Cumulative Dosage of PA Instruction

To answer our first research question, the overall model was fit to 35 effect sizes. As predicted, the nonlinear cumulative dosage response model took a concave parabolic form with the maximal effect size for early PA instruction at $d = 0.74$. This effect was estimated to occur at 10.20 hours ($p = .0005$) of PA instruction, given the concave and parabolic pattern in the data.

Figure 2 depicts the nonlinear optimal cumulative dosage curve estimated by the model (blue curve) both without (left panel (a)) and with superimposed individual data points (right panel (b)). As indicated in the left panel of Figure 2, the nonlinear function which describes the estimated effects of PA instruction increases from the shortest included PA instruction (2 hours) until the maximal time point of approximately 10 hours, after which it plateaus and decreases until it reaches the longest included PA instruction in this meta-analysis (35 hours). The right panel of Figure 2 includes bubbles

Table 1. Description of studies included in the meta-analysis with effects on PA outcomes.

Study	Independent sample number	*N* (T, C)	Grade level	Study design	Cumulative dosage (in hours)	Group size[a]	PA skills taught, use of letters during instruction	At-risk status	ES (standardized mean-change)	PA outcome measure	Brief description
Anthony (2016)	1	122, 125	K	RCT	31.5	1:1	B, NL	AR	−0.15	deletion	Earobics Step 1 compared to math computer games.
	1	122, 125	K	RCT	31.5	1:1	B, NL	AR	0.07	blending	
Barker and Torgesen (1995)	1	18, 18	1	RCT	8	1:1	B, NL	AR	1.80	deletion	Daisy's Quest & Daisy's Castle compared to math computer games.
	1	18, 18	1	RCT	8	1:1	B, NL	AR	1.36	segmentation	
	1	18, 18	1	RCT	8	1:1	B, NL	AR	0.61	blending	
	1	18, 18	1	RCT	8	1:1	B, NL	AR	0.52	categorizing	
Gale (2006)	1	13, 13	K	RCT	8.3	1:1	B, L	AR	2.36	isolation	Earobics compared to no treatment.
	2	13, 12	1	RCT	8.3	1:1	B, L	AR	1.39	segmentation	Earobics compared to no treatment.
	3	12, 12	1	RCT	8.3	1:1	B, L	AR	0.50	segmentation	Lexia Reading compared to no treatment.
Hatcher et al. (2004)	1	82, 61	Mixed (PreK & K)	RCT	20.46	SG (10–15)	A, L	LR	0.21	deletion	PA instruction compared to BAU.
Hecht & Close, 2002	1	42, 34	K	QED	21.43	1:1	A, L	AR	1.79	segmentation	Waterford Early Reading compared to BAU.
	1	42, 34	K	QED	21.43	1:1	A, L	AR	0.86	blending	
Hudson et al., (2017)	1	42, 44	PreK	RCT	12	1:1	A, NL	AR	0.59	composite/ total	Road to the Code by tutors compared to BAU.
Hund-Reid (2008)	1	22, 8	K	RCT	18.45	SG (2)	B, L	AR	0.67	composite/ total	Road to the Code by paraeducators compared to BAU.
Kelly et al. (2019)	1	51, 53	PreK	RCT	24	SG (4–7)	NR, L	LR	0.62	composite/ total	Cracking the Code compared to grammar control group.
Kyle et al. (2013)	1	10, 10	1	QED	10.73	1:1	B, L	AR	0.54	deletion	GraphoGame phoneme compared to no treatment.
Lonigan et al., (2003)	1	20, 21	PreK	QED	5.46	1:1	B, L	AR	−0.09	blending	Daisy's Quest & Daisy's Castle compared to BAU.
	1	20, 21	PreK	QED	5.46	1:1	B, L	AR	−0.31	deletion	
	1	20, 21	PreK	QED	5.46	1:1	B, L	AR	0.22	blending	

(*Continued*)

Table 1. (Continued).

Study	Independent sample number	*N* (T, C)	Grade level	Study design	Cumulative dosage (in hours)	Group size[a]	PA skills taught, use of letters during instruction	At-risk status	ES (standardized mean-change)	PA outcome measure	Brief description
Mitchell & Fox (2001)	1	24, 24	Mixed (K &1)	RCT	5	SG (6)	B, NL	AR	0.39	blending	Phonological Awareness Kit compared to a math computer games.
	1	24, 24	Mixed (K &1)	RCT	5	SG (6)	B, NL	AR	1.07	isolation	
	1	24, 24	Mixed (K &1)	RCT	5	SG (6)	B, NL	AR	0.75	segmentation	
	1	24, 24	Mixed (K &1)	RCT	5	SG (6)	B, NL	AR	0.86	composite/ total	
	2	24, 24	Mixed (K & 1)	RCT	5	1:1	B, NL	AR	0.27	blending	Daisy's Quest & Daisy's Castle compared to math computer games.
	2	24, 24	Mixed (K & 1)	RCT	5	1:1	B, NL	AR	1.02	isolation	
	2	24, 24	Mixed (K & 1)	RCT	5	1:1	B, NL	AR	0.30	segmentation	
	2	24, 24	Mixed (K & 1)	RCT	5	1:1	B, NL	AR	0.86	composite/ total	
Nelson et al., (2009)	1	41, 47	PreK	RCT	16.67	SG (2–6)	A, L	AR	0.22	combined/ total	Stepping Stones to Literacy with tutors compared to vocabulary group.
Nutkins (2004)	1	21, 10	PreK	RCT	2	SG (NR)	B, L	AR	2.23	isolation	PA instruction by teachers compared to BAU.
Rehmann (2005)	1	14, 16	K	RCT	10	1:1	B, L	LR	0.33	isolation	Earobics compared to BAU.
	2	16, 19	1	RCT	10	1:1	B, L	LR	0.10	segmentation	Earobics compared to BAU.
Travis (1997)	1	21, 22	1	RCT	12	1:1	B, NL	AR	0.29	segmentation	Daisy's Quest & Daisy's Castle compared to no treatment.
	1	21, 22	1	RCT	12	1:1	B, NL	AR	0.19	deletion	
	1	21, 22	1	RCT	12	1:1	B, NL	AR	0.15	blending	
P. Warren (2009)	1	5, 5	Mixed (K &1)	RCT	35	1:1	A, NL	AR	−0.10	segmentation	PA instruction by parents compared to read-aloud only group.

