Efficacy

Examining the Efficacy of mCLASS Intervention

Sandra Pappas

Allison York

Ye Wang

Kathleen Richards

Amplify.

Amplify pioneered the adaptation of mobile technologies, including handheld computers and digital pens, for use in managing and improving teaching and learning in grades PreK–6. The company's commitment to listening to educators and gaining a deep understanding of their challenges has led to the development of offerings that optimally combine mobile tools, Internet technology, and in-person services to help improve student achievement in reading and math. Amplify's mCLASS products and services streamline the collection of data about student learning needs and school operations, facilitate data analysis and interpretation, and build educators' capacity to implement data-driven instructional programs that deliver better outcomes for children. State and district school systems across the country and overseas now rely upon these offerings to achieve and sustain growth in their classrooms. For more information about Amplify, visit us on the Web at www.amplify.com.

For more information on this study, contact:
Sandra Pappas, Ph.D., spappas@amplify.com

Amplify.

Table of contents

Abstract

This paper presents a randomized matched pair study examining the effectiveness of mCLASS Intervention in improving Tier 2 student reading achievement for elementary students in grades K–3. Results from our 2015 study suggest that student achievement in mCLASS Intervention schools is higher than similar students not in mCLASS Intervention schools on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS Next) and STAR Early Literacy (SEL) measures across one full year in grades K–3. The findings suggest that mCLASS Intervention, a data-driven, differentiated instructional intervention, promotes student achievement.

Introduction

Reading skills developed during the primary school years constitute the base upon which future student learning is built. However, many students struggle to read because they do not receive the amount and type of instruction they need (e.g., Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Morrison, Bachman, & Connor, 2005). Further, several longitudinal studies have shown that reading trajectories are distressingly stable: Students who are poor readers at the end of first or second grade almost never acquire average-level reading skills by the end of elementary school (e.g., Spira, Bracken, & Fischel, 2005; Torgesen & Burgess, 1998); these findings have important implications for early intervention approaches. Tailoring instruction to individual student abilities should maximize each student's literacy growth. However, data-driven, differentiated instruction is both challenging and time-consuming to implement in the classroom (Moats & Foorman, 2003; Cunningham, Perry, Stanovich, & Stanovich, 2004).

mCLASS Intervention helps educators become more efficient and effective at providing appropriate instruction to struggling students in an intervention setting. Students attending schools that implement mCLASS Intervention are first screened with a multi-battery assessment that (a) provides cross-skill information about a student's reading ability and (b) identifies students who are below expectations for specific skills at appropriate grade levels. mCLASS Intervention's technology-driven grouping algorithm then uses this assessment data across multiple skill categories to prioritize grouping and instructional recommendations for students who need assistance in reaching grade- and time-specific targets and to prescribe highest priority skill combinations to small groups of students with similar ability profiles.

Background

Student performance on the multi-battery screening assessment administered within mCLASS Intervention first provides information about skills that contribute to the successful development of reading comprehension and includes all of the relevant measures from Dynamic Indicators of Basic Early Literacy Skills (DIBELS Next; Good et al., 2013) that assess letter name knowledge, phonological awareness, decoding, fluency, and comprehension. Three additional mCLASS Intervention measures are also included to address vocabulary, decoding, and comprehension skills. The recommendations subsequently generated by the mCLASS Intervention algorithm incorporate instructional prioritization rules based on grade, time of year (TOY), and class-wide skill profiles. For students who have decoding needs, the mCLASS Intervention algorithm starts with the earliest component skills needed, groups students who have these needs in common, and assigns decoding instruction to match. Students who have both decoding and language comprehension needs receive skill instruction in both of these areas as early as kindergarten. The Institute of Education Sciences (IES) panel report on Response to Intervention (RTI) instruction (Gersten et al., 2008) recommends a focus on three or fewer skills at a time so as not to overwhelm students with too much information at once. mCLASS Intervention instruction is, therefore, composed of instruction that targets a maximum of two skills and designed to fit into a 30-minute intervention block in order to devote enough time to each skill. Furthermore, a focus on two skills at a time rather than three facilitates the mCLASS Intervention algorithm's ability to create groups that are more homogeneous in terms of instructional needs.

The mCLASS Intervention Skills-Based Model is based on the Simple View of Reading, which defines reading ability as a product of language comprehension and word reading skills (Gough & Tunmer, 1986). Reading with comprehension is seen as the product of decoding and language comprehension. The mCLASS Intervention Skills-Based Model takes each of those areas and breaks out the constituent skills as shown in Figure 1: Decoding skills include the phonological awareness and alphabetic principle skills shown in the inner hexagons, while language comprehension skills are shown on the upper and lower portions, including vocabulary and comprehension (both oral and written).

Each strand of instruction in mCLASS Intervention corresponds to one of the hexagons in the mCLASS Intervention Skills-Based Model as shown in Table 1. This model shows how mCLASS Intervention identifies the earliest skills that the students have yet to acquire, teaches those skills, builds on that foundation with skills acquired later, and connects each component skill to the overall goal of reading with comprehension. For example, for a group of students who can identify most letter sounds but are not blending orally or from print, the mCLASS Intervention system would recommend instruction in Phonological Awareness and Sounding Out & Blending.

Small-group mCLASS Intervention lessons include important English Language Arts skills from the Common Core State Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and are designed to facilitate growth in the five reading skills highlighted by the National Reading Panel (National Institute of Child Health and Human Development, 2000) as being empirically related to future literacy outcomes: phonemic awareness, alphabetic principle (both early and late decoding), vocabulary, fluency (both rate and prosody), and comprehension (both reading and aural). In addition to being aligned with the Simple View of Reading, each hexagon represents a component skill of reading. Skills to the left of the diagram are generally precursors of skills to those on the right of the diagram. Of course, there are exceptions; some children might become fluent readers without phonological awareness. And there is reciprocity between skills; for example, reading fluency practice improves phonological awareness. This skill-profile approach to grouping is a clear advance over methods that group students based on overall levels of ability or risk, because low-ability students can have widely varying needs (Valencia & Buly, 2005).

Teachers download sets of mCLASS Intervention lessons for their group, which are formatted to support daily 30-minute intervention sessions (Gersten et al., 2008). For ease of implementation, all mCLASS Intervention lesson templates include a familiar format (introduce, model, practice), and the lessons are organized around an effective, direct instruction approach (Carnine, Silbert, Kame'enui, & Tarver, 2004; Kamps et al., 2008; Block & Israel, 2004; Oster, 2001; Swanson, 2000). Each lesson also includes instructions for adjusting lesson pacing according to students' speed of mastery. Resulting mCLASS Intervention lessons are administered by trained interventionists and involve a high level of student-teacher interaction through the use of activities, games, and manipulatives. All intervention groups are led through a number of 10-day instructional mCLASS Intervention cycles and are progress monitored for skill development at the end of each cycle so that the subsequent lesson sequence can be tailored to changing student needs, and students can be regrouped to make intervention groups as homogeneous as possible.

The algorithm also incorporates information regarding staff capacity along with school-wide data results so that intervention time is maximized. For example, the algorithm recommends a larger number of groups in school environments where intervention time is prescribed both during and outside of the language arts block. This means that both "highest priority" intervention cases (e.g., students struggling with fluency in third grade) and "high priority" intervention cases (e.g., students who are fluent but struggling with comprehension in third grade) can be grouped for differentiated instruction.

Finally, teachers have the opportunity to exercise their own expertise and reassign students to other groups with similar needs as they see fit. At the end of one semester of instruction, typically 12 weeks after initial implementation, benchmark data is once again collected to measure the growth of all students across classrooms. This step ensures that (a) the progress of students who participated in interventions is not determined solely by submeasures targeting the particular skills they worked on in intervention, and (b) any students who were not treated, but fell behind during the initial semester, are again screened and identified as needing intervention.

A final key feature of using the mCLASS Intervention program is the professional development (PD) component. The professional development offerings prepare district and school leaders to implement mCLASS Intervention. The sessions familiarize participants with the mCLASS Intervention process, including assessments, lessons, and components of the mCLASS Intervention kit. Instructional leaders make decisions about implementation models and logistics and learn to support mCLASS Intervention instruction and monitor fidelity. Overall, mCLASS Intervention PD teaches school personnel to effectively use student performance data to plan for and monitor progress toward improved student achievement, to engage in collective decision-making and collaborative support to achieve this end, and to develop norms and expectations that support high achievement for all students, especially language minority students and those from low-income households. In the current study, schools implementing mCLASS Intervention received one to two fidelity monitoring sessions during the course of the year.

A previous quasi-experimental study design with matched treatment and control groups was conducted to examine the effectiveness of mCLASS Intervention in improving student reading achievement in kindergarten through third grade (Dubal et al., 2012). Student achievement was measured in terms of differences in raw score growth and student growth percentiles (SGPs; Betebenner, 2009) on DIBELS Next. Generally positive effects were found for mCLASS Intervention; the biggest and most important effects were found in the spring semesters of kindergarten through third grade. Results for students in mCLASS Intervention for a full year of intervention in kindergarten through third grade suggest that this group of students might benefit additionally from more intensive forms of intervention characteristic of Tier 3.

While the results from the prior study are promising, a more rigorous study design is needed to provide a deeper understanding of mCLASS Intervention's capacity to help teachers improve instructional outcomes for students. This report extends previous findings by describing the results from year one of a four-year longitudinal, cluster-randomized field study examining the effectiveness of mCLASS Intervention to improve literacy outcomes in grades K–3 (Pappas, York, & Richards, 2015). Specifically, the researchers sought to examine whether students receiving mCLASS Intervention across a full year showed higher growth on DIBELS Next and STAR Early Literacy (SEL) measures than similar students in control schools not receiving mCLASS Intervention in grades K–3.

Study design

A longitudinal matched pair randomized controlled trial experimental design was used wherein schools were randomly assigned to treatment or control groups after blocking on low socioeconomic status ([SES]; i.e., schools selected based on at least 50 percent of students eligible for free or reduced-price lunch).

As soon as schools in a single geographic area within a district fully agreed to participate, they were randomly assigned to treatment or control conditions by researchers at the University of Michigan Institute for Social Research (ISR). Random assignment procedures were as follows:

1. ISR staff received relevant school-level demographic and prior achievement data for the group of schools to be randomly assigned. The actual data used in matching varied depending on jurisdiction, but typically included multiple prior years of data on school achievement in reading, mathematics, and writing (if available); data on school size; and data on the percentage of White and free lunch students at a school.
2. With this data in hand, ISR researchers used the nbpMatching package in the R statistical environment to produce matches using the covariates at hand (Cole, Bo, & Greevy, 2015; R Core Team, 2013).
3. Once matched pairs were formed, schools were randomly assigned to treatment or control conditions.
4. After schools were randomly assigned, they were notified of their condition.

