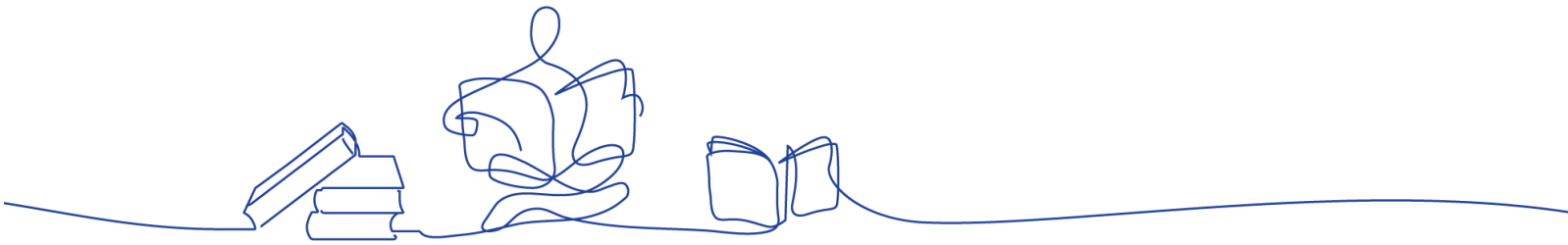


RESEARCH REPORT

Luke Waites Center for Dyslexia and Learning Disorders

Take Flight: Growth in Student Literacy Skills When Implemented in Routine Practice

Anna E. Middleton, Ph.D.

Jeremiah Ring, Ph.D.

Emily Farris, Ph.D.

Timothy N. Odegard, Ph.D.

Jeffrey L. Black, M.D.

Sheryl L. Frierson, M.D.

SCOTTISH RITE FOR CHILDREN®

Intervention studies of children with dyslexia have shown that teaching both phonological awareness and letter-sound correspondences can improve the primary deficits associated with developmental dyslexia. Comprehensive reading instruction, however, involves more than just an emphasis on the structure of English orthography but also includes vocabulary development, comprehension strategy instruction and extended practice of reading for developing fluency. A report from the National Reading Panel (NICHD, 2000) outlined five critical components for comprehensive reading instruction, which include: phonemic awareness, phonics, fluency, vocabulary, and reading comprehension. *Take Flight: A Comprehensive Intervention for Children with Dyslexia* represents such a comprehensive program that integrates those best practices and has proven efficacy for remediating the reading deficits of children with dyslexia. This report documents the effectiveness of *Take Flight* as implemented in routine practice, i.e., a public school environment.

The Curriculum

Take Flight is derived from Orton Gillingham-based principles of instruction and integrates instructional activities in each of the five component areas within each lesson (Avrit et al., 2006; Ring, Avrit & Black, 2017). The curriculum was developed for use as a pull-out program in public schools for small groups of four to six students with dyslexia, four to five days per week. The curriculum scope was developed according to evidence-based best practices for teaching the important components of a comprehensive reading program (e.g., NICHD, 2000). The sequence of the curriculum introduces reading and spelling concepts in a cumulative fashion, starting with stable and high frequency concepts including grapheme-phoneme correspondences (e.g., *p/b* and *t/d*), syllable patterns (e.g., VC), and common morphemes (e.g., *-ing*). Each lesson introduces 1-2 new concepts and integrates visual, auditory, and kinesthetic practice through phonemic awareness activities, phonics application (i.e., decoding words and sentences), repeated reading to bolster reading fluency, spelling, and comprehension strategies. After foundational skills and procedures have been taught (Lesson 36), the lesson sequence changes to include Application Lessons on alternating class days. Application Lessons provide students with an opportunity to apply previously learned skills and strategies in repeated reading exercises, spelling, and vocabulary development and comprehension strategy use with authentic text reading. Fluency practice in the curriculum is integrated throughout the lesson plans, and is addressed by oral readings at the word, sentence, and passage levels during Application Lessons. Fluency practice is organized in 18 packets of practice materials which are focused around a subset of reading concepts. Each packet begins with a 1-minute oral reading of a passage, is followed by several days of repeated practice in oral reading of words and phrases contained in the passage and ends with a second reading of the passage. Student growth is measured in number of words correct per minute.

Intervention studies conducted at the Luke Waites Center for Dyslexia Lab School have documented efficacy of the *Take Flight* program through a series of quasi-experimental designs. A historical-control study compared intervention effects across two samples of students: a group of students who received *Take Flight* instruction, and a historical-control group of students who received instruction in its predecessor curriculum, the *Dyslexia Training Program*. Results indicated significant and similar growth in word-reading and decoding ability across the two groups. In addition to growth in these word-level skills, students in the *Take Flight* group demonstrated significant growth in comprehension ability, a finding which was not documented in the DTP sample (Ring, Avrit, & Black, 2017). Data collected from the same *Take Flight* sample during the wait-list period prior to their enrollment in the Center's treatment program demonstrated reliable gains in word reading, decoding, and comprehension that were not evident during the baseline period, and

thus may be attributed to skills developed during the intervention. Furthermore, longitudinal results supported maintenance of acquired skills in word recognition and reading comprehension for 4 years following the conclusion of the intervention (Black & Ring, 2017).