Note. *N* = number of students; *T* = treatment group; C = control group; PA = phonemic awareness; PreK = preschool or Head Start; K=kindergarten; 1= first grade; RCT = randomized control trial; QED = quasi-experimental design; 1:1 = one-on-one; SG = small group; B = basic PA skills were taught during instruction; A = advanced PA skills were taught during instruction; L = letters included in instruction; NL = no letters included in instruction; AR = at-risk; LR = low risk; NR = not reported; ES = effect size; BAU = business as usual.

[a]= The number in parentheses after SG represents the number of students in the small groups.

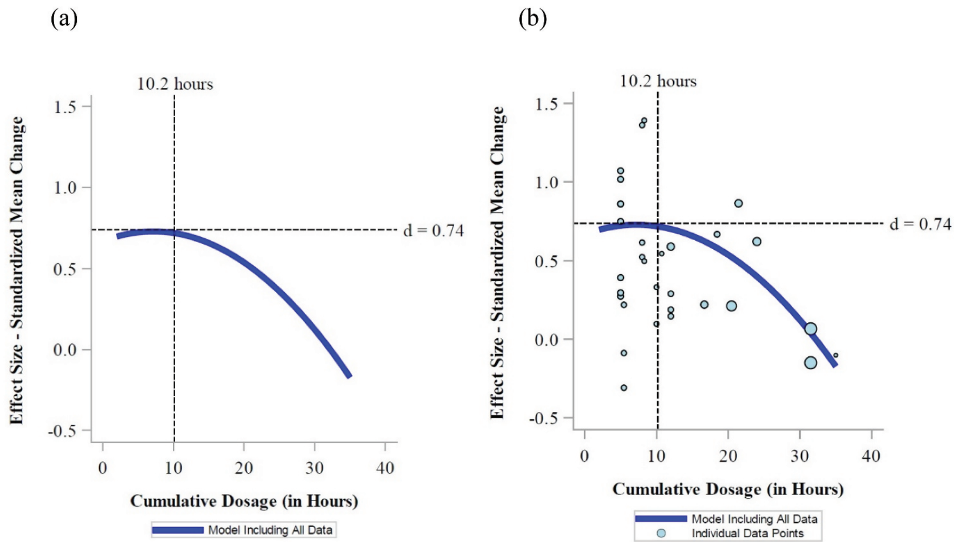

Figure 2. Nonlinear optimal cumulative dosage curves for the overall Model without (a) and with individual data points (b). *Note*. The blue curve is identical in both panels. The overall model estimated on 35 effect sizes had a maximal effect of $d = 0.74$ with the optimal cumulative dosage of 10.20 hours. The maximal estimated effect is shown by the horizontal dotted line, and the optimal cumulative dosage is shown by the vertical dotted line.

to represent individual effect sizes. The size of the bubbles varies, such that larger bubbles represent larger samples. The blue curve is the same nonlinear optimal cumulative dosage curve as the one in the left panel. The parameters of the overall model, as depicted in Figure 2, were then used as a point of comparison for the moderator analyses.

Moderator Analyses

To answer our second research question, we conducted moderator analyses. We conducted a separate nonlinear model for the subset of effect sizes relevant to each moderator.

Group Size

The optimal cumulative dosage curve for individual PA instruction (26 effect sizes) is depicted in Figure 3 (blue curve). For this subset of PA instruction, the maximal effect was larger ($d_{max} = 0.99$) than the overall maximal effect ($d = 0.74$), however, this difference was not statistically significant, $t(15) = 1.18$, $p = .26$. As predicted, the optimal cumulative dosage for individual PA instruction was 16.63 hours ($p < .0001$), which was significantly longer than the overall optimal cumulative dosage, $t(15) = 4.77$, $p < .0001$.

We were not able to model the nonlinear effects of small-group PA instruction with precision. Nine effect sizes were not concentrated where the curve of the function was the greatest, suggesting the suboptimal recovery of parameters for this moderator.

Use of Letters in PA Instruction

The curve for PA instruction with no letters (22 effect sizes) is shown in the left panel of Figure 4. As indicated in the figure, PA instruction with no letters was not associated with a significantly distinct optimal maximum. Neither the effect ($t(15) = -0.86$, $p = .41$) nor the dosage ($t(15) = 0.54$, $p = .60$) were significantly different from the overall model.

The right panel of Figure 4 shows the optimal cumulative dosage curve for PA instruction that used letters (blue curve; 13 effect sizes). Unlike the form of the overall model (black

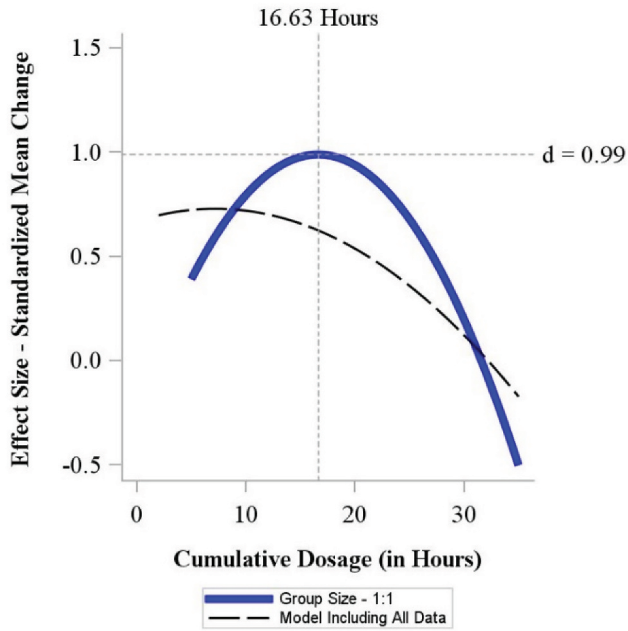

Figure 3. Nonlinear optimal cumulative dosage curves for the group size moderator. *Note*. The black dashed curve shows the overall model, which is also depicted in Figure 2. The black dashed curve is identical in both panels. The overall model had a maximal effect of $d = 0.74$ with the optimal cumulative dosage of 10.20 hours. The blue curves are unique to each panel and show the nonlinear cumulative dosage curves for the specific set of PA instructions featuring the specific moderator (in this case, group size). Maximal estimated effects are shown by the horizontal dotted lines, and the optimal cumulative dosage is shown by the vertical dotted lines. The model for individual PA instruction was estimated on 26 effect sizes.