Using these procedures, 29 schools were randomly assigned to the treatment group and 28 schools were assigned to the control group. Given the odd number of schools participating in the study, one treatment group school was matched to two control schools. Schools assigned to the treatment group implemented mCLASS Intervention with Tier 2 students considered to be struggling based on DIBELS Next composite scores and therefore eligible to receive additional support.

The data presented within this report represent year one of a four-year longitudinal study.

Participants

Research staff were responsible for recruiting districts and/or individual schools for participation in the study. Initial contact was in the form of a brief description of the study sent via email to either district leaders or individual principals within districts. A meeting either on the phone or in person was then held with districts/schools interested in hearing more about the study and mCLASS Intervention. Research staff also worked with account managers to recruit schools using various methods such as recruitment flyers and existing relationships with customer accounts. The final sample included 57 elementary schools that serve ethnically diverse, lower-income student populations. Participating schools represented approximately 12,000 kindergarten through third grade students from 10 states and included 14 school districts. See information about the setting of these districts in Table 2.

Treatment schools

Intervention procedures

As described in the introduction, mCLASS Intervention consists of the following main components: (a) universal benchmark screening for all participating students during the Beginning-of-Year (BOY), Middle-of-Year (MOY), and End-of-Year (EOY) assessment administration periods; (b) algorithmic assignment of at-risk students to homogeneous intervention groups; (c) 10-day sequences of differentiated, small-group, teacher-led intervention for mCLASS Intervention groups; (d) progress monitoring after each intervention sequence followed by regrouping and instructional modifications; and (e) sustained professional development to cultivate teachers' knowledge of data-driven, differentiated intervention.

Training of school personnel

Prior to implementing mCLASS Intervention, school personnel participated in a standardized training series that included a one-day, on-site session hosted by professional development staff to prepare teachers or interventionists. This training followed a common "see one, do one" model in the class with students, so teachers could quickly learn, through context, how the mCLASS Intervention instruction should be delivered. Teachers implemented mCLASS Intervention instruction from August through May, or for the duration of their academic school year. Treatment schools also received one-to-two on-site support visits by a member of the educational support team. The support visit served as an opportunity to observe fidelity of mCLASS Intervention implementation and provide teachers with immediate constructive feedback.

Following mCLASS Intervention fidelity observations, research staff conducted teacher focus groups in both treatment and control schools. Teacher focus groups in treatment schools were led by the educational support team and focused on feedback within schools on mCLASS Intervention practices, fidelity issues, comparing experiences both within and across school settings, and providing guidance and encouragement to teachers administering the intervention. All questions were answered, and combining teachers from all grade levels provided teachers with an opportunity to learn from one another's experience with mCLASS Intervention and/or DIBELS Next.

Assessment and intervention schedule

All students assigned to mCLASS Intervention schools were assessed with the suite of Benchmark measures at BOY, MOY, and EOY during district/school pre-established assessment windows lasting approximately two weeks. Students who were selected for mCLASS Intervention received 10-day sequences of 30-minute intervention lessons, and they were assessed with progress monitoring measures corresponding to their instructed skills on the last day of each sequence. This schedule served as a guideline for educators, and analytical results should be interpreted with an understanding that strict adherence to this schedule did not always occur.

Control schools

Schools assigned to the control group implemented business-as-usual practices, which likely included district- or school-directed interventions. Interventions implemented within non-mCLASS Intervention control group schools were district/school-selected and may have differed between districts as well as between schools within a district. In addition, control schools were granted access to Now What? Tools, a supplement to DIBELS Next which supports struggling students. Now What? Tools include Item-Level Advisor, which highlights student error patterns in DIBELS Next and makes instructional recommendations; Small Group Advisor, which suggests student groupings for intervention based on DIBELS Next performance; and Home Connect, which generates letters that teachers can send to parents regarding their child's DIBELS Next performance and how they can help their children with additional reading practice at home.

Measures

Students in both treatment and control schools were administered the DIBELS Next assessment at BOY, MOY, and EOY and STAR Early Literacy (SEL; Renaissance Learning, 2008) at EOY. Treatment students were also administered a series of mCLASS Intervention-specific measures (Gushta, Parisi, Richards, Wang, & York, 2014) for both benchmark and progress monitoring purposes and intervention placement as previously described. Each student was assessed on DIBELS Next and the computer-adaptive SEL. Students in the mCLASS Intervention schools were administered three additional mCLASS Intervention assessments. Descriptions of the specific measures follow.

STAR Early Literacy (SEL)

STAR Early Literacy (SEL; Renaissance Learning, 2008) is a computer-adaptive test intended for students in prekindergarten through third grade and may be administered on a screening or progress monitoring basis. Students complete the measure independently on a computer, using a mouse or keyboard to select answers. Each administration consists of 25 items presented in multiple-choice format (three answer choices per item). Each item includes a graphic display and is dictated by audio recordings. Examinees choose answers with the computer mouse or keyboard.

SEL content was originally constructed using an item blueprint that organized item types according to seven literacy domains (general readiness, graphophonemic knowledge, phonemic awareness, phonics, comprehension, structural analysis, and vocabulary). Items within the comprehension and structural analysis domains are omitted in kindergarten administrations. An adaptive branching process is used to select individual test items based initially on the examinee's age or grade, and subsequent item selection is based on a Rasch one-parameter logistic response model. A Rasch scale is then used to express students' abilities, which are transformed to scaled scores ranging from 300 to 900 (Renaissance Learning, 2008).

Each test administration begins with an exercise to verify that the student understands the use of the mouse or keyboard, followed by a brief practice exercise that the student must pass prior to starting the test. Students complete the SEL test in approximately 10 minutes, during which time the software establishes a total scaled score ranging from 300 to 900.

The SEL technical manual reports a kindergarten split-half reliability of 0.75, one-week testretest reliability of 0.66, and "generic" reliability (i.e., an upper-bound estimate of overall reliability estimated by calculating the ratio of error variance and scaled score variance and subtracting from 1) of 0.77 (Renaissance Learning, 2008). SEL's scaled score has generic reliability ranges from 0.78 to 0.86, split-half reliability ranges from 0.75 to 0.85, and alternate form reliability ranges from 0.63 to 0.78. Concurrent validity ranges from 0.50 to 0.88 as measured by the correlations with teachers' ratings of students' skills, Brigance scale, Developing Skills Checklist, Metropolitan Early Childhood Assessment, Texas Primary Reading Inventory, and Test of Phonological Awareness; the concurrent validity ranges from 0.42 to 0.73 in predicting subsequent STAR Reading scores (Renaissance Learning, 2008). SEL total scaled scores were used in the present analyses rather than scores from the seven subscales within SEL because students may only see a limited number of items in some domains based on their item responses. Thus, scaled scores are considered the strongest estimate of a student's overall reading skills at a particular time (Renaissance Learning, 2008).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS Next)

Students in both treatment and control schools were administered the DIBELS Next assessment at BOY, MOY, and EOY. DIBELS Next is composed of six measures: First Sound Fluency (FSF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), DIBELS Oral Reading Fluency (DORF), and Daze. The DIBELS Next measures that were administered at each benchmark period were specific to the student's grade and TOY, progressing from measures of lower-level phonological awareness and phonics skills to measures of higher-level fluency and comprehension skills. Together, the measures administered at each benchmark period comprise a DIBELS Composite Score. The DIBELS Composite Score is a combination of multiple DIBELS Next scores and provides the best overall estimate of the student's reading proficiency (Good et al., 2013). While mCLASS Intervention students were also progress monitored with DIBELS measures, for the sake of comparison between mCLASS Intervention and non-mCLASS Intervention students (who may have been progress monitored on different measures or with different frequencies), only benchmark results were analyzed.

The DIBELS Next technical manual reports strong reliability support for the measures and for the overall Composite Score. Specific evidence for the reliability and validity of each of the measures and the Composite Score is presented below.

First Sound Fluency (FSF)

FSF is a test of a student's fluency in identifying the initial sounds in words. The ability to isolate the first sound in a word is an important phonemic awareness skill that is highly related to reading acquisition and reading achievement (Yopp, 1988). The ability to isolate and identify the first phoneme in a word is an easier skill than segmenting all the phonemes in words or manipulating phonemes in words, thus FSF is used as a measure of developing phonemic awareness at the beginning and middle of kindergarten.

Using standardized directions, the assessor says a series of words one at a time to the student and asks the student to say the first sound in the word. Students receive either two points for saying the initial phoneme of a word (e.g., saying the /s/ sound as the first sound in the word street) or one point for saying the initial consonant blend, consonant plus vowel, or consonant blend plus vowel (e.g., /st/, /str/, or /strea/ for street). A response is scored as correct as long as the student provides any of the correct responses listed for the word. The total score is the sum of the points the student receives in one minute.

Examining the FSF measure among kindergartners, the authors report a two-week single-form alternate form reliability of 0.85, two-week three-form alternate form reliability of 0.95, and inter-rater reliability of 0.95, indicating strong reliability for the measure (Dewey, Powell-Smith, Good, & Kaminski, 2015). Predictive validity, as measured as the correlation between the measure and the GRADE (Group Reading Assessment and Diagnostic Evaluation) end-of-year test was 0.52; the correlations with CTOPP (Comprehensive Test of Phonological Processing) end-of-year range from 0.19 to 0.49, suggesting moderate to strong validity (Good et al., 2013).

Letter Naming Fluency (LNF)

LNF is a brief, direct measure of a student's fluency with naming letters. LNF assesses a student's ability to recognize individual letters and say their letter names. The purpose of LNF is to measure students' automaticity with letter naming. Fluency in naming letters is a strong and robust predictor of later reading achievement (Adams, 1990). All letters are included on the LNF materials, but they appear in random order.

Using standardized directions, the assessor presents a page of uppercase and lowercase letters arranged in random order and asks the student to name the letters. The assessor marks letter names that are read incorrectly or skipped. The total score is the number of correct letter names that the student says in one minute.

The authors report alternate form reliability ranges from 0.86 to 0.95 and inter-rater reliability ranges from 0.99 to 1.00 (Good et al., 2013). Predictive validity, as measured as the correlation with GRADE end-of-year results, ranges from 0.35 to 0.39. This suggests moderate validity and strong reliability evidence.