Students in the Lab School benefit from a number of factors that are important to effective instruction (Vaughn, Denton & Fletcher, 2010). First, all therapists providing intervention therapies in LWCD classrooms are Certified Academic Language Therapists. The training course for a dyslexia therapist, or Certified Academic Language Therapist, is a certification that requires completing an accredited two-year Orton Gillingham-based training program. The therapist training curriculum focuses on teaching the structure of written language, multisensory structured language-based teaching methods, instructional strategies, reading development, and assessment of reading disability. Certification requires 200 instructional hours from qualified instructors and observation of experienced therapists teaching the curriculum. The training program also requires the teachers complete 700 supervised clinical hours working with children with dyslexia in their schools. In addition to providing services to children with dyslexia, LWCD serves as a dyslexia therapist training center, accredited by the International Multisensory Structured Language Education Council (IMSLEC), and live classes are used as observation opportunities for teachers in training. Therefore, therapists providing *Take Flight* instruction within LWCD classrooms must adhere strictly to the curriculum's content and procedure. Student attendance within LWCD is mandatory at 90% or greater. All students are under supervisory care of the LWCD Medical Director; instructors are given the opportunity through monthly meetings with the Medical Director to address the behavioral and emotional needs of their students.

Scale-up Issues in Routine Practice

There are a number of challenges that arise when taking evidence-based interventions from well-controlled research environments and implementing in routine practice such as in public schools (Klingner, Boardmann, & McMaster, 2013). In fact, the National Center for Education Research in the Institute of Education Sciences acknowledged the challenges of this type of research, noting that very few research programs achieve this level of evaluation (IES, 2016). Perhaps the central issue concerns implementation fidelity, that is, the extent the intervention is delivered as originally designed and tested (e.g., O'Donnell, 2008). Implementation effectiveness can be affected by many factors that can be summarized in two primary dimensions (Stein et al., 2008). The first refers to characteristics of the program itself such as professional development, on-going teacher support and supplemental support materials (e.g., manuals, workbooks, etc.). The second category of influences concerns the contextual setting such as school/organizational, teacher/classroom, and student characteristics. Each category can independently or interactively affect the quality of intervention implementation. Generalization of efficacy data to observed effects in public schools, therefore, is not necessarily assured as there are many environmental factors that are mitigated in laboratory studies. Research does indicate, however, that students in school environments benefit from well-designed interventions when both characteristics of the program and the contextual setting are accounted for in the implementation plan (e.g., Stein et al., 2008).

The District

Frisco Independent School District is a large suburban school district located in north Texas. The district enrolls over 60,000 students K-12, across 42 elementary schools, 17 middle schools, and 10 high schools. Within the district, approximately 800 children are referred for dyslexia evaluations each year; of these, approximately 500 are identified with dyslexia and placed in Tier 3 intervention classrooms to receive dyslexia therapy.

Frisco ISD recently became the first public school district in the nation to achieve IMSLEC accreditation as a dyslexia therapist training center. As such, Frisco ISD dyslexia personnel train and monitor progress of all teachers providing dyslexia therapy within their district. The district has adopted *Take Flight* as their standard dyslexia curriculum, and have modeled their teacher training, dyslexia identification, and remedial processes after guidelines provided from LWCD. Through their preparation course, Frisco ISD therapists are trained in the *Take Flight* curriculum and use *Take Flight* in all dyslexia classrooms district wide. In developing this training model, Frisco ISD provides the structure and support necessary to implement the curriculum and maximize student success.

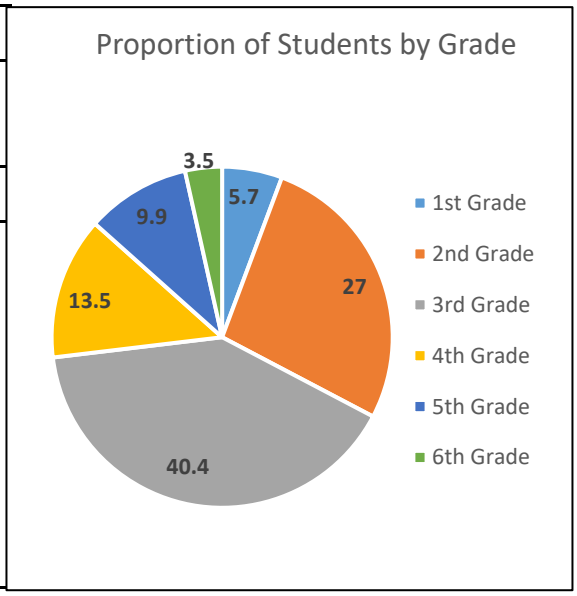
Monitoring of fidelity to intervention practices was integrated into the Frisco ISD dyslexia program implementation. Adherence to program structure and process are continuously monitored by school district dyslexia personnel. Each year, qualified instructors observed each therapist delivering the *Take Flight* program twice during scheduled classroom observations. Unscheduled walkthroughs were completed at random. Finally, all therapists were required to submit weekly monitoring forms to their supervisory team, documenting teacher and student attendance and daily lessons taught. The supervisory team ensured no lessons were omitted and addressed concerns in attendance or program adherence with the therapists directly.