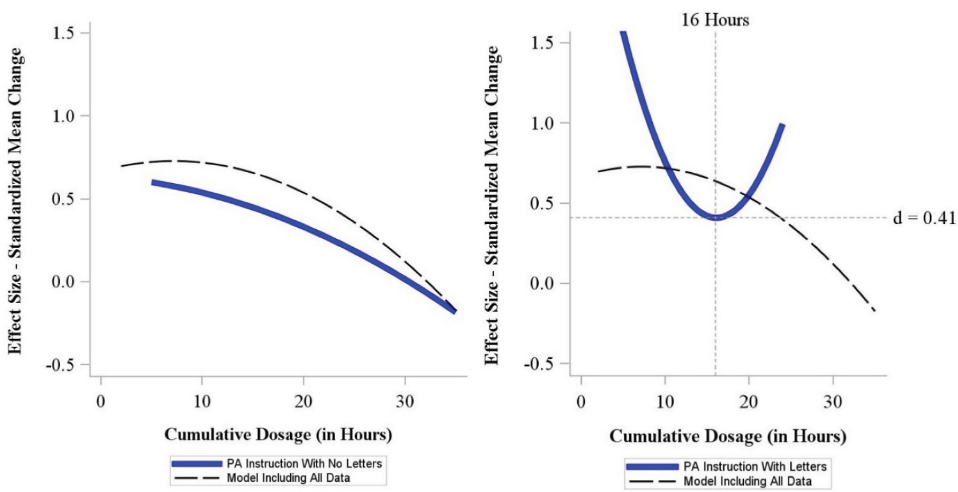

Figure 4. Nonlinear optimal cumulative dosage curves for the use of letters in PA instruction moderator. *Note*. The black dashed curve shows the overall model, which is also depicted in Figure 2. The black dashed curve is identical in both panels. The overall model had a maximal effect of $d = 0.74$ with the optimal cumulative dosage of 10.20 hours. The blue curves are unique to each panel and show the nonlinear cumulative dosage curves for the specific set of PA instructions featuring the specific moderator (in this case, use of letters in PA instruction). Minimal estimated effects are shown by the horizontal dotted lines, and the optimal cumulative dosage is shown by the vertical dotted lines. The model for PA instruction with no letters (left panel) was estimated on 22 effect sizes, and for PA instruction that used letters (right panel) on 13 effect sizes.

dashed curve), the subset of PA instruction with letters displayed a convex shape. The minimum effect ($d_{min} = 0.41$) was estimated at 16.10 hours ($p < .0001$), after which it increased as the cumulative dosage increased. The differences from the overall model were not statistically significant for the effect ($t(15) = -1.37$, $p = .19$). However, there were differences in the optimal cumulative dosage, which was significantly longer than the overall model ($t(15) = 3.77$, $p = .002$).

At Risk for Reading Disabilities Status

The nonlinear optimal cumulative dosage curve for PA instruction for students at risk for reading disabilities (31 reported effect sizes) is shown in Figure 5. For this group of students, it appears that the maximal effect was higher ($d_{max} = 0.79$) than the overall effect ($d = 0.74$). However, this maximum was not significantly different from the overall model ($t(15) = 0.34$, $p = .74$). Similarly, the optimal cumulative dosage for this group was longer (10.84 hours; $p < .0001$). However, this estimate was not significantly different from the overall model (10.20 hours; $t(15) = 0.10$, $p = .92$). The findings were contrary to our hypothesis.

Unfortunately, we were not able to model the nonlinear effects (4 effect sizes) of PA instruction for students who were not at risk for reading disabilities with precision.

PA Skills Taught

The nonlinear optimal cumulative dosage curve for PA instruction that included PA identification, isolation, categorization, blending, and/or segmentation only ("basic" PA skills) is shown in Figure 6 (28 effect sizes). As predicted and indicated in the figure, PA instruction with basic PA skills did not significantly vary by cumulative dosage. The almost straight line represents a functional form with no associated optimal maximum or minimum. The estimates

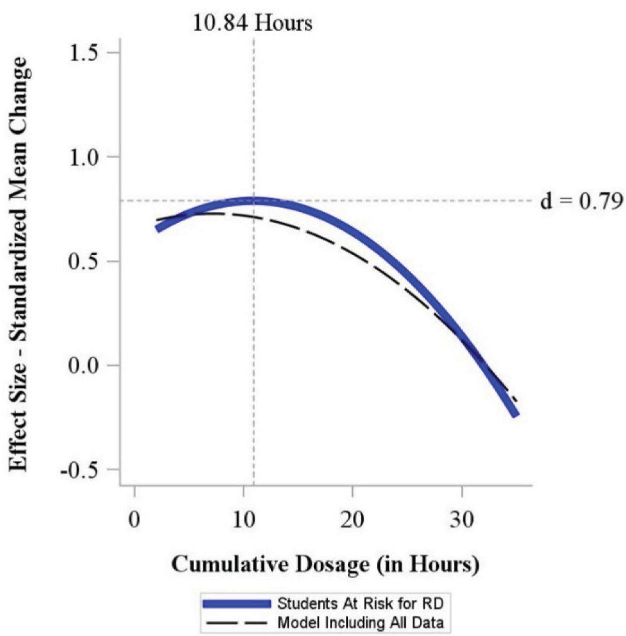

Figure 5. Nonlinear optimal cumulative dosage curves for the at risk for reading disabilities status moderator. *Note.* The black dashed curve shows the overall model, which is also depicted in Figure 2. The overall model had a maximal effect of $d = 0.74$ with the optimal cumulative dosage of 10.20 hours. The blue curve is unique to the panel and shows the nonlinear cumulative dosage curve for the specific set of PA instructions featuring the specific moderator (in this case, at risk for reading disabilities status). The maximal estimated effect is shown by the horizontal dotted lines, and the optimal cumulative dosage is shown by the vertical dotted lines. The model for PA instruction for students at risk for reading disabilities was estimated on 31 effect sizes.