Phoneme Segmentation Fluency (PSF)

PSF is a test of phonological awareness (Good, Gruba, & Kaminski, 2001). The PSF measure assesses a student's ability to fluently segment three- and four-phoneme words into their individual phonemes and is a good predictor of future reading achievement (Kaminski & Good, 1996).

During this task, an examiner orally presents words of three to four phonemes, and the student is asked to verbally produce the individual phonemes for each word. For example, when the examiner says "sat," the student should say "/s/ /a/ /t/" to receive three possible points for the word. After the student responds, the examiner presents the next word, and the number of correct phonemes produced in one minute determines the final score.

The authors report PSF alternate form reliability ranges from 0.44 to 0.78 and inter-rater reliability ranges from 0.95 to 0.98. Predictive validity, as measured as the correlation with GRADE end-of-year results, ranges from 0.24 to 0.34 (Good et al., 2013). This suggests moderate validity and strong reliability evidence.

Nonsense Word Fluency (NWF)

NWF is a brief, direct measure of the alphabetic principle and basic phonics. It assesses knowledge of basic letter-sound correspondences and the ability to blend letter sounds into consonant-vowel-consonant (CVC) and vowel-consonant (VC) words. The test items used for NWF are phonetically regular make-believe (nonsense or pseudo) words. To successfully complete the NWF task, students must rely on their knowledge of letter-sound correspondences and how to blend sounds into whole words. One reason that nonsense word measures are considered good indicators of the alphabetic principle is that "pseudowords have no lexical entry, [and thus] pseudo-word reading provides a relatively pure assessment of students' ability to apply grapheme-phoneme knowledge in decoding" (Rathvon, 2004, p. 138).

Following a model and a practice item, the student is presented with a sheet of randomly ordered VC and CVC nonsense words (e.g., dif, ik, nop). Standardized directions are used to ask the student to read the make-believe words the best they can, reading either the whole word or saying any sounds they know. For example, if the stimulus word is tof, the student could say "/t/ /o/ /f/" or "tof." The assessor underlines each correct letter sound produced either in isolation or blended together. Whole words read without sounding out are underlined in their entirety.

There are two separate scores reported for NWF: Correct Letter Sounds (CLS) is the number of letter sounds produced correctly in one minute, and Whole Words Read (WWR) is the number of pseudo-words read correctly as a whole word. These scores provide important information to educators about a student's place in the progression from sounding out individual letter-sounds to reading whole words.

The authors report that the NWF-CLS alternate form reliability ranges from 0.71 to 0.94, inter-rater reliability ranges from 0.99 to 1.00, and test-retest ranges from 0.76 to 0.90 (Good et al., 2013). Predictive validity, as measured as the correlation with GRADE end-of-year, ranges from 0.43 to 0.56. For NWF-WWR, alternate form reliability ranges from 0.90 to 0.97, interrater reliability ranges from 0.99 to 1.00, and test-retest ranges from 0.70 to 0.88, and predictive validity with GRADE ranges from 0.39 to 0.56. This suggests that NWF-CLS and NWF-WWR have moderate to strong validity and strong reliability evidence.

DIBELS Oral Reading Fluency (DORF)

DORF is a test of accuracy and fluency with connected text. The DORF passages and procedures are based on the research and development of the Curriculum-Based Measurement (CBM) of Reading by Stan Deno and colleagues at the University of Minnesota and use the procedures described in Shinn (1989). A version of CBM Reading also has been published as the Test of Reading Fluency (TORF; Children's Educational Services,1987).

DORF is a standardized set of grade-appropriate passages and administration procedures designed to identify students who may need additional instructional support and to monitor progress toward instructional goals. Students read three DORF passages (which include both literary and informational topics) aloud for one minute each, and their score for each passage is the number of words they read correctly. The final score is the median of the three passage scores; the median is used to account for slight variations in text complexity due to both text and reader characteristics.

The authors report alternate form reliability ranges from 0.88 to 0.98, inter-rater reliability ranges from 0.91 to 0.95 (Good et al., 2013), and test-retest reliability of 0.99. Predictive validity, as measured as the correlation with GRADE end of year, ranges from 0.59 to 0.77; the correlations with NAEP Oral Reading Passage range from 0.83 to 0.97. This suggests strong validity and reliability evidence.

Daze

Daze is the standardized DIBELS Next version of a maze testing procedure for measuring reading comprehension. Daze assesses the student's ability to construct meaning from text using comprehension strategies, word recognition skills, background information and prior knowledge, familiarity with linguistic properties such as syntax and morphology, and reasoning skills.

Daze can be given to a whole class at the same time, to a small group of students, or to individuals. Students are given a passage where approximately every seventh word has been replaced by a box containing the correct word and two distractor words. Using standardized directions, students are asked to read the passage silently and circle their word choices. The student receives credit for selecting the word that best fits the omitted word in the reading passage. The scores that are recorded are the number of correct and incorrect responses. An adjusted score, which compensates for guessing, is calculated based on the number of correct and incorrect responses.

The authors report alternate form reliability ranges from 0.81 to 0.93 and inter-rater reliability ranges from 0.99 to 1.00 (Good et al., 2013). Predictive validity, measured as the correlation with GRADE end of year, ranges from 0.56 to 0.68. This suggests strong validity and reliability evidence.

DIBELS Composite Score

The authors report alternate form reliability ranges from 0.66 to 0.97, inter-rater reliability ranges from 0.81 to 0.94, and test-retest reliability ranges from 0.97 to 0.99 (Good et al., 2013). Predictive validity, as measured as the correlation with GRADE end of year, ranges from 0.50 to 0.80. This suggests strong validity and reliability evidence.

mCLASS Intervention Assessments

An additional three measures were administered to students in mCLASS Intervention schools. The mCLASS Intervention-specific measures were designed to supplement DIBELS Next, assessing students' comprehension, vocabulary, and decoding knowledge of regular words, irregular words, letter combinations, and advanced phonics. These measures and evidence for reliability and validity are summarized below; for a detailed description of the measures, see the Burst Reading Assessment Technical Manual (Gushta, Parisi, Richards, Wang, & York, 2014).

Decoding (DEC)

DEC is a measure of the alphabetic principle. DEC assesses the degree to which students acquired different skills that both pave the way for the eventual consolidation of sight words and facilitate on-the-spot decoding when necessary. Four DEC submeasures were used for the mCLASS Intervention Assessments in grades 1–3. Each of the four submeasures targets different word reading skills developed in the full alphabetic, consolidated alphabetic, and sight word phases of reading that contribute to fluent reading (Carnine, Silbert, Kame'enui, & Tarver, 2004). The four DEC submeasures are:

- Regular Words (RW) assesses a student's ability to make use of regular, one-to-one letter sound correspondences in decoding real, monosyllabic words.
- Letter Combinations (LC) assesses student's ability to decode monosyllabic words containing basic letter combinations.
- Advanced Phonics (AP) assesses student's ability to decode mono-, di-, and multisyllabic words that can be broken down into parts, including morphemes and phonograms.
- Irregular Words (IW) assesses holistic word recognition skills for high-frequency irregular words.

Each submeasure consists of a printed list of eight words the student reads aloud during an independent timed reading, and the administrator marks words as correct, incorrect, or no response if the student is unable to respond to a word within five seconds. The maximum score on each DEC submeasure is eight.

Internal consistency reliability for DEC RW ranges from 0.62 to 0.86; inter-rater reliability (as measured by intraclass correlation) ranges from 0.68 to 0.93; alternate form reliability ranges from 0.53 to 0.60. Concurrent validity with SEL and DIBELS Next ranges from 0.31 to 0.71; predictive validity with SEL and DIBELS Next ranges from 0.48 to 0.78. This suggests moderate to strong reliability and validity evidence.

Internal consistency reliability for DEC LC ranges from 0.63 to 0.82; inter-rater reliability (as measured by intraclass correlation) ranges from 0.70 to 0.84; alternate form reliability ranges from 0.54 to 0.63. Concurrent validity with SEL and DIBELS Next ranges from 0.37 to 0.73; predictive validity with SEL and DIBELS Next ranges from 0.44 to 0.71. This suggests moderate to strong reliability and validity evidence.

Internal consistency reliability for DEC AP ranges from 0.63 to 0.89; inter-rater reliability (as measured by intraclass correlation) ranges from 0.83 to 0.92; alternate form reliability ranges from 0.45 to 0.69. Concurrent validity with SEL and DIBELS Next ranges from 0.28 to 0.77; predictive validity with SEL and DIBELS Next ranges from 0.46 to 0.80. This

suggests moderate to strong reliability and validity evidence. Internal consistency reliability for DEC IW ranges from 0.40 to 0.86; inter-rater reliability (as measured by intraclass correlation) ranges from 0.71 to 0.92; alternate form reliability ranges from 0.47 to 0.64. Concurrent validity with SEL and DIBELS Next ranges from 0.18 to 0.74; predictive validity with SEL and DIBELS Next ranges from 0.44 to 0.79. This suggests moderate to strong reliability and validity evidence.

Vocabulary (VOC)

VOC assesses the breadth of students' basic receptive vocabulary in kindergarten through grade 3. VOC is a picture-matching assessment in which a student is shown a set of images and asked to point to the image that exemplifies a word the test administrator read aloud. The picture-matching format isolates a student's receptive vocabulary knowledge from both reading and speaking skills, which may be confounded with vocabulary knowledge in the case of either a word-matching or an expressive vocabulary task. Choosing the correct image among distractors indicates a student's familiarity with the word's meaning. The VOC measure consists of five pages with three words, three target images, and three distractor images on each page, for a total of 15 words and 30 images per form. The test administrator reads each word aloud and asks the student to match the word to its exemplifying image. Student responses are scored as correct, incorrect, or no response if the student does not provide an answer within five seconds. The maximum score on VOC is 15.

Internal consistency reliability for VOC ranges from 0.39 to 0.93; inter-rater reliability (as measured by intraclass correlation) ranges from 0.35 to 0.99; alternate form reliability ranges from 0.22 to 0.69. Concurrent validity with SEL and DIBELS Next ranges from 0.20 to 0.50; predictive validity with SEL and DIBELS Next ranges from 0.26 to 0.72. This suggests moderate to strong reliability and validity evidence.

Comprehension skills (mCLASS Intervention CS)

mCLASS Intervention CS is a measure that screens for difficulties in reading comprehension at both the literal and the inferential levels. The mCLASS Intervention CS measure consists of grade-appropriate fiction and nonfiction texts and accompanying comprehension measures for grades 1–3. The measure is administered individually and begins with the administrator selecting either a fiction or nonfiction text. The student is then instructed to "read to yourself," implicitly encouraging the student to read silently so that cognitive resources can be maximally focused on interpreting meaning, although out loud reading is permitted.