Method

All instruction was delivered by Frisco ISD dyslexia therapists in daily lessons of approximately 45-minute duration for a total of 230 hours of instruction over two academic years. Longitudinal data were collected on demographic, school-level, class-level, and student-level factors and correlated with developmental trends on both summative assessments (e.g., pre- and post-tests) and more fine-grained measures of progress in component reading skills.

Table 1. *Sample Demographics*

	Mean(SD)	Range
Age (years)	8.9 (1.12)	6.7-13.0
Grade	3.06 (1.16)	1-6
	Percentage	
Gender (Female %)	54%	
Race:		
African American	7.1%	
Asian	2.1%	
Hispanic	11.3%	
White	78.7%	
Co-Existing Conditions:		
ADHD	15%	
ASD	0%	
ODD	0%	



Retrospective data from 231 individual students were entered by school personnel into an online digital database created by the research team. Of these, 90 cases were dropped due to irreconcilable data entry errors or missing data points. The final dataset entered into analyses included 141 cases with complete data across 2 years of the *Take Flight* program. These 141 cases were collected from 48 therapists across 38 schools. The sample was 54% female, the majority of cases were within 2nd-4th grades, with an average age of 8.9 years, and was mostly Caucasian (78.7%; see Table 1). Fifteen percent of the sample was reported to have co-occurring ADHD; there were no reported cases of Autism Spectrum Disorder or Oppositional-Defiant Disorder.

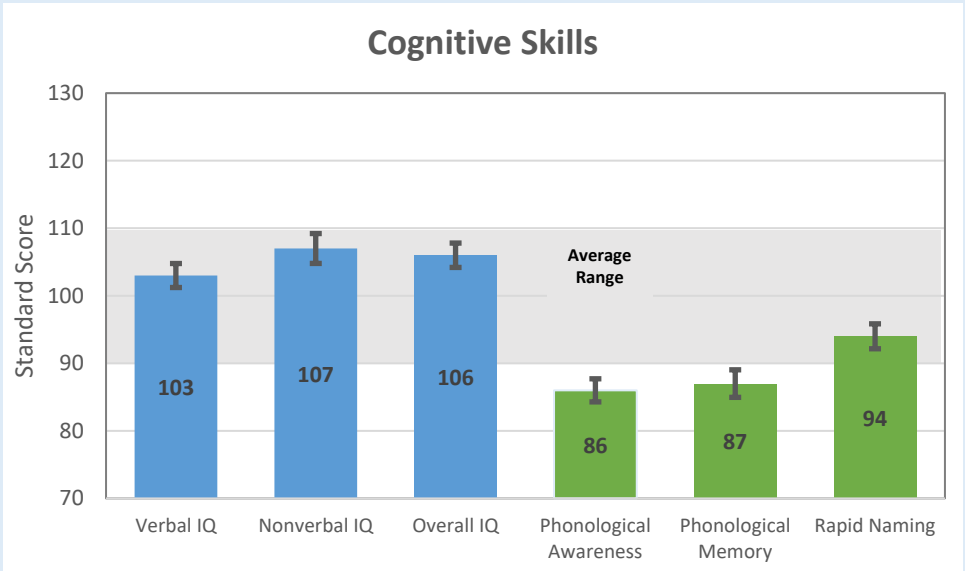


Figure 1. Mean standard scores on cognitive and phonological processing skills at Baseline. Error bars reflect 95% Confidence Intervals . The shaded area represents the Average Range (middle 50%) based on student age at time of testing.

Results

Baseline Cognitive Performance

Students in the current sample were of typical intelligence, with relative weaknesses in phonological processing abilities. On average, students in the sample scored within the Average range (25th-74th percentile) on measures of Nonverbal intelligence and Verbal intelligence, for a mean overall IQ score of 106 (SD = 11), as measured by the Kaufman Brief Intelligence Test. Performance on the Comprehensive Test of Phonological Processing was significantly weaker in comparison to general cognitive abilities (*p*s < .001). On average, students exhibited below-average performance on tests of Phonological Awareness and Phonological Memory, weaknesses which are common in individuals with dyslexia (see Figure 1). The sample performance was significantly stronger on tests of Rapid Naming, with average performance falling within the average range.