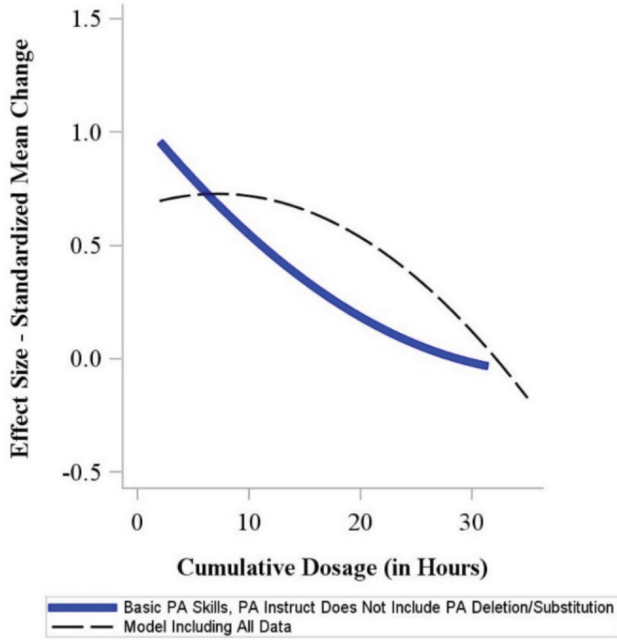

Figure 6. Nonlinear optimal cumulative dosage curves for the PA skills taught moderator. *Note*. The black dashed curve shows the overall model, which is also depicted in Figure 2. The black dashed curve is identical in both panels. The overall model had a maximal effect of $d = 0.74$ with the optimal cumulative dosage of 10.20 hours. The blue curves are unique to each panel and show the nonlinear cumulative dosage curves for the specific set of PA instructions featuring the specific moderator (in this case, PA skills taught). The model for PA instruction that included basic PA skills was estimated on 28 effect sizes.

were not significantly different from the overall model for the effect ($t(15) = 0.71$, $p = .49$) or the cumulative dosage ($t(15) = -0.40$, $p = .70$).

We were not able to model the nonlinear effects of PA instruction that included "advanced" PA skills with precision. Six effect sizes were not concentrated where the curve of the function was the greatest, suggesting the suboptimal recovery of parameters for this moderator.

Discussion

When students experience difficulties with PA or do not demonstrate adequate progress following PA instruction, a seemingly straightforward way to enhance PA skills acquisition is by increasing PA instruction dosage or intensifying instruction in another manner (e.g., providing small-group instruction). However, studies utilizing linear models have failed to provide evidence that extending PA instruction cumulative dosage yields a significantly greater impact than PA instruction with a shorter cumulative dosage. This is potentially due to diminishing returns in the magnitude of effects once students reach a certain level of proficiency in PA skills. As a result, the present study sought to model the nonlinear relations of PA instruction cumulative dosage to gain a better understanding of the association between optimal cumulative dosage of PA instruction and PA instruction effectiveness. There are several noteworthy results in our study. The optimal cumulative dosage of PA instruction across all effect sizes was 10.20 hours, with a maximal associated effect size of $d_{max} = 0.74$. This optimal cumulative dosage differed as a function of one moderator, the use of letters in PA instruction. Our findings have important implications for PA instruction in school settings.

As expected, our finding of identifying the optimal cumulative dosage of PA instruction appears to be in convergence with the theoretical perspectives and empirical research. Based on reading theories, typically developing readers will acquire PA skills and go on to develop phonemic-orthographic associations with increased instruction and print experience (C. A. Perfetti, 1992). As students develop automaticity in PA and sublexical orthographic-phonological representations, there might be a point after which additional PA instruction would no longer yield significant PA gains. The results of our study suggest that this might happen at around 10.20 hours of PA instruction in studies reported here (all studies reported small group or one-on-one PA instruction, mostly with students at-risk for reading disabilities). Given the studies' characteristics, we might assume that this optimal estimate of 10.20 hours indicates an additional 10.20 hours of instruction on top of typical whole group PA instruction. However, note that our findings should not be used to dictate an oversimplified prescription regarding dosage. Students will differ in the time they need to acquire PA and bond orthographic and phonological representations. Moreover, there are many other factors that govern the effectiveness of instruction (see also PA instruction with letters and limitations sections).

Empirically, our results align with prior research. In the most recent nonlinear meta-analysis on the optimal cumulative dosage of reading interventions, Roberts and colleagues (Roberts et al., 2022) identified the optimal cumulative dosage for reading interventions, which was estimated to happen at 39.92 hours. The results of our report and Roberts et al. (2022) study align in that students appear to make the highest gains up to a certain point in instruction and reading interventions, after which the effects of continued instruction decline. As such, the findings suggest that any sizable gains in PA outcomes after the optimal cumulative dosage points will not increase substantially. Our findings are in keeping with other reports as well. The National Reading Panel (2000) indicated that effect sizes were significantly larger for the cumulative dosage of PA instruction lasting from 5 to 9.3 hours and from 10 to 18 hours. PA instruction dosage that was either shorter or longer was less effective for teaching PA skills.

As to moderators, the maximum effect appears to take longer to achieve in a one-on-one setting. Extrapolating the findings from research and practice that students with the most significant learning needs are commonly taught in one-on-one settings within the response-to-intervention framework, the increased dosage to detect a non-significantly different effect from the overall model seems plausible for the 1:1 subset of studies. However, this finding is not in line with some of the previous research (e.g., Vaughn et al., 2010) that found that higher cumulative dosage was associated with larger effect sizes in one-on-one reading intervention settings. The differences between Vaughn et al. (2010) and our subset of studies might stem from differences in student characteristics across studies. The population of students that used one-on-one PA instruction in our study comprised both students who were and were not at risk for reading disabilities, whereas all students in Vaughn et al. (2010) were at risk for reading disabilities.