After the student finishes reading, the text is removed and the student is asked to retell (for fiction texts) or recall (for nonfiction texts) its content for the administrator. The text is then returned to the student, and the administrator asks the student five literal questions that refer to what the text explicitly states. The student is allowed to refer to the text so that literal comprehension skill can be assessed in isolation from individual differences in memory. Following the literal questions, the examiner asks the student a series of five inferential questions and again allows the student to refer back to the text when answering. The maximum score a student can achieve on the mCLASS Intervention CS Fiction and Nonfiction measures is 14. Internal consistency reliability for mCLASS Intervention CS ranges from 0.75 to 0.88; interrater reliability (as measured by intraclass correlation) ranges from 0.32 to 0.78; alternate form reliability ranges from 0.46 to 0.64. Concurrent validity with SEL and DIBELS Next ranges from 0.38 to 0.56; predictive validity with SEL and DIBELS Next ranges from 0.42 to 0.75. This suggests moderate to strong reliability and validity evidence.

Fidelity of implementation procedures

Fidelity observations allow us to know whether users are actually implementing mCLASS Intervention instruction as it was intended, and observations can guide specific improvements on fidelity. If students in mCLASS Intervention are not making the progress we expect, fidelity observations can help us determine whether this result is due to poor implementation fidelity or whether the lack of student progress is due to some other problem with mCLASS Intervention itself. Alternatively, if students make good progress and teachers are using mCLASS Intervention properly, a component of the progress can be attributed to fidelity of mCLASS Intervention implementation (Cordray & Pion, 2006; Durlak & DuPre, 2008).

Members of the Educational Support Team (EST) conducted mCLASS Intervention fidelity site visits to treatment schools during the 2014–2015 school year. The goal during each site visit was to observe one (or more) mCLASS Intervention sessions for each grade level, document fidelity of mCLASS Intervention instruction, and provide additional support to educators implementing mCLASS Intervention. Although not always possible due to scheduling conflicts, efforts were made to observe different teachers/interventionists implementing mCLASS Intervention during each site visit.

Fidelity data was collected via the mCLASS Intervention fidelity checklist, which was developed and piloted specifically for this study. The mCLASS Intervention fidelity checklist captures ratings on six dimensions of program delivery (i.e., teacher follows mCLASS Intervention as prescribed, teacher allows each student to demonstrate beginning skills' acquisition before processing, teacher employs management skills that support mCLASS Intervention, teacher addresses individual student needs, teacher promotes active student participation, teacher completes all activities within a lesson) and includes 21 items that capture key components of mCLASS Intervention implementation rated on a three-point Likert scale (0: "None or almost none/poor," 1: "Some or fair," 2: "All or almost all/good"). The fidelity checklist and summaries of these ratings are presented in Table 3 in the Appendix. EST members underwent a day-long training on using the checklist and demonstrated at least 85 percent inter-rater agreement prior to collecting data in the field. There were two methods for collecting data: Either the observer could complete the checklist in real time during or immediately following an observation, or the observer could use a structured notes guide during the observation and translate notes into ratings on the checklist immediately following the mCLASS Intervention observation.

The observations covered 16 schools by three observers for grades K–3, with 35 responses in total. The average number of students in mCLASS Intervention groups was 4.33 when observed. The average ratings on all six dimensions generally ranged from about 1.37 to 1.91 with a mean of 1.76 (68.50% to 95.50% with a mean of 88.12%). These results suggest a high level of adherence to the implementation of mCLASS Intervention as it was designed. Specifically, teachers completed the designated activities within a lesson while monitoring progress on the targeted skills, effectively encouraged student participation and engagement toward content, and employed strategies to support individual student needs.

Results

All the students in treatment and control schools identified as at risk (below benchmark) by DIBELS Composite Score at the pretest (i.e., eligible for the intervention) are included in the analyses. Intent to treat (ITT) analyses were used because this model estimates the causal effects of a school's assignment to treatments, where these effects are averaged over schools with possibly very different levels of treatment implementation and schools with very different pretreatment conditions in spite of matching. As such, ITT analyses provide policymakers with information about the size of treatment effects that can be expected when treatments are implemented in the field under typically imperfect levels of implementation or under varying pre-existing conditions in schools. Further, best practices in Response to Intervention (RTI) models suggest that students who are Red (intensive/at risk) or Yellow (strategic/below benchmark) on DIBELS Next should receive further instruction through interventions such as mCLASS Intervention (Shapiro, 2008). Therefore, the analyses evaluated the impact of mCLASS Intervention on only those students who were eligible to receive additional support, comparing only Red/Yellow students in the schools receiving mCLASS Intervention to Red/Yellow students in schools that did not receive mCLASS Intervention.

Participants

The demographic characteristics of the students in the treatment and control samples is similar across all variables (i.e., gender, race, special education, English as a second language, and free/reduced price lunch eligible). For example, a large proportion of students within both conditions were White (56.21% treatment and 42.74% control), followed by Black (17.57% treatment and 18.01% control) and Hispanic (9.27% treatment and 15.12% control). See Table 4 for a description of the student population in mCLASS Intervention (treatment) versus non-mCLASS Intervention (control) schools. The effect size of the demographic differences between control and treatment ranges from 0 to 0.26 with a mean of 0.03; small effect sizes indicate little difference in demographic characteristics between treatment and control.

The mCLASS Intervention algorithm determines individual student eligibility based on DIBELS Benchmark Status or risk category (i.e., Red, Yellow, and Green). Students who were identified as At Risk (Red) or Some Risk (Yellow) at the pretest are considered eligible for mCLASS Intervention. DIBELS Composite Scores corresponding to the 25th percentile and 30th percentile for each grade as determined in a national norming study conducted by Dewey, Kaminski, & Good (2014), as well as the percentage of students in the treatment and control groups whose scores fell below each of the percentile thresholds, are presented in Table 5.

It is important to note that DIBELS Benchmark Status for each grade was not determined according to norm-referenced methodology; that is, risk was determined through prediction of later success on an outcome measure and not determined according to percentile ranks. Students who are categorized as Red (Well Below Benchmark) likely need intensive support. Their odds of achieving subsequent early literacy goals are approximately 10–20 percent and those students are unlikely to achieve subsequent reading benchmarks unless provided with substantial, intensive instructional support. Students who are categorized as Yellow (Below Benchmark) likely need strategic support and their odds of achieving subsequent early literacy goals are roughly 40–60 percent. Students with scores in this range typically need strategic, targeted instructional support to ensure they make adequate progress and achieve subsequent reading benchmarks (Good et al., 2013). Therefore, while it is not expected that all students eligible for mCLASS Intervention would fall below the 25th or 30th percentiles, all students eligible for mCLASS Intervention were identified based on risk and need for additional instructional support.

Pretest

First a comparison of the DIBELS Next pretest composite scores (DIBELS CS) of the treatment and control groups was conducted to determine whether these two groups had significant academic differences prior to the start of the study. Analysis of DIBELS BOY 2013 pretest DIBELS CS suggest that there were no significant differences between the DIBELS CS of students in mCLASS Intervention and non-mCLASS Intervention schools in kindergarten, grade 2, and grade 3 [kindergarten: $t(1314) = 0.98$, n.s.; Grade 2: $t(1008) = 0.55$, n.s.; Grade 3: $t(1003) = 0.59$, n.s.]. However, the DIBELS CS of Grade 1 students in mCLASS Intervention schools was significantly higher than those who were in the nonmCLASS Intervention schools [$t(1253) = 2.82$, $p < 0.05$; see Table 6]. Thus, ANCOVA (Analysis of Covariance) and adjusted means will be used to examine post-test scores, thereby controlling for any pretest differences.

Growth and post-test results

We hypothesized greater growth and post-test performance among treatment group students as compared to control group students. Growth is represented according to Student Growth Percentiles (SGPs), which describe the "rank" of students' current assessment performance relative to other students who had the same score on a previous assessment. For example, an SGP of 90 indicates that a student's growth was better than 90 percent of students with similar start scores, representing exemplary growth and the range of scores is from 1 to 99. Post-test performance was examined using a one-way ANCOVA to examine whether there was a statistically significant difference between treatment and control group schools on DIBELS Next measures and STAR Early literacy scaled scores controlling for pretest DIBELS Next CS.

Student growth percentiles

Student Growth Percentiles (SGPs; Betebenner, 2009) use quantile regression to provide a normative context for interpreting student growth based on prior performance. SGPs were calculated using DIBELS Next Composite Score across beginning of the year (BOY) to the end of the year (EOY). SGP calculated based on DIBELS Next Composite Score provides an overall indicator for the growth in order to compare the control and treatment group.

Table 7 shows SGPs by grade and condition (treatment versus control). In kindergarten, grade 1, grade 2, and overall, mCLASS Intervention students achieved growth that was significantly higher than their matched pair control group counterparts from the beginning to the end of the year (Overall across grades: $Z = 3.83$, $p < 0.05$; kindergarten: $Z = 2.57$, $p < 0.05$; grade 1: $Z = 2.39$, $p < 0.05$; grade 2: $Z = 2.59$, $p < 0.05$). In grade 3, growth was not significantly different between mCLASS Intervention (i.e., treatment) and control groups ($Z = 0.11$, n.s.).

ANCOVA results

A one-way between-groups analysis of covariance (ANCOVA) was conducted to compare the effect of the intervention program in the treatment group to the control group's performance on post-test DIBELS Next measures and STAR Early Literacy at the end of the year. DIBELS Next pretest CS scores were used as the covariate in this analysis in order to obtain the adjusted post-test means. Table 8 shows the ANCOVA results by measure and condition (treatment versus control) for the full sample.

Overall, the ANCOVA results suggest that student scores in kindergarten, grade 1, and grade 2 were significantly higher for mCLASS Intervention treatment groups than control groups on the majority of the measures (effect sizes range from 0.11 to 0.50). As with the SGP results, the performance of students in grade 3 is not significantly different between treatment and control groups (effect sizes range from –0.08 to 0.03). The effect sizes are typically highest in kindergarten and grade 1. The highest effect was found in NWF-WWR results for kindergarten [$t(1152) = 5.16$, $p < 0.05$, effect size = 0.50].