Growth in Literacy Skills

Pre- and post-intervention data were used to evaluated growth in word-level and passage-level skills. Results of tests of statistical assumptions were satisfactory. Due to correlations between variables, omnibus multivariate analyses were run on all dependent variables and followed by univariate analyses. Mean-centered age was used as a covariate in all analyses.

Word-Level Skills. Given the nature of the characteristic deficits of dyslexia (i.e., poor word-recognition, decoding, and spelling skills), effective interventions must focus on developing these word-level skills. These skills are critical in developing accuracy and automaticity with which one can decipher written text. The desired outcome of any reading intervention is to bolster reading skills and reduce the severity of observed deficits. Response to treatment from this perspective may be operationalized as growth in measured skills over time (e.g., gain score). A loftier conceptualization of response to treatment is by considering the normalization, or "closing the gap," of reading and spelling skills that were previously deficient in comparison to age-based norms. Normalization refers to scores which fall within the average range on a given measure, i.e., a standard score of 90 or greater (25th percentile or greater based on age at test) after the intervention. Although normalization of function is not always attainable for children receiving remedial reading instruction, it is a paramount achievement for those who do (Torgesen, 2001). In this sample, average student performance on norm-referenced word-level reading measures closed the gap on Word Recognition and Word Attack, suggesting normalization of these skills. Figure 2 illustrates growth over time on each of the norm-referenced word-level measures. Baseline and end of program performance on all measures reported in Table 2. By the end of the intervention, students were performing on average as well or better than 34% (SS = 94) of their age-equivalent peers on word recognition and 42% (SS = 96) of their age-equivalent peers on word attack.

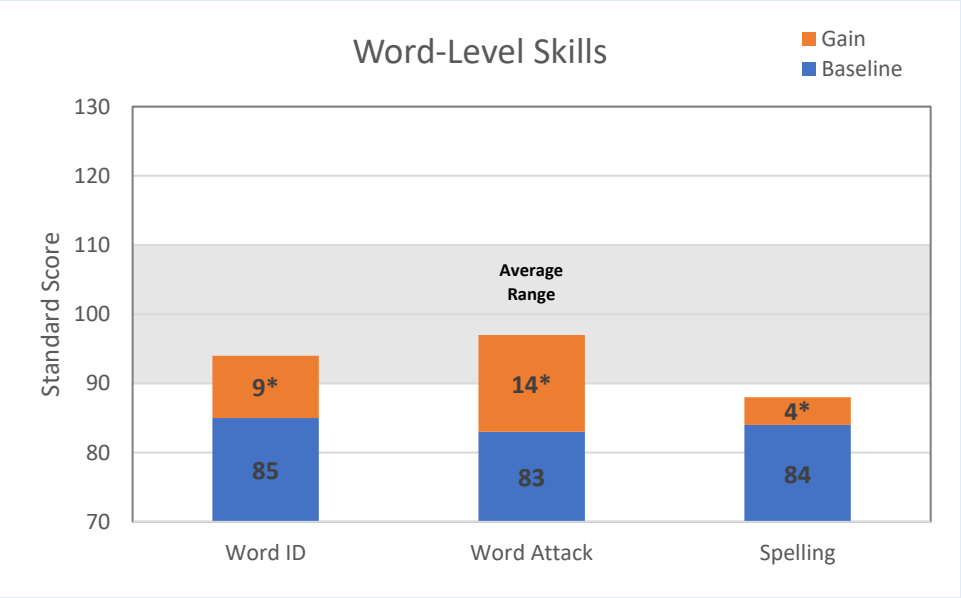


Figure 2. Baseline scores and growth on word-level reading and spelling measures. *Growth significant at p<.001. The shaded area represents the Average Range (middle 50%) based on student age at time of testing.

To take into account pre-intervention reading ability, these data were further analyzed to evaluate the amount of growth measured over the course of the program. A doubly multivariate analysis was run on word-level measures (Word ID, Word Attack, Spelling), indicating significant growth in standard scores, $F(3,137) = 63.74$, $p < .001$, $\eta^2 = .58$. Students in the *Take Flight* program exhibited significant growth in word-level skills over time compared to pre-intervention ability level.

Individual repeated-measures univariate ANCOVAs were then conducted to evaluate growth across each of the three word-level measures. Analyses indicated reliable growth in Word Identification, $F(1,139) = 97.25$, $p < .001$, $\eta^2 = .41$, Word Attack, $F(1,139) = 181.53$, $p < .001$, $\eta^2 = .57$, and Spelling, $F(1,139) = 34.07$, $p < .001$, $\eta^2 = .20$. Growth on tests of Word Recognition and Word Attack was significant and moved into the Average range after the intervention, closing the gap on these skills. At the end of the program, approximately 66% of students showed normalized scores on a test of word recognition, 75% were normalized on word attack, and 51% were normalized on spelling.

