This is a post-peer-review, pre-copyedit version of an article published in *Reading Psychology.* The final authenticated version was available online first on January 16, 2023 at https://doi.org/10.1080/02702711.2023.2166636

Cognitive flexibility + phonics intervention effects on reading gains

Patricia F. Vadasy

Oregon Research Institute

Elizabeth A. Sanders

University of Washington

Author Note

Patricia F. Vadasy, Oregon Research Institute. Elizabeth A. Sanders, University of Washington.
We wish to thank Sueanne Sluis, project manager, and the research assistant instructors for their contributions and dedication to this intervention study. This research was supported by the Institute of Education Sciences (IES), U.S. Department of Education, Grant No. R305A180005. Any opinions, findings, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Institute of Education Sciences.
Correspondence should be addressed to Patricia Vadasy, Oregon Research Institute, 3800 Sports Way, Springfield, OR 97477, USA. patriciav@ori.org

Abstract

This is the second in series of studies designed to test direct and conditional effects of embedded cognitive practice in phonics instruction. Students identified in winter of kindergarten with minimal alphabet knowledge were randomly assigned to one of two conditions: explicit phonics (Plain) (n = 28) or explicit phonics with embedded cognitive flexibility practice (Flex) (n = 29). The core of both conditions was an explicit structured literacy approach: the Flex condition was differentiated by brief cognitive flexibility practice switching letter or word dimensions. Instruction was delivered individually over a six-week period. In spite of Covid-19 impacts, both treatment groups exhibited significant gains on reading outcomes. However, there were no significant differences between the conditions on growth in decoding, encoding, or cognitive flexibility. Future research should consider the timing and design of instruction to determine how cognitive abilities, as well as alphabet knowledge, contribute to acquisition of early reading skills.

Cognitive Flexibility + Phonics Intervention Effects on Reading Gains

The most robust kindergarten predictor of later reading achievement is letter name knowledge, and a composite measure of letter name and letter sound knowledge at preschool corroborates this relationship (see Foulin, 2005). Letter sound knowledge also predicts decoding development (Schaars et al., 2017). The period during which children develop accurate knowledge of grapheme-phoneme correspondences (GPCs) and draw upon this knowledge to practice decoding is extended for many at-risk children. Before decoding becomes automatic, it draws heavily on executive functions (EF)—cognitive resources to control attention to the print, recall the relevant phonological codes, ignore irrelevant codes, and flexibly shift attention from one grapheme and sound to the next (van de Sande et al., 2017). Decoding also requires children to flexibly apply these skills to recognize the blended word. This is often difficult because phonemes become less distinct when they are coarticulated (Liberman et al., 1967). This is a slower and more effortful task for some children. Longitudinal findings indicate that children are typically learning these decoding skills while the cognitive abilities involved in this task are still developing (Yeniad et al., 2014).

Structured literacy approaches, featuring explicit, systematic, and sequenced instruction enable many at-risk children to overcome the most challenging obstacle to reading success, word identification (Spear-Swerling, 2018). Yet, a troubling proportion of at-risk beginning readers fail to respond to evidence-based phonics interventions and continue to struggle to learn the to decode (Al Otaiba & Fuchs, 2006; Torgesen, 2000). In our earlier research on supplemental kindergarten phonics instruction for students identified at risk for reading difficulties, between 8 to 36 percent of students remained below the 30th percentile in decoding at posttest (Vadasy, Sanders, & Peyton, 2006), and others have reported on children who fail to respond to explicit phonics interventions (Al Otaiba & Fuchs, 2006).

Two areas of research on the acquisition of early reading skills suggest features of early reading instruction that may increase responsiveness for beginning at-risk readers. One line of research on teaching alphabet knowledge suggests re-examining the content of early phonics instruction. Many programs introduce children to the letters of the alphabet, often in order, often one letter a week (Pressley et al., 1996). Single-letter grapheme-phoneme-correspondences (GPCs) are typically taught before introducing two-letter GPCs, some of which appear more frequently in grade-level texts than single letters (e.g., *sh* versus *j* and *y*), and these variations in GPCs characterize the English language (Ziegler & Goswami, 2005, 2006). Primary grade students introduced to multi-letter grapheme-phoneme correspondences (mixed grain size phonics content) are able to read words accurately and fluently (Blachman et al., 2004; Conrad & Levy, 2011), which benefits spelling (Wright & Ehri, 2007) and comprehension (Christensen & Bowey, 2005). A close examination of alphabet instruction recommends a teaching rate of one letter a day accompanied with judicious review and discrimination learning (Jones & Reutzel, 2012). In a series of experiments we conducted (summarized below), of which the present study represents the last, we examined how the types of GPCs introduced, and the rate at which they are introduced, may influence initial phonics learning.

A second line of research on the acquisition of alphabet and decoding skills highlights the role of basic cognitive processes, such as phonemic awareness and rapid naming skills, involved in young children's early reading development (Compton, 2000). One group of cognitive abilities in particular, described under the umbrella of executive function (EF) plays an important role in the development of young children's reading and math skills (Bierman et al., 2008; Bull et al., 2008; Haft et al., 2019; Swanson & O'Connor, 2009; Yeniad et al., 2013). Executive function (EF) comprises three primary and related cognitive processes: working memory, inhibitory control, and cognitive flexibility. These cognitive abilities, which are still developing during the early school years, are each positively related to academic performance (Best et al., 2011; Jacob & Parkinson, 2015).

The role of cognitive flexibility is most well documented in studies of reading comprehension, which requires the reader to simultaneously process semantic and phonological information. Most of this research included school-age children and adults (Cartwright, 2002; Jacobson et al., 2016; Kieffer et al., 2013; Sondergaard Knudsen et al., 2018). However, research indicates that cognitive processes also play a role in early reading development. In a study of both typically developing elementary school children and children with language impairments (M_{age} 9.75 years) Messer et al. (2016) found that several verbal EF tasks significantly predicted decoding: verbal fluency, inhibition, and planning. Ober et al. (2020) reviewed the research on associations between EF components and decoding in children and adolescents. In their review of 65 studies, task switching was significantly associated with word and nonword reading, with moderate effect sizes. Research also suggests that training these cognitive abilities may support young children's learning of early reading skills (Cartwright et al., 2020; Dias & Seabra, 2016; Goodrich et al., 2021; Peng et al., 2019).

In this study, we focus specifically on cognitive flexibility, which is the "ability to consider multiple bits of information or ideas at one time and activity switch between them when engaging in a task" (Cartwright, 2012, p. 26). Cognitive flexibility is also sometimes referred to as "switching" (Davidson et al., 2006). Readers must coordinate the semantic and word-level features of print, and this flexibility significantly contributes to reading comprehension in school-age children (Cartwright, 2002, 2006; Cartwright et al., 2017).

Beginning readers must also develop the cognitive flexibility to attend to the phonological, printed, and semantic features of words, and more advanced word reading requires flexibility in processing multiletter units and generating alternative pronunciations (Gaskins, 2008). Others have found that EF skills play a role in early reading development, for preschoolers' learning emergent literacy skills (Bierman et al., 2008), and for young children learning decoding and single-word reading (Cartwright, 2002; Cole et al., 2014; van de Sande et al., 2018). Van de Sande et al. (2017) examined the influence of attentional and action control on early reading skills in children kindergarten through second grade. These EF component skills were most advantageous in kindergarten when children are developing decoding skills. Attention and action control have been found to mediate reading development in children kindergarten to Grade 2 (van de Sande et al., 2017). Haft et al. (2019) found that EF significantly indirectly influenced reading comprehension through word decoding in kindergarten children. There are challenges in measuring these cognitive abilities in young children when individual EF components are less differentiated.