Further, we provide ANCOVA results for disaggregated samples in Table 9, to Table 14. Specifically, in Table 9, the results for students who were categorized as low-performing during the pre-test (i.e., their DIBELS Next composite scores were below the 20th percentile at the beginning of the year) were presented. In Table 10 to Table 14, the results for students who were categorized as White, Black, Hispanic, learning English as a second language, and subsidized lunch are presented respectively. Overall, results for the disaggregated samples are consistent with the full sample; the results for students in kindergarten, grade 1, and grade 2 were significantly higher for mCLASS Intervention treatment groups than control groups on most measures.

Discussion

The purpose of this paper is to examine the effectiveness of mCLASS Intervention by comparing the reading achievement of students within schools implementing mCLASS Intervention (treatment) with similar students in schools that do not use mCLASS Intervention (control). First, treatment fidelity within the mCLASS Intervention schools was examined to ascertain the degree to which mCLASS Intervention was implemented as intended. Next, differences in the reading achievement of students in mCLASS Intervention compared to their business-as-usual counterparts were investigated.

Fidelity of implementation procedures

A critical component of any study examining the effectiveness of an intervention is determining fidelity of implementation: whether the intervention was implemented as it was designed (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). The assumption is that schools implementing with high fidelity will result in improved student outcomes over and above those schools with low fidelity of implementation regardless of the intervention used. In the current study, treatment fidelity was examined through direct observation of mCLASS Intervention lessons at least once over the course of the year. Observation of teachers or coaches during 30-minute mCLASS Intervention lessons and the completion of the mCLASS Intervention fidelity observation checklist served as a measure of teacher or coach adherence to the dimensions considered critical for effective implementation of mCLASS Intervention. Results indicate a high degree of adherence to prescribed mCLASS Intervention lesson content and progression through skills within a lesson. In addition, teachers were rated on the mCLASS Intervention fidelity checklist as engaging during the delivery of lesson content, providing encouragement and reinforcement of student participation, and supporting or challenging students when appropriate. In brief, the teachers observed were found to implement mCLASS Intervention with fidelity.

Growth and post-test results

Overall, results suggest that mCLASS Intervention was successful in boosting student performance; the treatment group demonstrated performance over and above that which was achieved by the control group.

The SGP findings presented in this paper suggest that participation in mCLASS Intervention leads to improved student performance over and above that which is achieved by the control group. Student Growth Percentiles examining whether students receiving mCLASS Intervention across a full year show larger growth on DIBELS Next measures than similar students in control schools in grades K–2, suggesting that the mCLASS Intervention may have a stronger impact in grades K–2 than in grade 3. Additionally, the impact of mCLASS Intervention was evident in students' performance on SEL, an external post-test reading achievement assessment. Specifically, students in mCLASS Intervention schools scored higher on SEL compared to their control group counterpart in grades K–2. However, this difference was small in grade 2 and no difference was noted in grade 3.

There are several possible explanations for the lack of mCLASS Intervention impact on grade 3. For example, this result may be related to higher DIBELS Next performance at BOY for grade 3 students in mCLASS Intervention schools, which limited the opportunity for growth (i.e., a ceiling effect). In addition, state-mandated testing and associated preparation in grade 3 may have been prioritized over participation in mCLASS Intervention thereby reducing the number of mCLASS Intervention cycles offered to struggling students in grade 3. Lastly, research supports the notion that growth in upper elementary school grades improves at a slower pace (e.g., Christ, Silberglitt, Yeo, & Cormier, 2010). The students who still need intervention in grade 3, and to a lesser extent in grade 2, may be further behind grade level than those in the lower grades, and the grade-level DIBELS Next measures that they were tested with may not have captured progress made on earlier skills that have not yet generalized to connected text reading (e.g., phonological awareness skills may not have yet impacted Oral Reading Fluency scores, but phonological awareness measures not administered during benchmark periods in grade 3).

Furthermore, homogeneity of instructional needs within intervention groups promotes intervention efficacy (Gersten et al., 2008), but the longer a school waits to provide highquality intervention to struggling students, the more heterogeneous the classroom profile of skills becomes, as some students may be much further behind in reading development than their classmates. If a class has made it all the way to third grade without effective intervention, the heterogeneity in skill levels may reduce the homogeneity of the small groups that the mCLASS Intervention algorithm generates, which means that some "outlier" students may not have had classmates to work with that had similar needs. In kindergarten and grade 1, where we see the most growth, greater homogeneity of skill profiles is likely. We predict that the earlier students have access to mCLASS Intervention instruction, the better the intervention program will work to improve achievement over time and reduce the overall need for intervention.

In sum, further examination of grade 3 is required to evaluate the validity of these hypothesized explanations for the grade 3 results, as are more years of mCLASS Intervention implementation. The data presented in this paper represent the first year of a four-year longitudinal study suggesting that during the next three years, a stronger impact in grade 3 may emerge as those students with reading difficulties will have the advantage of receiving mCLASS Intervention consistently as they progress from kindergarten through grade 2, and/or grade 3. Perhaps greater growth will be observed with a greater number of years of participating in the study and subsequently receiving mCLASS Intervention.

Next steps

In the coming year, attention will be given to several of the patterns suggested by these results. First, the impact of mCLASS Intervention on the performance of students who are eligible to receive the intervention will be explored. This is particularly important for kindergartners as these students lack the historical data needed to inform mCLASS Intervention. mCLASS Intervention relies on a data-driven algorithm; therefore, as more student-level data is gathered through progress monitoring, the algorithm is better able to align student needs with instructional support. This is not a unique feature of mCLASS Intervention and is a common challenge encountered by any data-driven instructional model new to a school/district. As educators gain familiarity with the assessment model and mCLASS Intervention program, and as more information/data enters the system, the match between student needs and instruction becomes better.

Next, the differential impact of mCLASS Intervention during the fall versus the spring semesters will be monitored with particular attention to whether the plateau between the middle-and end-of the-year persists. Although there is some research suggesting that growth within the fall semester is larger than the spring (e.g., Ardoin & Christ, 2008; Christ, Silberglitt, Yeo, & Cormier, 2010), a previous quasi-experimental study examining the effectiveness of mCLASS Intervention found the opposite, that is, larger gains in the spring compared to the fall semester (Dubal et al., 2012). In an effort to eliminate extraneous factors that may have contributed to this result, mCLASS Intervention fidelity site visits will be conducted in the fall and spring with all treatment schools to further support and streamline implementation. In addition, other aspects of fidelity such as adherence to progress monitoring schedules and instructional downloads will also be examined. However, there may be other factors that contribute to this effect. For example, larger fall growth may be related to increased efforts on the part of the school to mitigate summer learning loss. If this is the case, growth may be related to improved performance as well as regaining what was learned in the spring of the previous year. Further, decreased intervention intensity in the spring may be related to an increased focus on standardized test preparations in the grades 3 and above and factors beyond the control of the schools (e.g., time off due to inclement weather, which uncharacteristically affected many areas of the country during the 2013–2014 school year).

Another consideration is the frequency of mCLASS Intervention cycles implemented both within and between treatment schools. As schools and teachers become more acclimated to mCLASS Intervention, the frequency of mCLASS Intervention cycles delivered both within and across years is expected to increase and positively impact student performance. Changes in the number of mCLASS Intervention cycles by time of year will be noted and further explored. This may be particularly important for grade 3 as the lack of impact of mCLASS Intervention on students in grade 3 suggests that significant effort over a longer period of time may be necessary before improvements in student performance occurs. Closer attention to grade 3 student performance and increased efforts to consistently implement mCLASS Intervention cycles in this grade despite external pressures (e.g., preparation for state testing) will be imperative to achieving SGP gains.

In addition to the aforementioned grade 3 focused efforts, hosting informal training and troubleshooting sessions with teachers throughout the course of a year is critical for the success of any newly introduced product or service across all grades. The most important lessons learned over the course of the past year relate to the importance of training and supporting teachers and schools to implement mCLASS Intervention with fidelity. These types of sessions afford teachers the opportunity to share information, resources, and easily implemented solutions such as creating words lists or using whiteboards to streamline instructional preparation. Customizing mCLASS Intervention training and support serve to address each school's current needs with the goal of increasing intervention compliance.

Acknowledgement

This work was supported in part by Grant R305A120639, the mCLASS Intervention Efficacy Study, the Institute of Education Sciences, U. S. Department of Education, and the University of Michigan (Brian Rowan, principal investigator).

Table of contents

References

Adams, M. J. (1990). Beginning to read: Learning and thinking about print. Cambridge, MA: MIT Press.

Ardoin, S. P., & Christ, T. J. (2008). Evaluating curriculum-based measurement slope estimate using data from tri-annual universal screening. *School Psychology Review, 37*, 109–125.

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4), 42–51.

Block, C., & Israel, S. (2004). The ABCs of performing highly effective think-alouds. *The Reading Teacher*, 58, (2), 154–167.

Carnine, D., Silbert, J., Kame'enui, E., & Tarver, S. (2004). *Direct Instruction Reading, Fourth Edition*. Upper Saddle River, NJ: Pearson Education, Inc.

Children's Educational Services (1987). *Test of Reading Fluency*. Minneapolis, MN: Author.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-Based measurement of Oral Reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review, 39*, 447–462.

Cole, B., Bo, L., & Greevy, R. (2015). *nbpMatching: Functions for Optimal Non-bipartite Matching*. R package version 1.4.4. **http://CRAN.R-project.org/package=nbpMatching**

Cordray, D. S., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research methodology. Psychological measurement and evaluation* (pp. 103–124). Washington, DC: American Psychological Association.

Cunningham, A., Perry, K., Stanovich, K., & Stanovich, P. (2004). Disciplinary knowledge of K–3 teachers and their knowledge calibration in the domain of early literacy. *Annals of Dyslexia*, 54(1), 139–163.

Dewey, E. N., Kaminski, R. A., & Good, R. H. (2014). *DIBELS Next® National Norms 2012– 2013* (Technical Report No. 17). Eugene, OR: Dynamic Measurement Group.

Dewey, E. N., Powell-Smith, K. A., Good, R. H., & Kaminski, R. A. (2015). *DIBELS Next Technical Adequacy Brief*. Eugene, OR: Dynamic Measurement Group, Inc.

Dubal, M., Harnly, A., Pavlov, M., Richards, K., Yambo, D., & Gushta, M. (2012). *Effects of Burst:Reading Early Literacy Intervention on Student Performance: 2012 Report*. Wireless Generation Research Report.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.