Turning to letters used in PA instruction, our findings suggested that the overall model was generalizable to the PA instruction with no letters. This finding was in line with our prediction, and implies that students may make the highest gains relatively early in PA instruction. However, by coupling PA instruction with letter instruction, our results imply that after around 16 hours of PA instruction with letters, the effectiveness of instruction will continually increase with increasing cumulative dosage. The interpretation of the curve, which exhibits a U-shaped pattern with diminishing effect sizes up to 16 hours of instruction, poses some challenges. A straightforward and mathematically plausible interpretation is that given the convex rather than concave pattern of a fairly small number of data points (13 effect sizes) for this subset, the model estimated a minimal rather than a maximal effect size. For comparison, previous research (Roberts et al., 2022) conducting the same type of nonlinear meta-analysis estimated the model for foundational skills on 107 effect sizes. Direct comparison with our results is difficult, though, given the referenced research's definition of foundational skills was phonological awareness, phonics, and word recognition. Additionally, some effect sizes were included across multiple categories (e.g., foundational skills and multicomponent interventions; Roberts et al., 2022).

The other possible interpretation for the U-shaped pattern is as follows. As students' phonemic and phonological representations become further refined with instruction and exposure to orthography, students need time to learn most of the letter names, many of the sounds that go with them, and a successful amalgamation of them. Then, once students show progress in mastering orthography-phonology mapping, including letters appears to accelerate the process of acquiring PA skills. We observe this increase in the curve's progress beyond the 16-hour mark, which suggests that incorporating letters to PA instruction and PA into alphabetic instruction would be considered a practically meaningful and effective way to maximize PA gains. It is important to note that the curve's endpoint corresponds to the latest available dosage data point for this subset of studies, which was at 24 hours (Kelly et al., 2019). This should not be misconstrued as a definitive limit. Had there been dosage data on PA instruction with letters extending beyond the currently reported dosage in the research literature, we might have observed a maximum effect followed by a subsequent decrease. Altogether, our results are interpreted here as lending further support for item-based reading theories (e.g., Harm & Seidenberg, 1999; C. A. Perfetti, 1992), including the bootstrapping hypothesis – a proposition highlighting the significant role of phonology in helping the child establish word-specific orthographic representations (see also Share, 1995, and the phonological recoding mechanism that plays a role in child's self-teaching in learning to read). In fact, upon scrutinizing the bootstrapping function of PA from our primary studies, positive effects were discerned across reading outcomes, including letter sound fluency, nonsense word reading, and word reading (see Table SI1 in Supplementary Materials). Such a finding is compatible with established reading practices, showing that providing students exposure to orthography and ample opportunities – also in the form of increased cumulative dosage – for the mastery of the orthography-phonology mapping is critical in learning to read and spell in English. Empirical evidence undisputedly shows that PA and knowledge of the alphabetic principle knowledge are powerful longitudinal predictors of how well children learn to read and spell (e.g., Foorman, 2003; Muter et al., 2004).

Regarding students at risk for reading disabilities, our results were contrary to our expectations. The results revealed that this population's nonlinear cumulative dosage response curve displayed a similar form to the overall cumulative dosage curve, such that the magnitude of the effect sizes reached a maximum at around 11 hours of PA instruction, which was not different from the overall model. On the one hand, this finding is striking. Because students at risk for reading disabilities, including dyslexia, have deficits in acquiring phonemic representations and linking them to corresponding graphemic representations (L. C. Ehri & Saltmarsh, 1995), they typically require additional and more intensive instructional practices to help build sublexical orthographic-phonological associations and display evidence of PA gains. For example, the Roberts et al. (2022) study identified that 46.46 hours of instruction on foundational reading skills, including PA, is optimal for reading interventions for students at-risk for disabilities. On the other hand, the lack of differences between the reading disabilities subset and the overall model is not as astonishing because 31 of the 35 total effect sizes were reported for students at-risk. It remains possible, of course, that a curve with a later decline would emerge if more data were available to us for the cumulative dosage for this population. However, it also seems likely that data on dosage are more often reported in articles that study at-risk populations receiving PA interventions than typically developing populations receiving only PA instruction with no interventions. For example, of the 26 studies excluded due to a lack of reporting on dosage data (see the flowchart in Figure SI1), 17 were conducted on students with no risk and 9 on at-risk populations. Overall, at this point, we can call for further research with systematically reported cumulative dosage data to be able to elucidate the conditions under which PA instruction would be maximally effective for typically developing as well as at-risk readers.

Our hypothesis was confirmed regarding the moderator of teaching basic PA skills. The optimal cumulative dosage to achieve maximal PA instruction effects was not associated with a different maximum and was not statistically significantly different in dosage from the overall model. Based on our results, we could argue it is reasonable to accept that children gain access and are able to manipulate phonemic representations after 10.20 hours of basic PA instruction. Namely, instruction of basic PA skills benefited not only lower-level PA skills, such as isolation, blending, and segmentation, but also higher-level skills, such as manipulating phonemes (e.g., deletion; Barker & Torgesen, 1995; Hatcher et al., 2004; Kyle et al., 2013). Moreover, in accordance with Share's self-teaching hypothesis (Share, 1995), it appears that basic PA instruction facilitated bootstrapping word-specific orthographic representation. For example, of the ten primary studies on basic PA instruction, nine reported positive outcomes on orthographic representations in nonsense as well as word reading (see Table SI1). However, one confound to consider is the use of letters in PA instruction. Because the number of effect sizes was somewhat limited across moderators, we were not able to statistically estimate the interaction effects among the type of PA instruction and use of letters.

Limitations

Although our meta-analysis provides an important foundation for considering to what extent instruction dosage is associated with the efficacy of PA instruction, we must be cautious in interpreting our findings. First, there were 16 studies with 35 effect sizes retained for the analysis. A larger sample size would have allowed us to model additional moderators in the nonlinear analysis (e.g., small group, advanced PA skills). The main reason our sample size was reduced lies in insufficient dosage and effect size reporting. Of 64 eligible studies, more than half were excluded for not adequately reporting data needed to compute cumulative dosage and effect sizes (26 and 10, respectively).