Passage-Level Measures. In addition to teaching word-level skills, comprehensive reading instruction should also address higher-level linguistic skills, such as fluency and comprehension. The *Take Flight* program addresses these by integrating repeated oral reading, wide reading, and explicit vocabulary and reading comprehension instruction into every lesson. To measure growth in these skills, a doubly multivariate analysis was performed on passage-level measures (Fluency, Timed Reading Comprehension, Untimed Reading Comprehension), indicating significant growth in standard scores, $F(3,137) = 14.92$, $p < .001$, $\eta^2 = .25$. Students in the *Take Flight* program exhibited significant growth in passage-level skills over time compared to pre-intervention ability level.

Individual repeated-measures univariate ANCOVAs were conducted to evaluate growth across each of the three passage-level measures. Analyses indicated significant growth in Timed Reading Comprehension, $F(1,139) = 13.65$, $p < .001$, $\eta^2 = .09$, and Untimed Reading Comprehension, $F(1,139) = 40.85$, $p < .001$, $\eta^2 = .23$. Figure 3 illustrates student performance on passage-level measures. Although growth on both timed and untimed tests of reading comprehension was significant, a direct comparison of growth between timed and untimed reading comprehension revealed significantly greater improvement in untimed reading comprehension than its timed counterpart $t(140)=4.15$, $p < .001$, 95% CI = 2.75-7.77. By the end of the program, 81% of students showed normalized scores on a test of Untimed Reading Comprehension, whereas 62% were within the normal range on Timed Reading Comprehension. An important consideration in comparing growth across these two tests come from way administration protocol impact scores. The Timed Comprehension score is produced from a passage-level test which measures both fluency (speed and accuracy) and comprehension. Ceilings, or discontinue criteria, are based on student performance on fluency measures, not comprehension. As a result, testing can be discontinued before the child reaches their true comprehension ceiling if they are performing poorly on reading speed and accuracy measures. Thus, performance on the comprehension subtest of the GORT-5 is subject to false-ceiling effects, which may underestimate student ability on reading comprehension.

Unlike performance on tests of comprehension, there was not a reliable change in Fluency scores over time, $F(1,139) = 3.58$, $p = .06$, $\eta^2 = .03$. Post-intervention data revealed that 46% of the sample was within the Average range on Fluency by the end of the program.

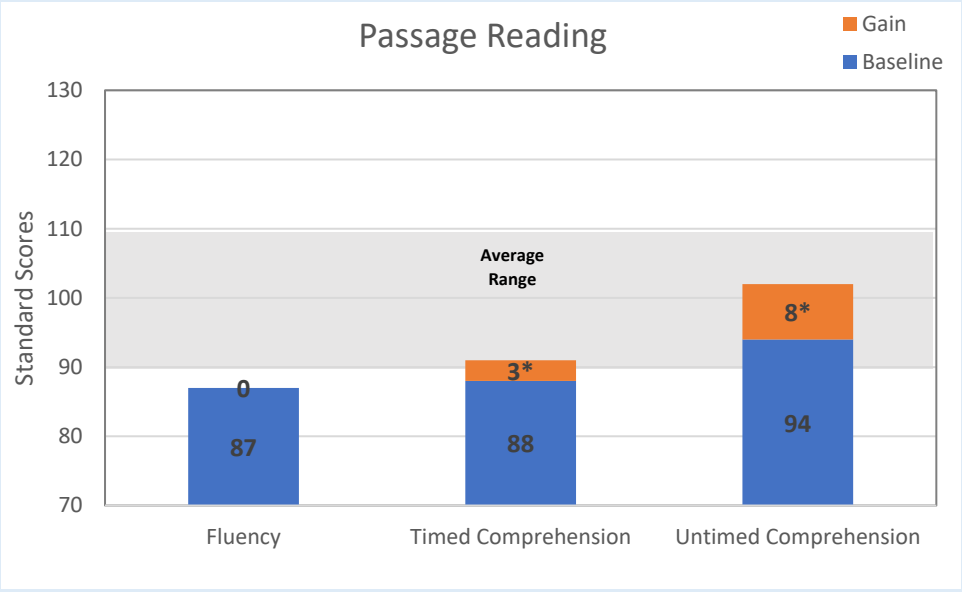


Figure 3. Baseline scores and growth on word-level reading and spelling measures. *Growth significant at p<.001.

The question of whether meaningful growth was achieved in reading fluency may not be fully apparent when evaluating standardized scores for a number of reasons. First, a standard score represents an individual's performance on a test relative to their age-equivalent peers. A 10-year-old child who scores at the 20th percentile on a test of reading will have correctly responded to a greater number of test items than an 8-year-old child with the same score and percentile ranking, as scores are leveraged based on the types of skills that are developmentally appropriate for a given age. In the current sample, students received a two-year intervention and therefore are ranked at the end of the program based on the performance that is expected given the individual's current age. Thus, a student who scores within the 20th percentile prior to an intervention and again at the 20th percentile after the intervention has increased in skill, but has not increased this skill at a rate that is faster than his or her age-equivalent peers, and therefore retains the same rank status. Raw scores can be used to evaluate absolute growth but were not available in the current dataset.