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

Gersten, R., Compton, D., Connor, C., Dimino, J., Linan-Thompson, S., & Tilly, W. (2008). *Assisting Students Struggling With Reading: Response to Intervention (RTI) and Multi- Tier Intervention in the Primary Grades* (No. NCEE 2009–4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Good, R.H., Gruba, J., & Kaminski, R.A. (2001). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds), *Best Practices in School Psychology IV* (pp. 679–700). Washington, DC: National Association of School Psychologists.

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. A. (2013). *DIBELS Next Technical Manual*. Eugene, OR: Dynamic Measurement Group, Inc.

Gough, P. & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10.

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice*, 15, 198–205.

Gushta, M., Parisi, D., Richards, K., Wang, Y., & York, A. (2014). *Burst:Reading Assessment Technical Manual*. Amplify Assessment Research Report.

Kaminski, R., & Good, R. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25(2), 215–227.

Kamps, D., Abbott, M., Greenwood, C., Wills, H., Veerkamp, M., & Kaufman, J. (2008). Effects of small-group reading instruction and curriculum differences for students most at risk in kindergarten two-year results for secondary- and tertiary-level interventions. *Journal of Learning Disabilities*, 41(2), 101–114.

Moats, L., & Foorman, B. (2003). Measuring teachers' content knowledge of language and reading. *Annals of Dyslexia*, 53(1), 23–46.

Morrison, F., Bachman, H., & Connor, C. (2005). *Improving Literacy in America: Guidelines From Research*. New Haven, CT: Yale University.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.

National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 000-4769). Washington, DC: U. S. Government Printing Office.

Oster, L. (2001). Using the think-aloud for reading instruction. *The Reading Teacher*, 55, 65– 69.

Pappas, S., York, A. E., & Richards, K. (2015). *Examining the efficacy of the mCLASS Intervention*. Poster presented at the annual National Association of School Psychologists (NASP) Convention, Orlando, FL.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rathvon, N. (2004). Early reading assessment: A handbook for practitioners. New York: Guilford Press.

Renaissance Learning. (2008). *STAR Early Literacy: Technical manual*. Wisconsin Rapids, WI: Author.

Shapiro, E. S. (2008). Best Practices in Setting Progress Monitoring Goals for Academic Skill Improvement. *Best Practices in School Psychology* V, 141–157.

Shinn, M.R. (1989). Curriculum-based measurement: Assessing special children. New York: Guilford.

Spira, E., Bracken, S., & Fischel, J. (2005). Predicting improvement after first-grade reading difficulties: The effects of oral language, emergent literacy, and behavior skills. *Developmental Psychology*, 41(1), 225–234.

Swanson, H. (2000). What instruction works for students with learning disabilities? Summarizing the results from a meta-analysis of intervention studies. In R.Gersten, E. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues*, pp.1–30. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Torgesen, J., & Burgess, S. (1998). Consistence of reading-related phonological process throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. Metsala, & L. Ehri, (Eds.) Word Recognition in Beginning Literacy. 161–188. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Valencia, S.W., & Buly, M.R. (2005). What struggling readers REALLY need. In S. Barrentine & S. Stokes (Eds.). *Reading assessment: Principles and practices for elementary teachers*, 2nd edition. Newark, DE: International Reading Association. (reprinted from *The Reading Teacher*). Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. Reading Research Quarterly, 159–177.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly*, 159–177.

Figures and tables

Figure 1. The mCLASS Intervention Skills-Based Model.

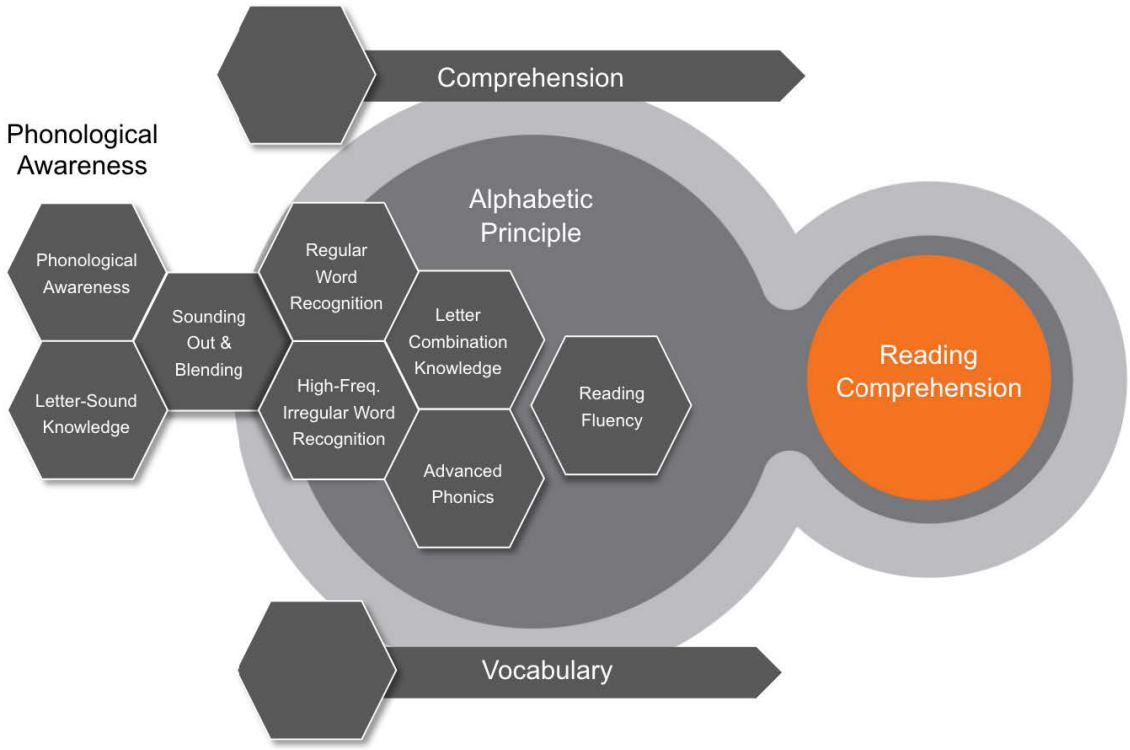

Table of contents

Table 1 | Alignment of Instructional Strands to Hexagons in the mCLASS Intervention Skills- Based Model.

Skill	Strand Abbrev.	Strand	Strand Goal
Phonologica Awareness	PhoA	Phonological Awareness Instruction	Students can blend and segment sounds in spoken words.
Letter-Sound Knowledge	LS	Letter-Sound Instruction	Students can say the most common sound for single letters.
Sounding Out & Blending	SO	Sounding-Out Instruction	Students can sound out and read regular words with simple patterns such asVC and CVC.
Regular Word Recognition	CT	Connected Text Instruction	Students can read and understand short sentences that are 100% decodable.
Irregular Word Recognition	IW	"High-Frequency Irregular Word Instruction"	Students can read 120 high-frequency irregular words.
Letter Combination Knowledge	LC	Letter Combination Instruction	Students can decode words containing common letter combinations (such as sh and oa) as well as VCe words.
Advanced Phonics	AP	Advanced Phonics Instruction	Students can decode words containing advanced text features such as common word families, contractions, compound words, etc.
Reading Fluency	Flu	Reading Fluency Instruction	Students can read connected text with grade- level automaticity.
Vocabulary	Voc	Vocabulary Instruction	Students know the meaning of grade- level words and high-frequency prefixes and suffixes.
Comprehension Strategies	CompS	Comprehension Strategies Instruction	Students know when and how to apply core comprehension strategies such as summarizing, making inferences, and identifying the author's purpose.

In accordance with the National Center for Education Statistics, the 57 schools participating in the study represent multiple settings (e.g., city, rural, suburban). The following is a list of school districts by state and NCES categories:

Table 2

State	District	NCES's urban-centric locale categories
AR	Conway Public Schools	City, Small
CA	Lodi Unified School District	City, Large
CO	Lake County Schools	Town, Remote
CO	North Conejos School District	Rural, Remote
CO	Holly School District No. RE–3	Rural, Remote
CO	Summer Scholars	Suburban
CT	Ansonia Public Schools	Suburban, Large
GA	Murray County Schools	Suburban, Small
KY	Elliott County Schools	Rural, Remote
LA	Recovery School District RSD	City, Midsize
LA	Terrebonne Parish School District	Suburban, Midsize
MD	Baltimore City Public Schools	City, Large
NY	Teaching Firms of America	City, Large
TN	Morgan County Schools	Rural, Distant
WV	Logan County Schools	Town, Distant

Table of contents

Table 3 | Fidelity implementation checklist results.

Fidelity Questions	N	Mean	SD	Percent Meet Fidelity
I. Teacher follows Burst as prescribed.				
Has all materials ready/prepared	35	1.71	0.46	85.50%
Uses the model/practice and instruction models	33	1.88	0.33	94.00%
Completes all parts of the instructions within the activity in order (e.g., Preview – Practice – Model – Wrap-Up)	35	1.89	0.32	94.50%
II. Teacher allows each student to demonstrate beginning skills' acquisition before processing.				
Provides each individual student with an individual turn	35	1.89	0.32	94.50%
Ensures each student provides at least one accurate response within the activity	35	1.91	0.28	95.50%
III. Teacher employs management skills that support Burst.				
Transitions smoothly and quickly between parts of activities (e.g., transitions smoothly from Preview to Wrap-Up)	35	1.37	0.6	68.50%
Promotes student engagement and precludes disruptions	35	1.83	0.57	91.50%
Paces to fit student needs	34	1.59	0.61	79.50%
Verbally and/or nonverbally praises students for appropriate behavior (e.g., nonverbal behavior: teacher smiles at student, nods his/her head, etc.)	35	1.86	0.43	93.00%
IV. Teacher addresses individual student needs.				
Differentiates as needed using Support/Challenge activities	7	1.43	0.98	71.50%
Reinforces correct student responses (e.g., teacher says, "Yes, the sound is mmmmm")	34	1.85	0.5	92.50%
Provides corrective feedback for incorrect student responses	35	1.86	0.36	93.00%
Re-engages off-task students	22	1.73	0.63	86.50%
Encourages students who are unsure	33	1.88	0.48	94.00%
V. Teacher promotes active student participation.				
Engages students with expression	35	1.71	0.57	85.50%
Encourages students to be involved in activities	34	1.91	0.38	95.50%
VI. Across all activities within one lesson.				
Completes all activities in order (i.e., A, B, C, D, E)	35	1.66	0.48	83.00%
Summary	-	1.76	-	88.12%

Table 3 | Student demographic characteristics by condition.