The present research represents the last in a sequence of three experiments with small samples to investigate malleable cognitive elements that may improve early phonics instruction for kindergarten children who respond inadequately to standard systematic phonics approaches. In the first experiment, we compared phonics elements and rates for teaching alphabetic and reading skills, and identified a rate for introducing a set of mixed size (one and two-letter) GPCs associated with significantly greater than average reading gains on mixed items (Vadasy & Sanders, 2021a). In the second experiment (Vadasy & Sanders, 2021b), we randomized younger children with low alphabetic knowledge in the fall of kindergarten to one of two six-week one-to-one training conditions that took place during the child's school day: one that featured basic explicit and effective phonics instruction ("Plain") and one that also included embedded cognitive practice ("Flex"). Controlling for pretest skills and treatment dosage (attendance), those results showed that children in the basic systematic phonics condition made significantly greater gains than Flex students on encoding outcomes: writing taught letter-sound correspondences and spelling. There were no other significant differences between conditions. A question remained whether children in the fall of kindergarten did not yet have a level of cognitive maturity to benefit from the embedded cognitive flexibility practice.

The Present Study

In the present study, we replicate and extend the previous study with a slightly older cohort of at-risk kindergarten children poised at the initial stage of decoding in the winter of their school year – after four months of formal classroom kindergarten learning. This was in-person, pre-pandemic; participating children were from the same school as those from the fall who had either been randomly selected to wait to be part of the winter cohort, or who had moved to the school after the fall study had begun. Taken together, the experiments examine components, timing, and intensity of initial phonics instruction to support children who struggle with initial decoding acquisition.

Specifically, the current study's research questions were: 1) Does brief phonics instruction with embedded cognitive practice improve older kindergarteners' cognitive and early reading skills growth, controlling for Emergent Bilingual status, initial skills, and treatment dosage, compared with phonics instruction alone? 2) Are treatment differences, if any, moderated by our covariates? In other words, is one treatment more or less beneficial for certain students?

Methods

Participants

Participants were a second cohort of students from one school in the Pacific Northwest, U.S. Students had low English alphabet knowledge (fewer than half of letter names or sounds) in December of their kindergarten year, averaging *M* = 9.11 letter names (*SD* = 2.85, *range* = 0 – 13) and *M* = 2.49 letter sounds (*SD* = 2.26, *range* = 0 – 8). The initial set of participants screened in December for alphabet knowledge included a mix of: 1) students previously identified by the school in September with low alphabetic knowledge but who were *randomly selected* to not take part in a fall study (Vadasy & Sanders, 2021b) due to time and budget constraints, and 2) students new to the school by December and whose teachers identified them as having low alphabetic knowledge. In other words, none had participated in a previous study from the fall of the academic year. For this pool of students, the school sent home a parent information letter (translated into three major languages), and teachers informed us of any parents who chose to opt out of their child's participation.

Treatment assignment. The school sent home a parent information letter (translated into three major languages) for all identified students, and teachers informed us of any parents who chose to "opt out." The initial sample included *N* = 68 consented children with low alphabet skills from 21 classrooms who were randomly assigned, within classroom and emergent bilingual (EB) status, to one of two experimental conditions (Flex *n* = 34, or Plain instruction, *n* = 34). EB status was determined by whether a language other than English was spoken at home (1 = yes, 0 = otherwise), and that random assignment was carried out by sorting students by classroom number identifier, EB status, and random number (uniform distribution), and then assigning treatment condition to each student in an alternate fashion. This stratified random assignment ensured proportional representation of classrooms and EB students in each group.

Covid-19 impact. Like others, the Covid-19 March 2020 school closures affected our study in two primary ways: some students missed some or all of week 6 of the tutoring intervention, and some students missed some or all of their posttesting. When we gained insight that the pandemic had begun (before schools actually closed), we stopped all remaining tutoring and tried to finish posttesting quickly (students were on slightly staggered schedules so some had already completed the intervention period). Even with our modified schedule, we were unable to test nine students (four Flex and five Plain), and for those we were able to test, there were some who missed their assessments for the writing and cognitive flexibility measures due to the order of the testing tasks. Specifically, nine students missed the cognitive flexibility posttest (two Flex and seven Plain), four were missing letter sound writing (all Plain), and six were missing spelling (all Plain). Ultimately, we chose to analyze data for any student who had any posttest data.

Final sample. Attrition included a total of $n = 11$ children: two were removed from the study early due to behavior problems during tutoring (one from each condition), and nine missed the last week of the treatment and all of posttesting (see above) due to the Covid-19 pandemic school closures co-occurring with the end of our study. The final sample included $N = 57$ children from 21 classrooms: 29 in the Flex condition and 28 in the Plain. As a check on attrition, we compared the 11 attritted students with the 57 non-attritted students, and found no evidence for differences on any demographic or pretest variable (2-group chi-square and *t*-test *p*s $> .10$). The final sample included 17 (30%) females, 26 (47%) EB children, and 15 (26%) children of color. The mean age at pretest was 5.86 years ($SD = 0.33$). Comparison of Flex and

Plain conditions also showed no evidence of differences among groups on any demographic or pretest variable, or on attrition (2-group chi-square and *t*-test *p*s > .10). There was substantially more missingness on the writing and cognitive flexibility measures for the Plain condition compared to Flex. Nevertheless, our statistical models included all 57 students because we use full information maximum likelihood estimation, which treats outcome data as missing-at-random (which can be assumed as long as there is complete data on covariates and the missingness mechanism is included in the model). Specifically, our models control for pretest levels (for which we have complete data) and treatment attendance (which serves as an indicator of the mechanism involved in the missingness – i.e., Covid 19 pandemic disruption).

Tutors. Children's assignment to their tutor was based on school scheduling convenience. There were eight tutors; each tutor served six to nine children daily (tutors served the same set of children throughout the intervention period), with approximately half of each tutor's assigned children in each of the two conditions.

General Procedures

The intervention was delivered by a team of eight research assistant instructors with previous training and classroom experience, including former classroom teachers and paraeducators. Instructors practiced the lessons as homework prior to an all-day training session in which the first author modeled lesson delivery, correction procedures, and pacing. All instructors were observed implementing lessons, were provided feedback. During and after their homework and training, the instructors shared questions by e-mail, and answers and feedback were shared with the entire team. Training emphasized instructional delivery, including, teaching letter content and reading skills with precision, including new letter sounds, and new or difficult skills like blending or segmenting. Instructors were trained to model the sounds and skills, require the student to repeat and practice the skills, and provide scaffolding and practice for each student. When students struggled to provide a letter sound, the instructor isolated the letter, and asked the student “what sound.” If the student could not supply the sound, the instructor provided the sound, had the student say it in unison, say it alone, and then the instructor provided added practice opportunities for difficult sounds. Researchers were on site daily, met with the team for informal lunch meetings and individual coaching, and shared instructional tips in weekly e-mails. A team of six research assistant assessors had previous testing experience and included former teachers, a school psychologist, a school principal, and teacher trainer. Assessors were trained in a half-day session in which assessments were demonstrated by the researchers, and the assessors paired up to practice measures while researchers observed and shared feedback. Assessors also had to complete homework to practice administering the entire assessment package with a young child, and researchers provided added feedback to address questions. Similar homework and review were assigned prior to posttests. Research staff included a project coordinator with over 15 years’ experience overseeing school-based early reading interventions.