Another important consideration is the way in which reading fluency is measured and, therefore, what the score represents. Fluency scores in this sample represent combined performance on a measure of reading speed (Rate) and a measure of words read correctly within a passage (Accuracy). To better understand the nature of the changes in these component measures and how they contributed to fluency, repeated-measures ANCOVAs were run on passage-level Rate and Accuracy individually. Results revealed a reliable increase in Accuracy over the course of the intervention, $F(1,139) = 22.03$, $p < .001$, $\eta^2 = .14$. Reading Rate did not change over time, $p = .84$. The lack of growth in Reading Fluency may be attributable to stagnation in reading speed due to a number of factors. First, the students in this sample may not have responded well to instructional methods targeted towards developing reading fluency. Additionally, improvements in decoding ability may have led students to utilize a more deliberate decoding strategy that promoted

accuracy but hindered reading speed. Finally, it is possible that the amount of fluency practice the students received was insufficient to affect reading speed. On average, students completed 5.6 (out of 18) fluency packets over the course of the program. For reference, students in the LWCD lab school typically complete 12 packets in their program. Insufficient data exists to evaluate the relationship between fluency practice and reading outcomes. Over the course of the intervention, however, students' ability to apply word recognition and decoding abilities during passage reading is significantly improved.

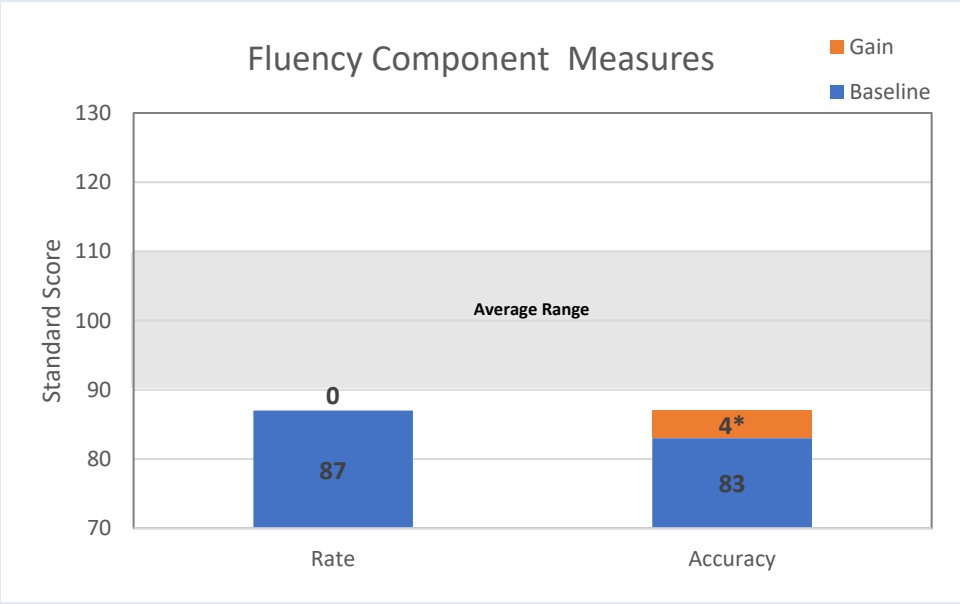


Figure 4. Baseline scores and growth on component measures of reading fluency. *Growth significant at p<.001.

Measuring Response to Decoding Instruction

Although performance on norm-reference tests can provide important information regarding students' rank status and normalization of generalized reading skill, they do not necessarily gauge how well students are able to learn the information presented in a curriculum. Therefore, scores on a curriculum-aligned progress monitoring tool were evaluated before and after the intervention. First, the proportion of correct items (50 items total) was calculated for each student's total raw score at baseline and at the end of the 2-year intervention program. A series of paired-samples *t*-tests was used to examine changes in performance on the curriculum-aligned probe over the course of the intervention. Results revealed a significant increase in decoding performance as measured by the progress monitoring measure, $t(140) = 22.73$, $p < .001$, 95% CI = 34.97-41.63. On average, students demonstrated an increase in approximately 19 additional correct items on the progress monitor after completing the *Take Flight* curriculum. This sample of students, who were performing at approximately 50% accuracy on decoding probe items prior to the intervention, reached mastery levels (90% accuracy) by the end of the program. Over the course of the

intervention, 82% of students' scores on the progress monitoring measure increased by at least 1 standard deviation (20 percentage points).