Demographic Characteristics	Treatment		Control	
	Number	Percentage	Number	Percentage
VII. Grade Level				
Kindergarten	666	29.25%	654	27.78%
Grade 1	621	27.27%	643	27.32%
Grade 2	516	22.66%	497	21.11%
Grade 3	474	20.82%	560	23.79%
VIII. Race–Ethnicity				
Black	400	17.57%	424	18.01%
American Indian	22	0.97%	31	1.32%
Asian/Pacific Islander	6	0.26%	5	0.21%
Hispanic	211	9.27%	356	15.12%
White	1280	56.21%	1006	42.74%
Other	358	15.71%	532	22.59%
IX. Socioeconomic Status				
Subsidized Lunch	1332	58.50%	1187	50.42%
No Subsidized Lunch	515	22.62%	597	25.36%
Not Specified	430	18.88%	570	24.21%
X. Disability Status				
With a Disability	254	11.16%	213	9.05%
Not Identified With a Disability	1216	53.40%	1373	58.33%
Not Specified	807	35.44%	768	32.63%
XI. ELL Status				
English Language Learner	711	31.23%	460	19.54%
Not English Language Learner	639	28.06%	658	27.95%
Not Specified	927	40.71%	1236	52.51%
XII. Gender				
Female	873	38.34%	898	38.15%
Male	1133	49.76%	1032	43.84%
Not Specified	271	11.90%	424	18.01%

Table of contents

Table 5 | Percentage of students below national norm percentiles by experimental group and grade

Grade (DIBELS CS at Percentile)	Treatment	Control
XIII. 25th Percentile		
K (10)	49.70%	46.64%
1 (94)	62.96%	55.68%
2 (125)	84.88%	83.30%
3 (191)	78.48%	78.93%
XIV. 30th Percentile		
K (14)	62.46%	60.70%
1 (100)	71.01%	67.03%
2 (138)	97.29%	97.59%
3 (209)	91.14%	91.61%

Table 6 | DIBELS Next pretest composite scores by grade and condition.

Measures (Name)	Treatment			Control		
	n	Mean	Standard Deviation	n	Mean	Standard Deviation
VII. Grade Level						
DIBELS Next CS Grade K	666	10.34	8.08	654	10.98	8.36
DIBELS Next CS Grade 1	621	79.11	26.4	643	83.2	25.16
DIBELS Next CS Grade 2	516	74.51	45.12	497	72.94	45.92
DIBELS Next CS Grade 3	474	127.16	67.89	560	124.66	67.77

Table 7 | Student growth percentiles by grade, TOY, and condition.

Grade	N of Students (mCLASS Intervention Schools)	N of Students (Control Schools)	Median SGP (mCLASS Intervention Schools)	Median SGP (Control Schools)	Significance Test (Wilcox Z)
All	2003	2019	53	48	3.83 ($p < 0.05$)
K	588	567	59	51	2.57 ($p < 0.05$)
1	556	557	50	43	2.39 ($p < 0.05$)
2	458	446	53.5	45.5	2.59 ($p < 0.05$)
3	401	449	50	51	0.11 (n.s.)

Table of contents

Table 8 | Post-test results for DIBELS Next and SEL: full sample.

Measure Name	Variables	Burst	Control	Significance Test
DORF G1	Unadjusted mean	33.93	31.41	t (1110) = 3.38, p < 0.05, effect size = 0.30
	Adjusted mean	34.8	30.55	
	Unadjusted standard deviation	24.1	22.74	
	N	556	557	
DORF G2	Unadjusted mean	54.91	49.63	t (901) = 4.15, p < 0.05, effect size = 0.27
	Adjusted mean	54.48	50.07	
	Unadjusted standard deviation	24.93	23.94	
	N	458	446	
DORF G3	Unadjusted mean	70.63	71.19	t (847) = –1.19, n.s., effect size = –0.08
	Adjusted mean	70.14	71.63	
	Unadjusted standard deviation	28.56	28.47	
	N	401	449	
DAZE G3	Unadjusted mean	12.43	12.08	t (847) = 0.41, n.s., effect size = 0.03
	Adjusted mean	12.33	12.17	
	Unadjusted standard deviation	6.98	7.2	
	n	401	449	

Table of contents

Table 9 | Post-test results for DIBELS Next and SEL: Subgroup DIBELS CS BOY below 20th percentile.

Measure Name	Variables	Burst	Control	Significance Test
DIBELS Composite Score GK	Unadjusted mean	102.73	93.58	t (431) = 1.68, n.s., effect size = 0.17
	Adjusted mean	102.15	94.23	
	Unadjusted standard deviation	51.1	46.95	
	n	230	204	
DIBELS Composite Score G1	Unadjusted mean	75.88	68.86	t (524) = 1.81, n.s., effect size = 0.16
	Adjusted mean	77.19	67.32	
	Unadjusted standard deviation	71.17	65.1	
	n	285	242	
DIBELS Composite Score G2	Unadjusted mean	128.65	112.4	t (636) = 2.37, p < 0.05, effect size = 0.17
	Adjusted mean	126.14	114.99	
	Unadjusted standard deviation	84.88	78.3	
	n	324	315	
DIBELS Composite Score G3	Unadjusted mean	200.23	205.49	t (522) = –1.14, n.s., effect size = –0.07
	Adjusted mean	199.33	206.27	
	Unadjusted standard deviation	110.72	109.44	
	n	244	281	
STAR Early Literacy Scaled Score GK	Unadjusted mean	592.76	573.86	t (386) = 1.29, n.s., effect size = 0.13
	Adjusted mean	590.41	576.64	
	Unadjusted standard deviation	100.6	112.16	
	n	211	178	
STAR Early Literacy Scaled Score G1	Unadjusted mean	659.61	642.77	t (473) = 2.42, p < 0.05, effect size = 0.21
	Adjusted mean	661.49	640.46	
	Unadjusted standard deviation	106.48	105.77	
	n	263	213	
STAR Early Literacy Scaled Score G2	Unadjusted mean	725.73	714.42	t (548) = 1.21, n.s., effect size = 0.10
	Adjusted mean	724.12	716.2	
	Unadjusted standard deviation	92.26	88.23	
	n	289	262	
STAR Early Literacy Scaled Score G3	Unadjusted mean	767.45	775.84	t (473) = –1.48, n.s., effect size = –0.13
	Adjusted mean	766.39	776.69	
	Unadjusted standard deviation	91.42	81.81	
	n	211	265	

Table of contents

Table 9 | Post-test results for DIBELS Next and SEL: Subgroup DIBELS CS BOY below 20th percentile.

Measure Name	Variables	Burst	Control	Significance Test
LNF GK	Unadjusted mean	38.92	34.97	t (431) = 1.88, n.s., effect size = 0.19
	Adjusted mean	38.68	35.24	
	Unadjusted standard deviation	20.08	18.04	
	n	230	204	
PSF GK	Unadjusted mean	37.79	36.23	t (431) = 0.56, n.s., effect size = 0.06
	Adjusted mean	37.59	36.45	
	Unadjusted standard deviation	20.44	21.62	
	n	230	204	
NWF-CLS GK	Unadjusted mean	26.02	22.38	t (431) = 2.20, p < 0.05, effect size = 0.22
	Adjusted mean	25.88	22.54	
	Unadjusted standard deviation	16.86	14.5	
	n	230	204	
NWF-CLS G1	Unadjusted mean	41.08	40.17	t (524) = 1.05, n.s., effect size = 0.10
	Adjusted mean	41.44	39.75	
	Unadjusted standard deviation	20.33	19.5	
	n	285	242	
NWF-WWR GK	Unadjusted mean	2.16	0.35	t (431) = 5.07, p < 0.05, effect size = 0.50
	Adjusted mean	2.13	0.39	
	Unadjusted standard deviation	4.75	1.34	
	N	230	204	
NWF-WWR G1	Unadjusted mean	8.13	8.38	t (524) = 0.03, n.s., effect size = 0.00
	Adjusted mean	8.25	8.23	
	Unadjusted standard deviation	8.28	7.95	
	n	285	242	
DORF G1	Unadjusted mean	25.23	21.59	t (524) = 2.75, p < 0.05, effect size = 0.25
	Adjusted mean	25.61	21.14	
	Unadjusted standard deviation	21.83	18.4	
	n	285	242	

Table 9 | Post-test results for DIBELS Next and SEL: Subgroup DIBELS CS BOY below 20th percentile.

Measure Name	Variables	Burst	Control	Significance Test
DORF G2	Unadjusted mean	46.24	41.2	t (636) = 2.95, p < 0.05, effect size = 0.21
	Adjusted mean	45.56	41.9	
	Unadjusted standard deviation	22.91	20.63	
	n	324	315	
DORF G3	Unadjusted mean	58.02	60.09	t (522) = –1.61, n.s., effect size = –0.10
	Adjusted mean	57.8	60.27	
	Unadjusted standard deviation	27.44	26.94	
	n	244	281	
DAZE G3	Unadjusted mean	9.89	9.67	t (522) = 0.33, n.s., effect size = 0.03
	Adjusted mean	9.85	9.71	
	Unadjusted standard deviation	6.26	6.36	
	n	244	281	

Table of contents

Table 10 | Post-test Results for DIBELS Next and SEL: White students.

Measure Name	Variables	Burst	Control	Significance Test
DIBELS Composite Score GK	Unadjusted mean	121.25	120.06	t (642) = 0.75, n.s., effect size = 0.07
	Adjusted mean	121.78	119.37	
	Unadjusted standard deviation	44.53	41.83	
	n	366	279	
DIBELS Composite Score G1	Unadjusted mean	108.96	102.2	t (599) = 3.40, p < 0.05, effect size = 0.29
	Adjusted mean	114.34	96.06	
	Unadjusted standard deviation	78.11	72.46	
	n	321	281	
DIBELS Composite Score G2	Unadjusted mean	161.1	151.81	t (463) = 1.81, n.s., effect size = 0.12
	Adjusted mean	161.26	151.58	
	Unadjusted standard deviation	88.45	85.78	
	n	271	195	
DIBELS Composite Score G3	Unadjusted mean	251.44	272.92	t (457) = –2.22, p < 0.05, effect size = –0.13
	Adjusted mean	254.43	268.6	
	Unadjusted standard deviation	115.16	103.87	
	n	272	188	
Star Early Literacy Scaled Score GK	Unadjusted mean	625.92	626.2	t (607) = 0.32, n.s., effect size = 0.03
	Adjusted mean	627.12	624.59	
	Unadjusted standard deviation	98.36	103.63	
	n	350	260	
STAR Early Literacy Scaled Score G1	Unadjusted mean	690.39	697.47	t (567) = 0.70, n.s., effect size = 0.06
	Adjusted mean	696.26	690.71	
	Unadjusted standard deviation	103.87	104.96	
	n	305	265	
STAR Early Literacy Scaled Score G2	Unadjusted mean	747.04	743.53	t (434) = 0.78, n.s., effect size = 0.07
	Adjusted mean	747.91	742.27	
	Unadjusted standard deviation	85.52	91.96	
	n	257	180	

Table of contents

Table 10 | Post-test Results for DIBELS Next and SEL: White students.