Materials and Procedures

Two sets of lessons, Plain and Flex, were prepared, each with similar teaching activities. The lessons included 8-10 brief activities to explicitly teach and practice letter sounds, segmenting, spelling, and blending. The rate of introducing and teaching the set of GPCs was tested in earlier studies (Vadasy & Sanders, 2021a, 2021b). Lesson activities progressed from learning the isolated GPCs, application to blending and spelling, sorting words, letter fluency practice, and reading words in short sentence contexts. The focus in Plain lessons was on varied activities to practice matching letters and sounds, in isolation and in words. Flex lessons featured variations in four tasks that required students to practice flexible thinking and inhibitory control in the phonics activities. Three GPC learning activities required Flex students to switch retrieval of letter features by alternately identifying the letter name or letter sound; find a target letter in varied word positions (first, last, middle); and quickly identify letter names or sounds in a printed array of taught letters. Beginning readers with little or no early literacy knowledge may struggle with stimuli like letters that have two labels (names and sounds). They may initially confuse these labels, utilizing names rather than sounds to decode (Van deSande et al., 2017). Finally, Flex students also practiced sorting words not only by initial sound (as in Plain lessons) but by several phonological and meaning dimensions (last sound, semantic category). The sorting practice was similar to traditional sorting tasks that require children to sort picture cards according to two dimensions (e.g., color and kind).

Lessons were scripted, and six weeks of individual instruction was provided in 20-minute sessions delivered four days a week. Instruction occurred outside the classroom, in areas designated for tutoring or small group instruction, or in small rooms or spaces outside the classroom. At the end of every two weeks of instruction, an 8-item decoding test was administered. The student was shown a set of 8 words printed on 5 x 7 cards. For each probe the words were decodable based on the phonics content taught to that point. The probes were scored for total words and total phonemes correct. The probe data for each student were shared with the instructors with suggestions to support student progress.

Treatment Conditions

Children were taught the same 13 GPCs in both conditions (a, s, t, oo, c, m, b, i, o, ee, p, sh, d). The number of correspondences introduced each week were: 4, 2, 2, 2, 2, 1. Lesson pages displayed tutor directions in a column on the right side of a page, and the letter and word item stimuli for student practice in large print on the left side of the page. Other materials used were grid paper, letter tiles for phoneme-grapheme spelling practice, laminated segmenting boxes, and index cards. Children in both conditions practiced immediately applying knowledge of taught correspondences to decoding and spelling tasks. The ten basic teaching activities are described, with Flex variations described for four activities.

Letter sound learning. The instructor introduced and modeled new letter sounds, and the student repeated each sound. For the first row the student pointed to each letter and said the name and sound, then the student then pointed to each letter in an array of new and previously taught letters in 5-6 rows of 5 letters and said each sound. Letters that were newly introduced were always featured 4 times in the array, with cumulative review of previously taught letters. The instructor then dictated the new sounds for the student to write. In the Flex lessons, using the same array of letters, the instruction alternated for each row asking the student to point to each letter and say either the name or the sound.

Blending. For the first three weeks of instruction, the instructor demonstrated phoneme blending with four words composed of taught letters. After pointing to each letter and blending the sounds the instructor asked the student to point, blend the sounds, and say the word fast. Then the student demonstrated the task. The student then identified the first and last sounds in each word.

Read letters in different word positions. Referring to a set of nine words with taught letters, the instructor pointed to a letter in a word and asked the student to say the sound, and then read the word. The instructor varied pointing to letters in each word position, and practice included 3-5 rows of words. In the Flex lessons, the instructor dictated a letter sound (without pointing to the letter) and asked the student to find that letter in the set of words, point, and say the sound and read the word. The student practiced recognizing the letters in the contexts of the same 3-5 rows of words used in the Plain lessons. The Flex activity required the student to discriminate a larger set of words and letters to match the sound to the letter.

Word sort. Using a set of eight word cards, the instructor asked the student to read the words, sort the words by initial sound, and then say the first sound and read the words in each pile. In the Flex lessons, the student sorted the same word cards used in the Plain lessons by initial and last sound, and by semantic category (animal words, other categories). To scaffold students with limited vocabulary and English proficiency for the semantic sort, the instructor pointed out the word meaning after the first reading of the word. For example, if the semantic sort asked the student to put the words that were a boy's name in one pile, when the words were first read the tutor would point out that "*Sam* is a boy's name. *Tim* is a boy's name."

Segmenting. Displaying letter tiles for taught words, the instructor dictated 5-8 words. For each word, the student segmented the phonemes, pointing to each letter, saying the sound, and reading the word.

Spelling. The instructor dictated 4-8 words, the student repeated the word, and then the student spelled the word while placing the letter tiles in the segmenting boxes, and read the word.

Reading words in context. Using two sentences constructed with mostly taught letters, the tutor read the sentence aloud with the student following. Then the student and instructor pointed and read each sentence together, and the student read the sentence with support as needed.

Quick find. Using an array of taught letters arranged randomly in 5-8 rows, the instructor dictated the new letter sounds and the student pointed to the letters that match the sounds. Then the student pointed to each letter in each row saying each sound. In the Flex lessons, using the same letter grid used in the Plain lessons, the student pointed to each letter and alternated saying either the name or the sound for each letter in the row.

Find the sound. Using letter tiles and segmenting boxes, the instructor set out one letter tile at a time and prompted the student to point to the box that matched the sound in a word. For each taught sound, the student practiced with two words. For example, "Point to the box where you hear the sound /c/ in mac. In *cat*."

Find the word. Using two sentences that featured a word with a recently taught letter, the instructor helped the student select the missing word from three or four choices. For example, "This sentence is missing a word (tutor reads the sentence with the correct word that goes in the blank). Mac has a furry pet ____ (cat)" Pick the word that goes in the blank." The student must discriminate among: *tat, cat, sat, cot*. The student read the correct word, spelled the correct word, and the instructor and tutor read the sentence together with the correct word. The student then wrote the word on grid paper.

Treatment Fidelity and Attendance

Treatment fidelity. Each of the eight tutors was observed for each treatment approximately 20 times (Flex $M = 10.63$, $SD = 2.45$; Plain $M = 9.00$, $SD = 2.00$), and all children were formally observed by research staff at least once with their tutor (range of 1 to 5 per child), for a total of 157 observations conducted. Observations involved 20 yes/no items on a checklist, 15 for treatment content and 5 pertaining to general instructional behaviors. Across observations, Flex content fidelity averaged $M = 99\%$ ($SD = 1\%$) and Plain averaged $M = {>}99\%$ ($SD = 1\%$). General instructional behavior averaged $M = 99\%$ ($SD = 2\%$) for Flex and $M = 98\%$ ($SD = 3\%$) for Plain. Paired *t*-tests showed no significant differences between treatments on observation counts, content fidelity, or instructional behaviors (ps > .10).

Treatment dosage (attendance). Tutors recorded weekly attendance for each child (maximum of 4 days per week for 6 weeks = 24 days); children's attendance ranged from 13 to 23 days, with the Flex condition averaging 21.24 days ($SD = 1.77$), and the Plain condition averaging 20.29 days ($SD = 2.77$). Due to Covid-19, $n = 19$ students missed part of their final week of tutoring: (9 Flex and 10 Plain), with $n = 2$ missing all of week 6 (one in each condition). Although there were no significant differences between conditions on total attendance, or any individual week of attendance (2-group *t*-test *p*s > .10), we control for the variation in treatment attendance in our statistical models (discussed in Analysis Plan).

Measures

Students were individually assessed during the school day by trained testers unaware of treatment assignment. Testing took place in a quiet space in the school, over two sessions to minimize student fatigue. As mentioned earlier, due to Covid-19 co-occurring with the end of the study coupled with the order of the test measures, we were missing some posttest data on three measures: taught letter sound writing, taught word spelling, and cognitive flexibility. Nevertheless, we retained all available data for analyses. Each measure is described below, including sample-based reliability estimates (Cronbach's alpha).