Second, we cannot draw conclusions regarding the optimal cumulative dosage of PA instruction for individual students. Given that PA acquisition, bonding of orthographic-phonological representations, and building high-quality word representations are item- and child-specific (Harm & Seidenberg, 1999; C. Perfetti, 2007), individual student and item factors will result in different dosage response patterns. Nonetheless, to begin to build a picture of how much dosage is just enough to be associated with the largest effect, we examined the data at the group level, including moderators such as student risk status for reading disabilities. Even among studies of students at risk for reading disabilities, though, students are on a continuum regarding being at risk. Hence, it might be that more extensive cumulative dosages of treatment might be required to achieve the same effects for different students.

Future research should explore in greater detail the extent to which student factors influence the effectiveness of PA instruction and evaluate to what degree the cumulative dosage of instruction is differentially associated with the efficacy for various groups of students. Until then, our results should not be interpreted as suggesting the futility of providing an estimate of the optimal dosage of early PA instruction for students. We can offer two conclusions on the cumulative dosage of PA instruction for classrooms/individual students based on our study. One, the optimal cumulative dosage of PA instruction for a randomly selected classroom should be regulated by how long it takes those students to acquire the PA and orthography-phonology associations they are taught. Two, one way to determine that is to pretest the students for phonology and orthography and adjust the cumulative dosage of PA instruction to suit individual and class needs.

Third, in the current study, we only examined PA instruction and its association with PA outcomes. Future studies should investigate the optimal dosage of PA instruction required to master the reading of letter-sounds, onset-rimes, and words in kindergarten. Beginning-of-the-year kindergarten students could be randomly assigned to small group PA instruction (not 1:1 due to cost) based on their letter-sound knowledge (none vs. some). Based on prior empirical research (e.g., Foorman & Francis, 1994), we might hypothesize that children with no letter-sound knowledge would require a higher optimal cumulative dosage of PA instruction to master literacy skills taught and achieve grade-level proficiency.

Fourth, this meta-analysis focused on one treatment intensity factor – cumulative dosage – and its effectiveness on PA instruction. Other factors, such as dose and quality of PA instruction, may also play a role. It may be possible that an optimal treatment effect is different at different dose levels and depends on the PA instruction quality. Future research can estimate such scenarios using RCTs. Statistically speaking, it was not possible to estimate the interaction effects among cumulative dosage, dose, PA instruction quality, or other factors. Similarly, because the number of effect sizes was somewhat limited across moderators, we were not able to statistically estimate the interaction effects among various moderators (e.g., group size by reading disabilities status or type of PA skills by use of letters).

Fifth, we need to use caution regarding the classification of effect size magnitude. Note that even though the standardized mean-change measure effect size is on the same metric as Cohen's *d*, the traditional small, medium, and large rules of thumb (0.20, 0.50, and 0.80, respectively) do not generalize well to repeated measures designs whose standardized effect sizes tend to be larger (Kraft, 2020). Additionally, we chose this effect size because it represents the difference in the standardized change in the PA instruction group to the standardized change in the control group relative to the pretest standard deviation. As such, this effect size does not represent the standardized change itself. Standardizing the pretreatment to posttreatment change in the treatment group alone would overestimate the effect size because, theoretically, the control group might also improve to some degree in response to the control instruction. The nonlinear meta-analysis provides crucial information about how effect size is related to treatment cumulative dosage alongside a maximum effect size. However, the maximum effect size may not be the perfectly optimal effect size when considered alongside treatment cumulative dosage. For example, if 50 hours of instruction results in an effect size of 0.40, 100 hours of instruction results in an effect size of 0.80, and 150 hours of instruction results in an effect size of 0.85 (the maximum effect size), the additional 50 hours of instruction beyond 100 hours is likely not cost-effective, as it comes with a relatively small increase in PA outcomes effectiveness. For these reasons, we avoid classifying effect sizes according to traditional rules of thumb and instead use Kraft's (2020) guidelines related to the cost-effectiveness of educational interventions. Under these guidelines, a small-effect, low-cost intervention may be considered as valuable as a large-effect, high-cost intervention.

Conclusion

Given that using linear models in the presence of nonlinear relations between variables may lead to misleading findings (Bauer & Cai, 2009), prior literature has cautioned against using them in such scenarios. In this paper, we present a nonlinear meta-analysis, which can serve as one example of the potential benefits of analyzing reading research questions in a nonlinear fashion. Specifically, we used a nonlinear meta-analysis to identify the optimal cumulative dosage of early PA instruction on PA outcomes with an associated maximum effect size. The present study identified that the optimal cumulative dosage of early PA instruction was approximately 10.20 hours. This finding also held for students at risk for reading disabilities. Together, these findings highlight the importance of planning the optimal cumulative dosage of PA instruction in preschool through first-grade settings so that students acquire the PA and phonological-orthographic associations taught and show progress in learning to read.

Acknowledgments

The authors would like to thank Mary Rose Sallese for her assistance with study coding.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported, in part, by a grant from the U.S. Department of Education, Office of Special Education Programs (H325D210073). Views expressed herein are those of the authors and have neither been reviewed nor approved by the granting agencies.

ORCID

Florina Erbeli http://orcid.org/0000-0002-4955-3170
Marianne Rice http://orcid.org/0000-0001-8935-4734
Megan E. Bishop http://orcid.org/0000-0002-2862-8479
J. Marc Goodrich http://orcid.org/0000-0003-1072-8305

Data Availability Statement

Data can be requested from the first author.

References

*Primary studies meta-analyzed in the current report are listed in the Supplementary Materials.