Each word item on the progress monitoring measure is decodable by concepts explicitly introduced in the program. Therefore, student learning of curriculum content was measurable by evaluating performance on these items. *Take Flight* lessons introduce combinations of phonemic awareness, phonics concepts (i.e., grapheme-phoneme correspondences), syllable division rules, morphology, spelling rules, vocabulary, and comprehension strategies. The decoding focus for the first 5 of the 7 books of the curriculum is on developing accuracy through phonics instruction, common morphemes, and syllabication procedures. Each item on the progress monitor is linked to the book and lesson in which its concepts are taught. Performance on progress monitor items was analyzed by decomposing total scores into proportions of correct items that correspond to concepts taught in each of the first 5 books of the curriculum. Significant growth was found for words containing decoding concepts taught in each of the first 5 books (all *p*s < .001), with the amount of exhibited growth increasing with each book. By the end of the intervention, word items covering concepts taught in Books 1-3 were at mastery (≥90% accuracy); items with more advanced reading concepts taught in Books 4 and 5 were just shy of mastery levels (approximately 84% accuracy for both Book 4 and 5 words). Figure 5 illustrates baseline and end of program performance on words linked to each of the first 5 books of the curriculum. Paired-sample *t*-tests were used to compare performance across the books at the end of the intervention. Students performed significantly better on words from Books 1 and 2 than all other books (all *p*s < .003), though performance did not vary between the first two books ($p = .12$). Performance on Book 4 and 5 words at the end of the intervention also did not differ ($p = .86$). Thus, students were able to learn and apply phonics patterns, morphology, and word structure concepts introduced in the curriculum with high levels of mastery and performed better on words with simpler reading concepts than those with more complex concepts.

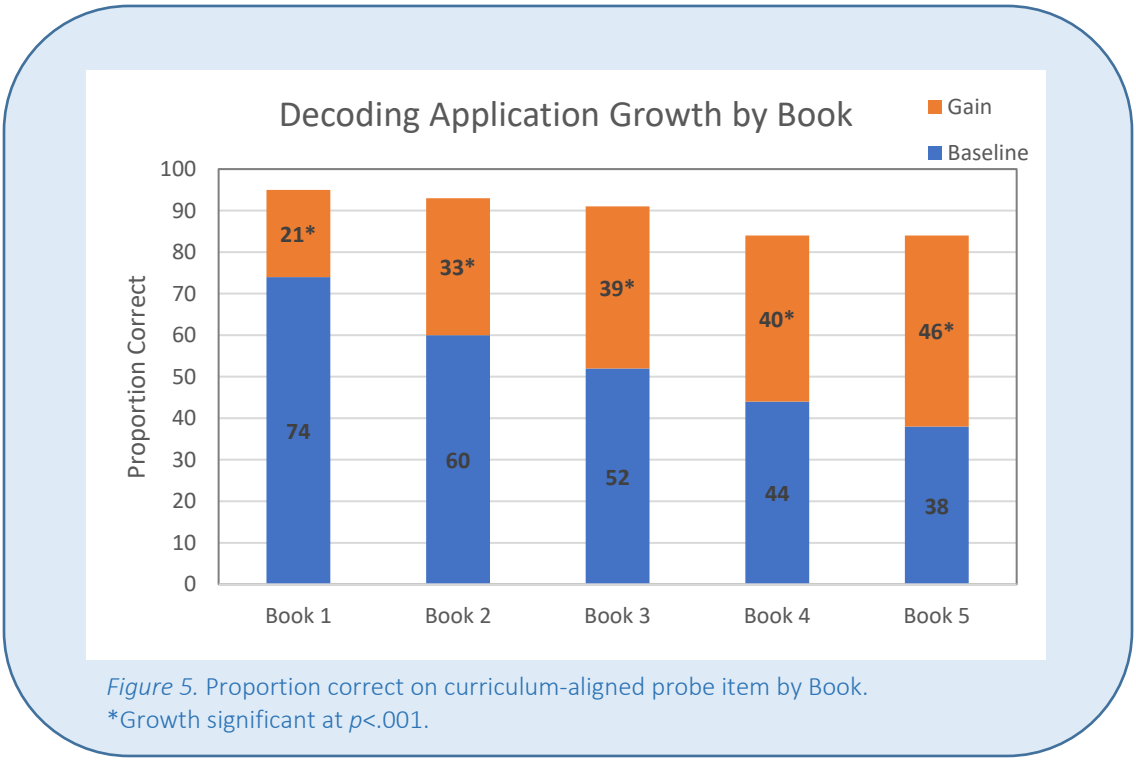


Figure 5. Proportion correct on curriculum-aligned probe item by Book.
*Growth significant at $p<.001$.