Receptive vocabulary. Receptive vocabulary was measured at pretest only for the purpose of describing the sample using the *Peabody Picture Vocabulary Test* (PPVT-4) (Dunn & Dunn, 2006). Reliability reported in the test manual is .97 for 5-year-olds; our sample internal consistency was estimated at .98.

Experimenter measures. For each of the following five experimenter measures, all letters taught in both groups were tested at pretest and posttest. The order in which the taught letter and word items (featuring taught letters) appeared in each test was randomized. Letter and word items matched the taught letter content. For each set of items, percent correct was calculated at pretest and posttest.

Taught letter names and sounds. For each test the tester presented a printed sheet of 13 taught letters. Letters were arrayed in four rows of three to four items per row. Students first completed two practice items with untaught letters. The tester directed the student to point and say the name (or sound) for each item, with 5 secs allowed for each item. If the student said the sound for the name (or vice versa), the tester prompted the student, "Yes, that's the sound, what is the name." The tester recorded 1 or 0 for each response, with a maximum score of 13. For letter names, sample internal consistencies were .78 and .91 at pretest and posttest, respectively; for letter sounds, internal consistencies were .72 at pretest and .78 at posttest.

Taught letter sound writing. For each of the taught letters, the tester dictated the taught sound for the student to write. The tester reminded the student that sometimes one letter makes the sound, sometimes two letters make the sound. Two practice items with untaught letters (z and wh) were first administered. The tester dictated each sound, and repeated the sound once, allowing 5 sec for each letter. If the student wrote only one letter of a two-letter sound, the tester prompted "This is a two-letter sound, write both letters that make this sound." If the student wrote two letters for a one- letter sound, the tester prompted "This is a one letter sound." The tester recorded 1 or 0 for each response, with a maximum score of 13. Sample reliabilities were estimated at .84 and .82 at pretest and posttest, respectively. As noted earlier, we were missing four students' data at posttest.

Word reading. Students were asked to read 16 cvc words constructed with taught letters that appeared in initial, medial, or final word positions. The tester first administered two practice items, demonstrating pointing to the word, blending, and reading the word fast as students learned to do in the intervention. The words were presented on a sheet in two columns and the tester directed the student to point to each word and read down each column, allowing 5 sec per word. If the student correctly said each of the sounds within 5 sec but did not blend the sounds, the tester prompted once to "Say it fast" and allowed 5 sec for the student to read the word. The tester recorded 1 or 0 for each response, with a maximum score of 16. Sample reliabilities were estimated at .73 at pretest and .92 at posttest.

Spelling. A set of 16 cvc words were used to test spelling with taught letters that appeared in initial, medial, and final word positions. The tester dictated each word, and repeated the word once upon request, allowing 5 sec per word. The tester recorded 1 or 0 for each response, with a maximum score of 16. Sample reliabilities were .87 and .94 at pretest and posttest, respectively. As noted earlier, we were missing six students' data at posttest.

Color card sort (cognitive flexibility). Cartwright's (2002) original sorting task and directions were adapted in this study for younger children with limited language skills to include three sets of picture cards used for one demonstration sort and two test sorts. Materials for each sort were a set of 12 picture cards. For the demonstration sort, the tester named each picture, had the child repeat the word, and showed the child how to make two piles of fruits and flowers, and then after shuffling the cards, how to make two piles of yellow and red things. Then the tester showed how to sort the cards by two dimensions on the 2 X 2 matrix, explained the sort, and asked the child to try to sort on the grid. If the child was unable to sort correctly, the tester demonstrated and explained the correct sort. For the two test sorts, the tester handed the child the set of shuffled cards, and began timing when the child looked at the first card. The tester named each picture as the child moved the card, and stopped timing when the child placed the last card. The tester recorded whether the sort was correct, the type of errors, and the total sort time. For a correct sort the tester asked the child to explain the sort, and recorded if the explanation was correct, and the type of errors. The total score was 1 point for a correct sort and 2 points for a correct explanation; across the two tasks, a total of 6 points was possible. The Pearson correlation between set 1 (maximum of 3 points) and set 2 (maximum of 3 points) was .52 (.44 using Kendall's tau) at pretest, and .75 at posttest (.65 using Kendall's tau). For analyses, we computed total percent correct. As noted earlier, we were missing nine students' data at posttest.

Analysis Plan

Pretest-posttest gains as outcomes. We modeled pretest-posttest gains, rather than posttests, because we were interested in overall child growth across treatments, as well as to ensure comparability of findings with our prior study of younger kindergartners. We note that modeling gain scores yields the same predictor effect results as modeling posttests as long as pretests are also included in the model, analogous to modeling residualized change (Petscher, & Schatschneider, 2011).

Data structure. We used two-level hierarchical regression models with children (Level 1, $N = 57$) nested in classrooms (Level 2, $N = 21$) to evaluate our research questions. Classrooms rather than tutors were treated as Level 2 because: (a) classroom teachers were expected to vary in their early literacy instruction and were therefore likely to have substantial effects on students' scores; and (b) tutors used scripted lessons and were observed to have very high fidelity, and were therefore unlikely to have large effects on student outcomes compared to classroom effects (we also note that there was insufficient data for cross-classified modeling). As an empirical check, we computed intraclass correlations using intercept-only 2-level linear models and found that classroom membership explained an average of 8.9% of the variance in student pretests, 5.9% in posttests, and 1.1% in pretest-posttest gains; comparatively, tutor membership explained an average of 0.3% in student pretests, 3.9% in posttests, and 1.9% in gains.

Predictor selection and coding. Similar to our prior study, we modeled student pretest-posttest gains as a function of treatment condition, Emergent Bilingual (EB) status, pretest, and treatment attendance. EB status was included because our randomization was stratified on EB status to ensure balanced representation in the two conditions. Each measure's corresponding pretest was included both to control for baseline levels in estimating treatment effects (in the prior study we used receptive vocabulary as a proxy for pretest literacy due to pretest floor effects with the younger children), as well as to provide sufficient covariate information for including all students in estimating model parameters for the three posttests affected by missingness. Finally, inclusion of treatment attendance served both to control for variation in the outcomes due to treatment dosage as well as to control for the source of the missingness for the three affected posttests (i.e., lower attendance being strongly related to school shutdowns due to the Covid 19 pandemic). All predictors were at the student level and treated as fixed effects. For ease of results interpretation, we effect-coded treatment condition (Flex = +1, Plain = –1) and EB status (EB = +1, otherwise = –1), and cluster-mean centered and standardized pretest and attendance (*z*-scores within classroom).

Model specification. A two-level random intercept linear model was used for all outcomes. Consistent with our study of younger kindergarteners, in addition to testing predictor main effects, we also included two-way interactions between treatment condition and the three covariates. Our general model was as follows.

$$\text{Pre-PostGain}_{ij} = \gamma_{00} + \gamma_{10}{*}\text{TreatEff} + \gamma_{20}{*}\text{EBEff} + \gamma_{30}{*}\text{ZPretestCMC} + \gamma_{40}{*}\text{ZAttendCMC}$$
$$+ \gamma_{50}{*}\text{Treat*EB} + \gamma_{60}{*}\text{Treat*Pre} + \gamma_{70}{*}\text{Treat*Attend} + U_{00} + r_{ij}$$

In the model above, the pretest-posttest gain for the i^{th} student in the j^{th} classroom on a given measure was estimated as a function of the conditional grand mean gain across all students and classrooms (γ_{00}). plus the main effects of treatment status (γ_{10}: difference between mean gain for children in the Flex treatment and the grand mean gain), EB status (γ_{20}: difference between EB children and the grand mean gain), pretest (γ_{30}: difference in gain between one standard deviation increase in pretest within classroom and the grand mean gain), and treatment attendance (γ_{40}: difference in gain between one standard deviation increase in treatment attendance within classroom and the grand mean gain), plus the effects of the joint effects of treatment and the three covariates ($\gamma_{50} - \gamma_{70}$), plus the random effects, the deviation between the student's classroom mean and the predicted grand mean (U_{00}) and the residual error between the student's score and their classroom's mean (r_{ij}). All models were implemented using *R* lme4 (Bates et al., 2015) with full information maximum likelihood estimation, and Satterthwaite *df* using *R* lmerTest (Kuznetsova et al., 2017).