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of literature. *Remedial and Special Education*, *23*(5), 300–316. https://doi.org/10.1177/07419325020230050501

Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough? *Exceptionality*, *13*(4), 195–208. https://doi.org/10.1207/s15327035ex1304_2

Anthony, J. L. (2016). For Which Children of Economic Disadvantage and in Which instructional contexts does earobics step 1 improve kindergarteners' literacy?. *Journal of Research on Educational Effectiveness*, *9*(1), 54–76. https://doi.org/10.1080/19345747.2015.1055637

Barker, T. A., & Torgesen, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research*, *13*(1), 89–103. https://doi.org/10.2190/TH3M-BTP7-JEJ5-JFNJ

Bauer, D. J., & Cai, L. (2009). Consequences of unmodeled nonlinear effects in multilevel models. *Journal of Educational and Behavioral Statistics*, *34*(1), 97–114. https://doi.org/10.3102/1076998607310504

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Brown, G. D. A., & Deavers, R. P. (1999). Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology*, *73*(3), 208–242. https://doi.org/10.1006/jecp.1999.2502

Bus, A., & van Ijzendoorn, M. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, *91*(3), 403–414. https://doi.org/10.1037/0022-0663.91.3.403

Clemens, N. H., Solari, E., Kearns, D. M., Fien, H., Nelson, N. J., Stelega, M., Burns, M., St. Martin, K., & Hoeft, F. (2021). They say you can do phonemic awareness instruction "in the dark," but should you? A critical evaluation of the trend toward advanced phonemic awareness training. https://doi.org/10.31234/osf.io/ajxbv

Cudeck, R., & du Toit, S. H. C. (2002). A version of quadratic regression with interpretable parameters. *Multivariate Behavioral Research*, *37*(4), 501–519. https://doi.org/10.1207/S15327906MBR3704_04

Denton, C. A., Cirino, P. T., Barth, A. E., Romain, M., Vaughn, S., Wexler, J., & Fletcher, J. M. (2011). An experimental study of scheduling and cumulative dosage of "Tier 2" first-grade reading intervention. *Journal of Research on Educational Effectiveness*, *4*(3), 208–230. https://doi.org/10.1080/19345747.2010.530127

Dumas, D. G., & McNeish, D. M. (2017). Dynamic measurement modeling: Using nonlinear growth models to estimate student learning capacity. *Educational Researcher*, *46*(6), 284–292. https://doi.org/10.3102/0013189X17725747

Dumas, D. G., & McNeish, D. M. (2018). Increasing the consequential validity of reading assessment using dynamic measurement modeling. *Educational Researcher*, *47*(9), 612–614. https://doi.org/10.3102/0013189X18797621

Ehri, L. (2005). Learning to read: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the national reading panel's meta-analysis. *Reading Research Quarterly*, *36*(3), 250–287. https://doi.org/10.1598/RRQ.36.3.2

Ehri, L. C., & Roberts, T. (2006). The roots of learning to read and write: Acquisition of letters and phonemic awareness. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 113–131). Guilford Press.

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing*, *7*(3), 295–326. https://doi.org/10.1007/BF03162082

Erbeli, F., Hart, S. A., & Taylor, J. (2018). Longitudinal associations among reading-related skills and reading comprehension: A twin study. *Child Development*, *89*(6), e480–e493. https://doi.org/10.1111/cdev.12853

Erbeli, F., He, K., Cheek, C., Rice, M., & Qian, X. (2023). Exploring the machine learning paradigm in determining risk for reading disability. *Scientific Studies of Reading*, *27*(1), 5–20. https://doi.org/10.1080/10888438.2022.2115914

Erbeli, F., & Rice, M. (2022). Examining the effects of silent independent reading on reading outcomes: A narrative synthesis review from 2000 to 2020. *Reading & Writing Quarterly*, *38*(3), 253–271. https://doi.org/10.1080/10573569.2021.1944830

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review*, *40*(1), 3–22. https://doi.org/10.1080/02796015.2011.12087725

Foorman, B. R. (2003). The necessity of the alphabetic principle to phonemic awareness instruction. *Reading and Writing*, *16*(4), 289–324. https://doi.org/10.1023/A:1023671702188

Foorman, B. R., & Francis, D. J. (1994). Exploring connections among reading, spelling, and phonemic segmentation during first grade. *Reading and Writing*, *6*(1), 65–91. https://doi.org/10.1007/BF01027278

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, *90*(1), 37–55. https://doi.org/10.1037/0022-0663.90.1.37

Foorman, B. R., Francis, D. J., Novy, D. M., & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology*, *83*(4), 456–469. https://doi.org/10.1037/0022-0663.83.4.456

Fuchs, D., & Fuchs, L. S. (2015). Rethinking service delivery for students with significant learning problems: Developing and implementing intensive instruction. *Remedial and Special Education*, *36*(2), 105–111. https://doi.org/10.1177/0741932514558337

Gale, D. (2006). *The effect of computer-delivered phonological awareness training on the early literacy skills of students identified as at-risk for reading failure*. [Doctoral dissertation, University of South Florida]. University of South Florida website. http://purl.fcla.edu/usf/dc/et/SFE0001531.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491–528. https://doi.org/10.1037/0033-295X.106.3.491

Hatcher, P. J., Hulme, C., Miles, J. N. V., Carroll, J. M., Hatcher, J., Gibbs, S., Smith, G., Bowyer-Crane, C., & Snowling, M. J. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *47*(8), 820–827. https://doi.org/10.1111/j.1469-7610.2005.01559.x

Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry*, *45*(2), 338–358. https://doi.org/10.1111/j.1469-7610.2004.00225.x

Hecht, S. A., & Close, L. (2002). Emergent literacy skills and training time uniquely predict variability in responses to phonemic awareness training in disadvantaged kindergartners. *Journal of Experimental Child Psychology*, *82*(2), 93–115. https://doi.org/10.1016/S0022-0965(02)00001-2

Hudson, R. F., Sanders, E. A., Greenway, R., Xie, S., Smith, M., Gasamis, C., Martini, J., Schwartz, I., & Hackett, J. (2017). Effects of emergent literacy interventions for preschoolers with autism spectrum disorder. *Exceptional children*, *84*(1), 55–75. https://doi.org/10.1177/0014402917705855

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science*, *23*(6), 572–577. https://doi.org/10.1177/0956797611435921

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on castles and coltheart. *Scientific Studies of Reading*, *9*(4), 351–365. https://doi.org/10.1207/s1532799xssr0904_2