Conclusions

Results from these analyses suggest *Take Flight* is a highly effective intervention when implemented in a public school setting with high fidelity to process and structure. Results indicated significant gains in all measured reading skills but reading fluency, which is notoriously resistant to treatment (e.g., see Norton & Wolf, 2012). By the end of the intervention, students had closed the gap with their age-equivalent peers on norm-referenced tests of word recognition, decoding, and untimed and timed comprehension. Passage level reading accuracy and spelling showed less robust improvements but were nevertheless significantly stronger after completing the intervention and fell just below the average range. Although untimed reading comprehension was already a relative strength, falling within the average range prior to the intervention, these skills were further strengthened over the course of the program.

Results also support the ability of students to learn and apply decoding concepts taught throughout the curriculum, increasing performance to achieve mastery levels on the progress monitor by the end of the intervention. Reliable growth was found for concepts taught in each of the first five books of the program, suggesting students are receptive to even the more advanced decoding concepts presented late in the curriculum. When compared across Books, End of Program performance was greatest for concepts introduced in Books 1 and 2, which likely reflects both the relative ease of learning the simple, stable, and high-frequency concepts introduced early in the program, as well as a greater amount of time to apply this information through distributed practice for the duration of the program.

Together, these findings present encouraging support for the scale-up use of *Take Flight* in routine practice. They highlight the effectiveness of the program as implemented in a large school district, where contextual setting and program support allow for strong implementation fidelity. Not only were significant gains observed, but the strength of the effects reported in student word reading growth are similar to the effect sizes documented in a LWCD lab sample (Black & Ring, 2017). Contrary to previous findings, which showed improvement in reading speed standard scores related the *Take Flight* program, reading speed standard scores did not improve in this sample over the course of treatment (Black, Ring, Middleton, & Frierson, 2019). Several factors may contribute to this disparate finding. The amount of fluency practice provided in Frisco ISD dyslexia classrooms may be insufficient for eliciting strong growth in reading fluency skills. Additionally, student demonstrated significant growth in decoding ability that may cause them to read more accurately but perhaps more slowly. Despite the relative weakness in implementation of fluency practice, the outcomes reported in this paper support strong fidelity practices and the provision of high-quality instruction by Frisco ISD dyslexia personnel.

An important caveat of the current data is the lack of a control group, which is necessary to contrast the effects of the treatment to those that may be the product of general skill maturation. However, it is reasonable to conclude that the growth documented in the current sample is beyond that of typical maturation, as reflected by significant changes to percentile rank status of children's reading and spelling performance over the course of the program. That is, the students' skills in this sample are not merely developing at a similar pace to their peers, as would be indicated by no change in standard scores, but they are developing much faster, which is highlighted by significant increases in standard scores across all reading measures excluding reading rate.

The findings support *Take Flight* as an effective intervention for children with dyslexia that improves both word-level and passage-level reading and spelling skills when implemented with fidelity in a public school setting. Student scores increased in both a statistically and clinically

meaningful way, with average performance within or just below the Average range by the end of the program. It is likely, however, that differential profiles of response to the intervention exist within this sample that are not evident at the group level. Previous literature documents a non-trivial number of children receiving intensive intervention in reading do not demonstrate adequate growth (e.g., Al Otaiba & Fuchs, 2002). In light of this, future efforts should be taken to not only identify inadequate response within an intervention setting but develop a method in which response can be identified and addressed in real time.

REFERENCES

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special education*, *23*(5), 300-316.

Avrit, K., Allen, C., Carlsen, K., Gross, M., Pierce, D., & Rumsey, M. (2006) *Take Flight: A comprehensive intervention for students with dyslexia*. Texas Scottish Rite Hospital for Children. Dallas, Texas.

Black, J., Ring, J., Middleton, A., & Frierson, S. (2019). *Technical Report on Treatment Efficacy*. Texas Scottish Rite Hospital for Children. Dallas, Texas.

Denton, C. A., Vaughn, S., & Fletcher, J. M. (2003). Bringing research-based practice in reading intervention to scale. *Learning Disabilities Research & Practice*, *18*(3), 201-211.

Institute of Educational Sciences (2016, October 14). *Building Evidence: What Comes After an Efficacy Study?* Retrieved from: https://ies.ed.gov/ncer/whatsnew/techworkinggroup/pdf/BuildingEvidenceTWG.pdf

Klingner, J. K., Boardman, A. G., & McMaster, K. L. (2013). What does it take to scale up and sustain evidence-based practices? *Exceptional Children*, *79*(2), 195-211.

National Institutes of Child Health and Development (2000). *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature and its implications for reading instruction.* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual review of psychology*, *63*, 427-452.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33-84.

Ring, J. J., Avrit, K. J., & Black, J. L. (2017). *Take Flight*: the evolution of an Orton Gillingham-based curriculum. *Annals of dyslexia*, *67*(3), 383-400.

Stein, M.L., Berends, M., Fuchs, D., McMaster, K., Saenz, L., Yen, L., Fuchs, L.S., & Compton, D.L. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis*, *30*, 368-388. doi.org/10.3102/0162373708322738

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of learning disabilities*, *34*(1), 33-58.

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.