Effect size computation. Effect sizes for all fixed effects were calculated as approximate Cohen's d = coefficient ÷ approximate SD_{pooled}, with the approximate $SD_{pooled} = SE$*sqrt(N) and where SE = model-estimated standard error and N = sample size for the measure.

Results

Descriptive statistics for each treatment condition (unadjusted for classroom membership) are shown in Table 1; zero-order correlations among variables included in analyses (also unadjusted for classroom membership) are provided in Table 2.

Mean pretest-posttest change. Table 3 presents the fixed effects results of our multilevel models for each outcome. For all measures, the intercept, which provides an estimate of the grand mean pretest-posttest change, controlling for all fixed and random effects, was positive and significantly greater than zero, indicating that students across both treatments made substantial growth during the intervention. The effect sizes were particularly large for taught letter sounds (growth of over four standard deviations) and taught letter sound writing (growth of over three standard deviations).

Flex vs. Plain treatment differences. No mean treatment differences were observed, nor were any significant treatment interactions apparent (coefficient test *p*s > .10). In other words, there was no evidence that added practice in cognitive flexibility skills benefitted students' growth in early literacy or cognitive flexibility beyond the basic phonics intervention.

Covariate effects. Pretest was found to negatively predict the three taught letter outcomes: generally, students who started out with higher early literacy skills were predicted to have 10-12% lower growth than average, all else held constant. For word-level skills, treatment attendance was found to uniquely predict pretest-posttest change: for each standard deviation increase in tutoring session attendance, there was an 11% predicted increase in word reading and spelling growth. Finally yet importantly, EB children were predicted to have 6% greater gains for letter sound writing compared to average, all else being equal; this said, there were no other patterns indicating an advantage (or disadvantage for EB children).

Treatment moderators. Results also showed no significant treatment moderators. In other words, treatment differences did not depend on student EB status, pretest levels, or total tutoring session attendance.

Discussion

Our current study replicates and extends a previous study with younger kindergarten students (identified at the beginning of the academic year) that compared a brief 6-week phonics-only intervention ("Plain") with phonics plus cognitive flexibility training ("Flex"). Specifically, the present study compares the two treatments for relatively older kindergarten students (i.e., perhaps more developmentally ready for the instruction) identified in the middle of the academic year. Consistent with the previous study, we found that all students made significant growth, irrespective of condition, but there were no differential benefits found for early literacy or cognitive flexibility skills growth. In contrast with the previous study, we also found no differential benefits between conditions on writing/spelling growth (the previous study found significant differences favoring the “Plain” condition), nor did we evidence of joint effects between treatment condition and any of our key factors (i.e., Emergent Bilingual status, pretest, or tutoring attendance).

In this study, kindergarten students identified with limited alphabet knowledge in the middle of the school year were randomly assigned to one of two brief supplemental explicit phonics interventions, one typical (“Plain”) and one embedded with brief cognitive flexibility practice (“Flex”). Our first question examined the benefits of 6 weeks of embedded cognitive flexibility practice. While students in both groups achieved significant gains in experimenter alphabet and word reading outcomes, there were no significant overall differences between the treatments. However, we found a consistent interaction between treatment type and week 6 attendance (recalling that the last week of attendance was affected by Covid-19): for students with 100% attendance in week 6, the Flex treatment demonstrated significantly greater gains than Plain on taught-letter alphabetics, decoding, and word reading.

Our second question considered whether initial pretest alphabetics and pretest cognitive flexibility moderated treatment effects. Although our results showed that these skills were uniquely predictive of pretest-posttest gains, we found no evidence of 2- or 3-way interactions among treatment and these pretests.

Although pretest cognitive flexibility and alphabetics did not moderate treatment effects, we found that students with higher pretest cognitive flexibility made significantly greater gains on the two decoding measures – both taught-letter (experimenter) and norm-referenced (Word Attack). Moreover, a significant interaction was observed between pretest cognitive flexibility and alphabetics on all three reading measures: the positive effect of pretest cognitive flexibility was apparent for children with lower pretest alphabetics. In other words, cognitive flexibility may compensate in part for limited pretest alphabet knowledge.

Students draw upon executive functions (EF) across areas of academic learning (Fuhs et al., 2014; Morgan et al., 2019). When first learning to decode, students must activate different response sets as they shift their focus in the decoding process: they retrieve the letter sound associated with each printed letter form in working memory, sometimes also inhibiting a stronger letter name association, and apply the letter sound to individual letters in the word, which requires segmenting the word into phonemes. Blending the retrieved sounds requires a degree of abstraction to coarticulate the sounds to pronounce the word (de Graaf et al., 2009). Children must learn to do this rapidly enough to be able to recognize the word. In the earliest stage of learning to decode, the student makes the transition from simple recall of newly learned letter-sound associations to incorporate that knowledge to the new and more complex task of blending sounds. The task calls upon multiple cognitive and EF skills. For example, working memory comes into play to manipulate letter sound information stored in memory. We did not assess working memory, and the skill is less differentiated with cognitive flexibility in young children ( ). Some children must inhibit their stronger letter name retrieval when letter sounds become required. As we note, timing, differentiating and measuring EF components in young learners is imprecise. Children may draw upon each component—shifting, working memory, and inhibitory control-- more heavily as they initially coordinate and apply their mental sets for letter sound correspondences to the new goal of phoneme blending. A larger body of research on interventions to develop or enlist component EF skills for the more cognitively demanding skill of reading comprehension, primarily with older students, have also reported measurement challenges (Cirino et al., 2016).

The present study examined the role of one EF component in early reading. Cognitive flexibility practice was embedded in one of the two phonics treatments that explicitly taught alphabetic knowledge and decoding skills to kindergarteners identified with limited alphabet knowledge and no decoding skills. As found in many previous phonics intervention studies, children in both groups made significant growth on the experimenter reading measures, but the added cognitive flexibility practice did not significantly influence gains. Children with higher pretest cognitive flexibility made greater growth in decoding, and this finding corroborates the involvement of this cognitive ability in the early trajectory of reading. The role of domain-specific alphabetic and phonological skills most clearly influence the development of decoding. Our results show that within the critical time window for this study cognitive flexibility skill compensated for limited alphabet knowledge. This executive skill might help children to advance slowly in decoding while they gain the necessary alphabet knowledge for accurate and efficient decoding.

Treatment intensity, as measured by intervention attendance during week 6, was associated with decoding gains, and was related to better outcomes for children in the Flex condition. The Flex intervention may not have been intense enough to detect overall treatment effects, or reveal differences in individual responses that could inform effective training conditions (Smid et al., 2020). The Flex specific activities primarily featured practice switching letter dimensions with limited exposure to graphophonological-semantic cognitive flexibility. The majority of time in both treatments was focused on phonics skills. Alternatively, the type of cognitive flexibility practice utilized in our current and previous studies (i.e., as embedded in phonics instruction) may not be an optimal instructional routine. Rather than embedding this practice within the phonics activities, it may be more effective to provide isolated practice in cognitive flexibility tasks, such as sorting, as described in Cartwright's (2016, 2019, 2020) interventions, or in the isolated working memory practice described in the training studies mentioned earlier.