Hund-Reid, C. (2008). *The effectiveness of phonological awareness intervention for kindergarten children with moderate to severe language impairment*. [Doctoral dissertation, University of Alberta]. Library and Archives Canada. https://central.bac-lac.gc.ca/.item?id=NR45449&op=pdf&app=Library&oclc_number=682256453

Kelly, C., Leitão, S., Smith-Lock, K., & Heritage, B. (2019). The effectiveness of a classroom-based phonological awareness program for 4–5-year-olds. *International Journal of Speech-Language Pathology*, *21*(1), 101–113. https://doi.org/10.1080/17549507.2017.1400589

Kim, Y. S., Petscher, Y., Foorman, B. R., & Zhou, C. (2010). The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition—a cross-classified multilevel model approach. *Journal of Educational Psychology*, *102*(2), 313–326. https://doi.org/10.1037/a0018449

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG rime and GG phoneme. *Reading Research Quarterly*, *48*(1), 61–76. https://doi.org/10.1002/rrq.038

Lonigan, C. J., Driscoll, K., Phillips, B. M., Cantor, B. G., Anthony, J. L., & Goldstein, H. (2003). A computer-assisted instruction phonological sensitivity program for preschool children At-risk for reading problems. *Journal of Early Intervention*, *25*(4), 248–262. https://doi.org/10.1177/105381510302500402

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Mitchell, M. J., & Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction*, *40*(4), 315–332. https://doi.org/10.1080/19388070109558353

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, Rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*(5), 665–681. https://doi.org/10.1037/0012-1649.40.5.665

National Reading Panel. (2000) . *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development.

Nelson, R. J., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research and Practice*, *18*(4), 255–267. https://doi.org/10.1111/1540-5826.00080

Nelson, J. R., Sanders, E. A., & Gonzalez, J. (2009). The efficacy of supplemental early literacy instruction by community-based tutors for preschoolers enrolled in head start. *Journal of Research on Educational Effectiveness*, *3*(1), 1–25. https://doi.org/10.1080/19345740903381031

Nutkins, C. L. (2004). *Examining the Effectiveness of Phonemic Awareness Training for At-Risk Preschool Children: Does Parental Involvement Promote Skill Development?* () [Doctoral dissertation, Central Michigan University]. ProQuest Dissertations and Theses Global.

Ogungbenro, K., Dokoumetzidis, A., & Aarons, L. (2009). Application of optimal design methodologies in clinical pharmacology experiments. *Pharmaceutical Statistics*, *8*(3), 239–252. https://doi.org/10.1002/pst.354

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Lawrence Erlbaum Associates.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*(2), 31–74. https://doi.org/10.1111/1529-1006.00004

Rehmann, R. (2005). *The effect of Earobics Step 1, software on student acquisition of phonological awareness skills.* [Doctoral dissertation, University of Oregon]. ProQuest Dissertations and Theses Global.

Rice, M., Erbeli, F., Thompson, C. G., Sallese, M. R., & Fogarty, M. (2022). Phonemic awareness: A meta-analysis for planning effective instruction. *Reading Research Quarterly*, *57*(4), 1259–1289. https://doi.org/10.1002/rrq.473

Rice, M., Erbeli, F., & Wijekumar, K. (2023). Phonemic awareness: Evidence-based instruction for students in need of intervention. *Intervention in School and Clinic*, 105345122311568. https://doi.org/10.1177/10534512231156881

Roberts, G. J., Dumas, D. G., McNeish, D., & Coté, B. (2022). Understanding the dynamics of dosage response: A nonlinear meta-analysis of recent reading interventions. *Review of Educational Research*, *92*(2), 209–248. https://doi.org/10.3102/00346543211051423

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*(2), 265–282. https://doi.org/10.1037/0022-0663.96.2.265

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology*, *91*(3), 439–449.

Schuele, C. M., & Boudreau, D. (2008). Phonological awareness intervention: Beyond the basics. *Language, Speech, and Hearing Services in Schools*, *39*(1), 3–20. https://doi.org/10.1044/0161-1461(2008/002)

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

Timmons, A. C., & Preacher, K. J. (2015). The importance of temporal design: How do measurement intervals affect the accuracy and efficiency of parameter estimates in longitudinal research? *Multivariate Behavioral Research*, *50*(1), 41–55. https://doi.org/10.1080/00273171.2014.961056

Travis, P. C. (1997). *Effects of computer-assisted and teacher-led phonological awareness instruction for first-grade students at risk for reading failure*. [Doctoral dissertation, University of Florida]. ProQuest Dissertations and Theses Global.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*(2), 107–136. https://doi.org/10.1037/0096-3445.124.2.107

Ukrainetz, T. A., Ross, C. L., & Harm, H. M. (2009). An investigation of treatment scheduling for phonemic awareness with kindergartners who are at risk for reading difficulties. *Language, Speech, and Hearing Services in Schools*, *40*(1), 86–100. https://doi.org/10.1044/0161-1461(2008/07-0077)

Vaughn, S., Denton, C. A., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools*, *47*(5), 432–444. https://doi.org/10.1002/pits.20481

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryant, D. P., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education*, *24*(5), 301–315. https://doi.org/10.1177/07419325030240050501

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology*, *30*(1), 73–87. https://doi.org/10.1037/0012-1649.30.1.73

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, *33*(3), 468–479. https://doi.org/10.1037/0012-1649.33.3.468

Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities*, *41*(2), 126–142. https://doi.org/10.1177/0022219407313426

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review*, *28*(3), 551–576. https://doi.org/10.1007/s10648-015-9321-7

Warren, P. (2009). *The Effects of Training Parents in Teaching Phonemic Awareness on the Phonemic Awareness and Early Reading of Struggling Readers.* [Doctoral dissertation, Auburn University]. ProQuest Dissertations and Theses Global.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*(1), 70–77. https://doi.org/10.1002/mrdd.20139

Webb, M. Y. L., Schwanenflugel, P. J., & Kim, S. H. (2004). A construct validation study of phonological awareness for children entering prekindergarten. *Journal of Psychoeducational Assessment*, *22*(4), 304–319. https://doi.org/10.1177/073428290402200402

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