Limitations

Our study has several limitations. First, the findings can be generalized only to kindergarten children with very low levels of alphabet knowledge. Second, the final week of the intervention coincided with the onset of the Covid-19 pandemic in the U.S., which left us unable to test the full sample on all measures, especially spelling. As such, we are unable to evaluate the impact of cognitive flexibility practice on children's encoding skills. Relatedly, part of the sample was unable to complete the last week of the intervention (week 6); nevertheless, we were able to control for that variation in our models.

Third, the present study targeted a single EF component. Our cognitive flexibility task was a card sort measure with verbal directions and responses that may have been too challenging for young kindergarten subjects in the sample, many with limited language or English proficiency being provided an intervention in English. In older children, the card sort task is scored on both accuracy and time, but a timed task was too difficult for our sample. Haft et al. (2019) suggested that an aggregate executive function measure may be administered more reliably with young children than measures of EF components skills. Others have cautioned in distinguishing EF components due to the “task impurity problem” (van der Sluis et al., 2007). These researchers report that EF significantly predicted word decoding in kindergarten children when controlling for age and oral language, and indirectly predicted reading comprehension through word decoding, at a stage of learning when “word decoding is a highly effortful process” (Haft et al., 2019, p. 7). Our findings confirm measurement challenges in determining the role of EF in young children’s reading when both academic skills and cognitive abilities are developing in parallel. Further, the verbal demands of EF tasks are quite difficult for younger children.

Related to the cognitive flexibility task limitation, we did not measure children on other executive functions that children draw upon in early literacy learning and that may have moderated treatment differences. For example, post-intervention reading and spelling trajectories three years after a first-grade phonics intervention have been positively predicted by pretest rapid automatized naming (a form of switching attention), above and beyond other pretest skills (e.g., Vadasy, Sanders, & Abbott, 2008).

Last, we expected the intervention to most benefit children on the cusp of decoding, with lower but some level of alphabet and phonological knowledge. Children were screened to enroll an adequate sample for testing the intervention at a time point in kindergarten when children are often beginning to consolidate these skills. However, treatment differences were not detected on gains (or at posttest), and a planned follow-up test at the end of kindergarten was cancelled due to Covid-19. Switching practice may have been more effective provided later in the kindergarten year, once children had developed stable letter associations, such that Flex condition benefits may have been detected at later follow-up.

Comparison of Present Study with Previous Study Results

When we previously compared these two treatments with a younger kindergarten cohort in fall of the school year, the Plain treatment (with no cognitive flexibility practice) was associated with significantly higher gains in encoding, and attendance predicted alphabet and reading gains (Vadasy & Sanders, 2021b). The Plain treatment may have been more effective in the fall when students entered with lower levels of alphabet knowledge. In the early kindergarten year, students were primarily engaged in acquiring this initial alphabet knowledge. In the fall, the Flex practice may have distracted students from the phonics content, and better alphabet learning in the Plain treatment was reflected in better spelling outcomes. Most students in the fall cohort had not yet acquired accuracy in alphabetics and phonemic awareness foundations necessary for decoding. These skills were more developed by winter, which may have helped students in this winter cohort who completed week 6 of the intervention benefit from the Flex practice. Findings suggest that a longer intervention and more time and practice to consolidate skills are needed to detect treatment effects for decoding and word reading. Finally, the embedded approach to practice cognitive flexibility may have created excessive learning demands on these students for acquiring both cognitive and literacy skills.

Conclusion

Two general approaches have been taken to help children develop and practice EF skills. One group of researchers have effectively incorporated working memory practice directly into reading interventions (Peng & Kievit, 2020). Others have provided general EF skill practice as one brief component of reading interventions (Cartwright et al., 2017, 2019, 2020). The difficulty of learning the system for decoding for some children, and the role of cognitive abilities in this learning warrant further examination of approaches that might support struggling learners, including children with limited literacy experience. The finding that cognitive flexibility moderates decoding outcomes for children with lower pretest alphabet knowledge offers guidance for early reading interventions to consider individual differences in maturation of cognitive abilities. This may recommend efforts to develop EF abilities in preschool in tandem with explicit preschool alphabet instruction to prepare all kindergartners to learn to decode.

Learning how to decode requires children to coordinate often-fragile literacy knowledge and understanding of the task. Individual differences in cognitive flexibility may contribute to the difficulty some children experience in first learning how to bring these skills together. Some children have difficulty learning the system for decoding. The findings suggest measurement and timing considerations in further study of the role of EF in young beginning readers.

References

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective?: An experimental and longitudinal study. *Journal of Learning Disabilities*, *39*, 441-431. https://doi.org/10.1177%2F00222194060390050401

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading*, 18(5), 325–346. https://doi.org/10.1080/10888438.2014.902461

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi: 10.18637/jss.v067.i01

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*, 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: impact, moderation, and mediation in the Head Start REDI program. *Developmental Psychopathology, 20*, 821-843. doi:10.1017/S0954579408000394

Blachman, B. A., Schatschneider, C., Fletcher, J., Francis, D., Clonan, S., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow up. *Journal of Educational Psychology, 96*(3)*,* 444-461. https://doi.org/10.1037/0022-0663.96.3.444

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205-228. https://doi.org/10.1080/87565640801982312

Cartwright, K. B. (2002). Cognitive development and reading: The relation of reading specific multiple classification skill to reading comprehension in elementary school children. *Journal of Educational Psychology, 94*, 56–63. https://doi.org/10.1037/0022-0663.94.1.56.

Cartwright, K. B. (2019). Insights from cognitive neuroscience: the importance of executive function for early reading development and education. *Early Education and Development, 23*, 24-36. http://dx.doi.org/10.1080/10409289.2100.615025

Cartwright, K. B., Bock, A. M., Clause, J. H., Coppage August, E. A., Saunders, H. G., & Schmidt, K. J. (2020). Near- and far-transfer effects of an executive function intervention for 2nd to 5th-grade struggling readers. *Cognitive Development, 56*. https://doi.org/10.1016/j.cogdev.2020.100932

Cartwright, K. B., Coppage, E. A., Lane, A. B., Singleton, T., Marshall, T. R., & Bentivegna, C. (2017). Cognitive flexibility deficits in children with specific reading comprehension difficulties. *Contemporary Educational Psychology, 50*, 33–44. https://doi.org/10.1016/j.cedpsych.2016.01.003

Cartwright, K. B., Marshall, T. R., Huemer, C. M., & Payne, J. B. (2019). Executive function in the classroom: Cognitive flexibility supports reading fluency for typical readers and teacher-identified low-achieving readers. *Research in Developmental Disabilities, 88*, 42–52. https://doi.org/10.1016/j.ridd.2019.01.011

Christensen, C. A., & Bowey, J. A. (2005). The efficacy of orthographic rime, grapheme-phoneme correspondence, and implicit phonics approaches to teaching decoding skills. *Scientific Studies of Reading, 9,* 327-349. https://doi.org/10.1207/s1532799xssr0904_1

Cirino, P. T., Miciak, J., Gerst, E., Barnes, M. A., Vaughn, S., Child, A., & Huston-Warren, E. (2017). Executive function, self-regulated learning, and reading comprehension: a training study. *Journal of Learning Disabilities, 50*(4), 450–467. https://doi.org/10.1177/0022219415618497

Cole, P., Duncan, L. G., & Blaye, A. (2014). Cognitive flexibility predicts early reading skills. *Frontiers in Psychology*, *5*, 1-8. doi:10.3389/fpsyg.2014.00565

Compton, D. L. (2000). Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading, 4*, 219-259. https://doi.org/10.1207/S1532799XSSR0403_3

Conrad, N. J., & Levy, B. A. (2011). Training letter and orthographic pattern recognition in children with slow naming speed. *Reading and Writing: An Interdisciplinary Journal, 24,* 91-115. doi: 10.1007/s11145-009-9202-x

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of systematic phonics instruction. *Scientific Studies of Reading, 13*, 318-333. doi:10.1080/10888430903001308

Dias, N. M., & Seabra, A. G. (2016). Intervention for executive functions development in early elementary school children: effects on learning and behaviour, and follow-up maintenance. *Educational Psychology, 37*(4), 468–486. https://doi.org/10.1080/01443410.2016.1214686

Dunn, L. M., & Dunn, D. M. (2006). *Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)*. Bloomington, MN: Pearson.

Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing: An Interdisciplinary Journal, 18*, 129-155. doi:10.1007/s11145-004-5892-2

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology, 50*, 1698-1709. doi: 10.1037/a0036633

Gaskins, I. W. (2008). Developing cognitive flexibility in word reading among beginning and struggling readers. In K. B. Cartwright (Ed.), Literacy processes: Cognitive flexibility in learning and teaching (pp. 90-113). NY: Guilford Press.

Haft, S. L., Caballero, J. N., Ganaka, H., Sekelman, L., Cutting, L. E., Uchikoshi, Y., & Hoeft, F. (2019). Direct and indirect contributions of executive function to word decoding and reading comprehension in kindergarten. *Learning and Individual Differences, 76*, 101783. https://doi.org/10.1016/j.lindif.2019.101783

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research, 85*, 512-552. doi:10.3102/0034654314561338

Jacobson, L. A., Koriakin, T., Lipkin, P., Boada, R., Frijters, J.C., Lovett, M. W., Hill, D., Willcutt, E., Gottwald, S., Wolf, M., Bosson-Heenan, J., Gruen, J. R., & Mahone, E. M. (2016). Executive functions contribute uniquely to reading competence in minority youth.

Journal of Learning Disabilities, 50, 422-433. https://doi.org/10.1177/0022219415618501

Jones, C. D., & Reutzel, D. R. (2012). Enhanced alphabet knowledge instruction: Exploring a change of frequency, focus, and distributed cycles of review. *Reading Psychology, 33*, 448-464. http://dx.doi.org/10.1080/02702711.2010.545260

Kieffer, M. J., Vucovic, R. K., & Berry, D. (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. *Reading Research Quarterly, 48*, 333-348. doi:10.1002/rrq.54

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. doi: 10.18637/jss.v082.i13

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431–61. https://doi.org/10.1037/h0020279

Messer, D., Henry, L. A., & Nash, G. (2016). The relation between executive functioning, reaction time, naming speed, and single word reading in children with typical development and language impairments. *British Journal of Educational Psychology, 86*(3), 412–428. https://doi.org/10.1111/bjep.12115

Morgan, P. L., Farkas, G., Millemeier, M. M., Pun, W. J., & Maczuga, S. (2019). Kindergarten children's executive functions predict their second-grade academic achievement and behavior. *Child Development, 90*, 1802-1816. https://doi.org/10.1111/cdev.13095 .

Ober, T. M., Brooks, P. J., Homer, B. D., & Rindskopf, D. (2020). Executive functions and decoding in children and adolescents: a meta-analytic investigation. *Educational Psychology Review, 32*, 735-763. https://doi.org/10.1007/s10648-020-09526-0

Peng, P., Fuchs, D., Fuchs, L., Elleman, A.M., Kearns, D.M., Gilbert, J., ... Patton, S. (2019). A longitudinal analysis of the trajectories and predictors of word reading and reading comprehension development among at-risk readers. *Journal of Learning Disabilities*, *52*, 195–208. doi:10.1177/0022219418809080

Peng, P., & Kievit, R. A. (2020). The development of academic achievement and cognitive abilities: a bidirectional perspective. *Child Development Perspectives, 14*, 15-20. https://doi.org/10.1111/cdep.12352

Petscher, Y., & Schatschneider, C. (2011). A simulation study on the performance of the simple difference and covariance-adjusted scores in randomized experimental designs. *Journal of Educational Measurement, 48*(1), 31-43. https://doi.org/10.1111/j.1745-3984.2010.00129.x

Pressley, M., Rankin, J., & Yokoi, L. (1996). A survey of instructional practices of primary teachers nominated as effective in promoting literacy. *Elementary School Journal, 96*, 363–384.

Schaars, M. M. H., Segers, E., & Verhoeven, L. (2017). Word decoding development in incremental phonics instruction in a transparent orthography. *Reading and Writing: An Interdisciplinary Journal, 30*, 1529-1550. doi:10.1007/s11145-017-9735-3

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a science of effective cognitive training. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721420951599

Sondergaard Knudsen, H. B., de Lopez, K. J., & Archibald, L. M. (2018). The contribution of cognitive flexibility to children’s reading comprehension- the case for Danish. *Journal of Research in Reading, 41*, S130-S148. doi: 10.1111/1467-9817.12251

Spear-Swerling, L. (2018). Structured literacy and typical literacy practices: Understanding differences to create instructional opportunities. *Teaching Exceptional Children, 51*, 201-211. https://doi.org/10.1177/0040059917750160

Swanson, H. L., & O'Connor, R. E. (2009). The role of working memory and fluency practice on the reading comprehension of students who are dysfluent readers. *Journal of Learning Disabilities, 42*, 548-574. https://doi.org/10.1177/0022219409338742

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice*, *15,* 55-64.

Vadasy, P. F., & Sanders, E. A. (2021a). Introducing grapheme-phoneme correspondences (GPCs): exploring rate and complexity in phonics instruction for kindergarteners with limited literacy skills. *Reading and Writing: An Interdisciplinary Journal*, *34*, 109–138. doi:10.1007/s11145-020-10064-y

Vadasy, P. F., & Sanders, E. A. (2021b). Introducing phonics to learners who struggle: content and embedded cognitive elements. *Reading and Writing: An Interdisciplinary Journal*, *34*, 2059–2080. doi:10.1007/s11145-021-10134-9

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, *98*(3), 508–528.

Vadasy, P. F., Sanders, E. A., & Abbott, R. D. (2008). Effects of supplemental early reading intervention at 2-year follow up: Reading skill growth patterns and predictors. *Scientific Studies of Reading*, *12* (1), 51–89.

Van de Sande, E., Segers, E., & Verhoeven, L. (2018). The role of executive functions for dyadic literacy learning in kindergarten. *Early Education and Development, 29*, 192-206. doi:10.1080/10409289.2017.1393739

Van de Sande, E., Segers, E., & Verhoeven, L. (2017). How executive control predicts early reading development. *Written Language and Literacy, 20*, 170-193. https://doi.org/10.1075/wll.00003.san

Van der Sluis, S., de Jong, P. F., & Van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*(5), 427–449. https://doi.org/10.1016/j.intell.2006.09.001

Wright, D. M., & Ehri, L. C. (2007). Beginners remember orthography when they learn to read words: The case of doubled letters. *Applied Psycholinguistics, 28,* 115-134. https://doi.org/10.1017/S0142716407070063

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., Emmen, R. A. G., & Prevoo, M. J. L. (2014). Cognitive flexibility children across the transition to school: A longitudinal study. *Cognitive Development, 31*, 35-47. https://doi.org/10.1016/j.cogdev.2014.02.004

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: a meta-analytical study. *Learning and Individual Differences, 23*, 1-9. https://doi.org/10.1016/j.lindif.2012.10.004

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3–29.

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Developmental Science, 9,* 429–436.

Table 1

Observed (Unadjusted) Sample Descriptive Statistics

Measure	Flex (*n* = 29)						Plain (*n* = 28)					
	Pretest		Posttest		Gain		Pretest		Posttest		Gain	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Recept Vocab	87.14	(22.63)	--		--		91.57	(19.59)	--		--	
Letter Names	0.73	(0.21)	0.90	(0.21)	0.18	(0.24)	0.67	(0.23)	0.83	(0.25)	0.16	(0.24)
Letter Sounds	0.19	(0.17)	0.91	(0.17)	0.73	(0.18)	0.20	(0.18)	0.90	(0.12)	0.70	(0.15)
Ltr Sound Wrt	0.34	(0.27)	0.89	(0.19)	0.55	(0.22)	0.29	(0.25)	0.86	(0.17)	0.54	(0.21)
Word Reading	0.02	(0.05)	0.56	(0.32)	0.54	(0.32)	0.02	(0.07)	0.50	(0.33)	0.48	(0.31)
Spelling	0.03	(0.10)	0.52	(0.36)	0.50	(0.35)	0.01	(0.05)	0.43	(0.38)	0.42	(0.37)
Col Card Sort	0.18	(0.29)	0.35	(0.42)	0.17	(0.39)	0.13	(0.21)	0.44	(0.33)	0.30	(0.37)

Note. *N* = 57 children across 21 teachers, except posttest/gain for: Ltr Sound Wrt (*N* = 53, *n* = 29 Flex and *n* = 24 Plain), Spelling (*N* = 51, *n* = 29 Flex and *n* = 22 Plain), and Col Card Sort (*N* = 48, *n* = 27 Flex and *n* = 21 Plain). Recept Vocab = PPTV-4 standard score; Letter Names = 13 taught letter names percent correct; Letter Sounds = 13 taught letter sounds percent correct; Ltr Sound Wrt = 13 taught letter sounds correctly written percent correct; Word Reading = 16 decodable words with taught letters percent correctly read; Spelling = 16 dictated decodable words with taught letters percent correctly written; Col Card Sort = Cartwright Color Card Sort percent correct out of 6 points. Gain = change from pretest to posttest.

Table 2

Observed (Unadjusted) Zero-Order Correlations for Variables included in Models

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. Treat (1 = Flex)	--														
2. EB (1 = yes)	-.09	--													
3. Treat Attend	.21	.15	--												
Pretests															
4. Letter Names	.12	-.26	**.28**	--											
5. Letter Sounds	-.04	-.25	.10	**.55**	--										
6. Ltr Sound Wrt	.09	-.20	.19	**.51**	**.66**	--									
7. Word Reading	-.06	.05	.10	.19	**.41**	**.33**	--								
8. Spelling	.11	-.04	.14	.18	**.42**	**.33**	**.62**	--							
9. Col Card Sort	.09	-.12	.25	.22	.24	**.35**	.10	.24	--						
Pre-Post Gains															
10. Letter Names	.03	.04	-.11	**-.48**	-.15	-.09	-.15	-.03	-.05	--					
11. Letter Sounds	.09	-.01	.12	-.09	**-.62**	-.25	**-.42**	**-.33**	-.04	**.36**	--				
12. Ltr Sound Wrt	.03	**.36**	-.02	-.22	**-.46**	**-.72**	-.25	-.26	-.21	.09	.23	--			
13. Word Reading	.10	-.04	**.29**	**.53**	**.38**	**.53**	.08	.10	.23	.13	.17	-.11	--		
14. Spelling	.11	-.08	.26	**.51**	**.53**	**.73**	.16	.07	**.36**	-.03	-.06	**-.37**	**.74**	--	
15. Col Card Sort	-.18	-.09	-.06	.07	**.30**	.06	.08	.16	**-.32**	-.01	-.20	-.07	-.06	.03	--

Note. *N* = 57 children across 21 teachers, except pre-post gains for: Ltr Sound Wrt (*N* = 53, *n* = 29 Flex and *n* = 24 Plain), Spelling (*N* = 51, *n* = 29 Flex and *n* = 22 Plain), and Col Card Sort (*N* = 48, *n* = 27 Flex and *n* = 21 Plain). Treat = treatment condition, dummy coded (+1 = Flex, 0 = Plain); EB = student Emergent Bilingual status, dummy coded (1 = yes, 0 = otherwise); Treat Attend = percent of tutoring sessions attended out of 23 sessions possible; Recept Vocab = PPTV-4 standard score; Letter Names = 13 taught letter names percent correct; Letter Sounds = 13 taught letter sounds percent correct; Ltr Sound Wrt = 13 taught letter sounds correctly written percent correct; Word Reading = 16 decodable words with taught letters percent correctly read; Spelling = 16 dictated decodable words with taught letters percent correctly written; Col Card Sort = Cartwright Color Card Sort percent correct out of 6 points. Gain = change from pretest to posttest. Pearson's *r* reported; correlations in boldface are significant at the .05 level, 2-tailed.

Table 3

Multilevel Model Fixed Effects Results Predicting Pre-Post Change

Fixed Effect	Letter Names		Letter Sounds		Ltr Sound Wrt		Word Reading		Spelling		Col Card Sort	
	Coeff	*d*	*Coeff*	*d*	*Coeff*	*d*	*Coeff*	*d*	*Coeff*	*d*	*Coeff*	*d*
Intcpt (Mean)	0.16***	.70	0.71***	4.07	0.55***	3.24	0.50***	1.67	0.45***	1.30	0.24***	.63
Treat (1 = Flex)	0.03	.13	0.00	.03	0.02	.13	0.01	.02	0.02	.05	-0.05	-.13
EB (1 = yes)	-0.02	-.08	-0.02	-.15	0.06*	.36	-0.02	-.08	-0.03	-.09	-0.05	-.14
Pretest (Z)	-0.12***	-.55	-0.10***	-.85	-0.12***	-.73	0.03	.05	0.01	.02	-0.09	-.19
Attend (Z)	0.00	.01	0.03	.26	-0.01	-.05	0.11**	.36	0.11*	.29	-0.02	-.04
Treat*EB	-0.04	-.18	-0.01	-.10	0.00	-.01	0.01	.03	-0.01	-.02	0.01	.01
Treat*Pre	-0.04	-.19	-0.01	-.10	0.01	.06	-0.03	-.06	-0.05	-.13	0.00	.00
Treat*Attend	0.03	.14	0.00	-.01	-0.03	-.19	0.04	.14	0.03	.07	-0.06	-.15

Note. N = 57 children across 21 teachers, except pre-post gains for: Ltr Sound Wrt (N = 53, n = 29 Flex and n = 24 Plain), Spelling (N = 51, n = 29 Flex and n = 22 Plain), and Col Card Sort (N = 48, n = 27 Flex and n = 21 Plain). Treat = treatment condition, effect coded (+1 = Flex, -1 = Plain); EB = student Emergent Bilingual status, effect coded (1 = yes, -1 = otherwise); Treat Attend = percent of tutoring sessions attended out of 23 sessions possible, centered within classrooms and *z*-scored; Letter Names = 13 taught letter names percent correct; Letter Sounds = 13 taught letter sounds percent correct; Ltr Sound Wrt = 13 taught letter sounds correctly written percent correct; Word Reading = 16 decodable words with taught letters percent correctly read; Spelling = 16 dictated decodable words with taught letters percent correctly written; Col Card Sort = Cartwright Color Card Sort percent correct out of 6 points. Gain = change from pretest to posttest. All pretests were centered within classrooms and standardized prior to model entry; all models were 2-level multilevel linear models; d = approximate Cohen's d = Coefficient / approximate SD_{pooled}, where approximate SD_{pooled} = SE*Sqrt(N).
*p < .05, **p < .01, ***p < .001.