Measure Name	Variables	Burst	Control	Significance Test
STAR Early Literacy Scaled Score G3	Unadjusted mean	790.39	796.67	t (432) = –0.76, n.s., effect size = –0.07
	Adjusted mean	790.75	796.18	
	Unadjusted standard deviation	81.84	81.32	
	n	252	183	
LNF GK	Unadjusted mean	45.02	44.3	t (642) = 0.89, n.s., effect size = 0.08
	Adjusted mean	45.22	44.05	
	Unadjusted standard deviation	18.43	16.19	
	n	366	279	
PSF GK	Unadjusted mean	44.99	46.53	t (642) = –0.90, n.s., effect size = –0.09
	Adjusted mean	45.15	46.32	
	Unadjusted standard deviation	16.94	17.39	
	n	366	279	
NWF-CLS GK	Unadjusted mean	31.24	29.23	t (642) = 2.03, p < 0.05, effect size = 0.19
	Adjusted mean	31.41	29	
	Unadjusted standard deviation	16.18	15.38	
	n	366	279	
NWF-CLS G1	Unadjusted mean	50.97	49.88	t (599) = 2.36, p < 0.05, effect size = 0.21
	Adjusted mean	52.35	48.31	
	Unadjusted standard deviation	24.62	20.76	
	n	321	281	
NWF-WWR GK	Unadjusted mean	3.7	2.13	t (642) = 4.28, p < 0.05, effect size = 0.41
	Adjusted mean	3.75	2.06	
	Unadjusted standard deviation	5.68	4.43	
	n	366	279	
NWF-WWR G1	Unadjusted mean	12.54	11.57	t (599) = 3.03, p < 0.05, effect size = 0.28
	Adjusted mean	13.13	10.89	
	Unadjusted standard deviation	10.32	9.29	
	n	321	281	

Table of contents

Table 10 | Post-test Results for DIBELS Next and SEL: White students.

Measure Name	Variables	Burst	Control	Significance Test
DORF G1	Unadjusted mean	34.34	32.02	t (599) = 3.38, p < 0.05, effect size = 0.30
	Adjusted mean	35.92	30.22	
	Unadjusted standard deviation	23.92	22.57	
	n	321	281	
DORF G2	Unadjusted mean	56.47	53.12	t (463) = 2.34, p < 0.05, effect size = 0.15
	Adjusted mean	56.52	53.05	
	Unadjusted standard deviation	25.31	24.72	
	n	271	195	
DORF G3	Unadjusted mean	70.66	76.39	t (457) = –2.34, p < 0.05, effect size = –0.14
	Adjusted mean	71.42	75.29	
	Unadjusted standard deviation	28.53	28	
	n	272	188	
DAZE G3	Unadjusted mean	12.85	13.78	t (457) = –1.09, n.s., effect size = –0.09
	Adjusted mean	12.99	13.57	
	Unadjusted standard deviation	7.17	6.92	
	n	272	188	

Table 11 | Post-test Results for DIBELS Next and SEL: Black students.

Measure Name	Variables	Burst	Control	Significance Test
DIBELS Composite Score GK	Unadjusted mean	123.71	99.1	t (166) = 3.77, p < 0.05, effect size = 0.35
	Adjusted mean	123.44	99.3	
	Unadjusted standard deviation	45.76	42.32	
	n	72	97	
DIBELS Composite Score G1	Unadjusted mean	80.8	96.91	t (204) = −2.00, p < 0.05, effect size = −0.17
	Adjusted mean	80.45	97.22	
	Unadjusted standard deviation	68.45	71.87	
	n	98	109	
DIBELS Composite Score G2	Unadjusted mean	142.95	113.7	t (199) = 2.57, p < 0.05, effect size = 0.16
	Adjusted mean	137.96	118.89	
	Unadjusted standard deviation	81.97	83.45	
	n	103	99	
DIBELS Composite Score G3	Unadjusted mean	236.3	216.98	t (115) = 1.11, n.s., effect size = 0.06
	Adjusted mean	233.54	219.84	
	Unadjusted standard deviation	120.9	111.19	
	n	60	58	
STAR Early Literacy Scaled Score GK	Unadjusted mean	600.26	578.99	t (124) = 1.12, n.s., effect size = 0.11
	Adjusted mean	599.87	579.3	
	Unadjusted standard deviation	91.27	112.42	
	n	57	70	
STAR Early Literacy Scaled Score G1	Unadjusted mean	652.11	654.12	t (150) = −0.16, n.s., effect size = −0.01
	Adjusted mean	651.95	654.3	
	Unadjusted standard deviation	102.13	101.55	
	n	80	73	
STAR Early Literacy Scaled Score G2	Unadjusted mean	717.54	715.89	t (138) = −0.02, n.s., effect size = 0.00
	Adjusted mean	716.64	716.94	
	Unadjusted standard deviation	94.61	86.55	
	n	76	65	
STAR Early Literacy Scaled Score G3	Unadjusted mean	790.59	753.05	t (75) = 1.81, n.s., effect size = 0.16
	Adjusted mean	786.95	756.69	
	Unadjusted standard deviation	70.13	93.58	
	n	39	39	

Table of contents

Table 11 | Post-test Results for DIBELS Next and SEL: Black students.

Measure Name	Variables	Burst	Control	Significance Test
LNF GK	Unadjusted mean	46.54	38.89	t (166) = 3.15, p < 0.05, effect size = 0.29
	Adjusted mean	46.44	38.96	
	Unadjusted standard deviation	15.31	16.94	
	n	72	97	
PSF GK	Unadjusted mean	44.25	31.51	t (166) = 4.32, p < 0.05, effect size = 0.41
	Adjusted mean	44.13	31.59	
	Unadjusted standard deviation	23.21	16.86	
	n	72	97	
NWF-CLS GK	Unadjusted mean	32.92	28.71	t (166) = 1.66, n.s., effect size = 0.16
	Adjusted mean	32.87	28.75	
	Unadjusted standard deviation	15.98	16.2	
	n	72	97	
NWF-CLS G1	Unadjusted mean	42.59	41.77	t (204) = 0.24, n.s., effect size = 0.02
	Adjusted mean	42.5	41.85	
	Unadjusted standard deviation	20.19	22.33	
	n	98	109	
NWF-WWR GK	Unadjusted mean	3.36	2.3	t (166) = 1.31, n.s., effect size = 0.13
	Adjusted mean	3.35	2.31	
	Unadjusted standard deviation	5.9	4.51	
	n	72	97	
NWF-WWR G1	Unadjusted mean	9.15	11.11	t (204) = –1.84, n.s., effect size = –0.17
	Adjusted mean	9.12	11.14	
	Unadjusted standard deviation	8.55	8.91	
	n	98	109	
DORF G1	Unadjusted mean	26.86	29.61	t (204) = –1.11, n.s., effect size = –0.10
	Adjusted mean	26.75	29.7	
	Unadjusted standard deviation	21.33	22.7	
	n	98	109	
DORF G2	Unadjusted mean	48.06	41.01	t (199) = 2.18, p < 0.05, effect size = 0.14
	Adjusted mean	46.79	42.33	
	Unadjusted standard deviation	23.4	19.98	
	n	103	99	

Table 11 | Post-test Results for DIBELS Next and SEL: Black students.

Measure Name	Variables	Burst	Control	Significance Test
DORF G3	Unadjusted mean	64.07	63.26	t (115) = −0.19, n.s., effect size = −0.01
	Adjusted mean	63.37	63.98	
	Unadjusted standard deviation	30.55	28.39	
	n	60	58	
DAZE G3	Unadjusted mean	11.02	10.48	t (115) = 0.33, n.s., effect size = 0.03
	Adjusted mean	10.9	10.6	
	Unadjusted standard deviation	6.4	6.17	
	n	60	58	

Table of contents

Table 12 | Post-test Results for DIBELS Next and SEL: Hispanic students.

Measure Name	Variables	Burst	Control	Significance Test
DIBELS Composite Score GK	Unadjusted mean	137.1	117.8	t (151) = 2.83, p < 0.05, effect size = 0.27
	Adjusted mean	137.28	117.68	
	Unadjusted standard deviation	46.18	41.86	
	n	61	93	
DIBELS Composite Score G1	Unadjusted mean	137.32	93.14	t (164) = 4.75, p < 0.05, effect size = 0.44
	Adjusted mean	141.17	90.62	
	Unadjusted standard deviation	73.1	73.36	
	n	66	101	
DIBELS Composite Score G2	Unadjusted mean	190.05	150.78	t (107) = 0.93, n.s., effect size = 0.06
	Adjusted mean	171.08	160.4	
	Unadjusted standard deviation	86	89.71	
	n	37	73	
DIBELS Composite Score G3	Unadjusted mean	263.39	244.54	t (87) = 0.46, n.s., effect size = 0.03
	Adjusted mean	254.65	249.13	
	Unadjusted standard deviation	96.98	97.73	
	n	31	59	
STAR Early Literacy Scaled Score GK	Unadjusted mean	612.67	594.2	t (144) = 1.25, n.s., effect size = 0.12
	Adjusted mean	613.88	593.36	
	Unadjusted standard deviation	106.86	98.31	
	n	60	87	
STAR Early Literacy Scaled Score G1	Unadjusted mean	707.87	661.63	t (157) = 3.76, p < 0.05, effect size = 0.34
	Adjusted mean	713.84	657.76	
	Unadjusted standard deviation	109.23	99.53	
	n	63	97	
STAR Early Literacy Scaled Score G2	Unadjusted mean	762.78	710.69	t (101) = 2.28, p < 0.05, effect size = 0.19
	Adjusted mean	751.23	716.81	
	Unadjusted standard deviation	81	84.73	
	n	36	68	
STAR Early Literacy Scaled Score G3	Unadjusted mean	793.48	787.53	t (88) = 0.14, n.s., effect size = 0.01
	Adjusted mean	790.79	788.93	
	Unadjusted standard deviation	68.83	64.79	
	n	31	60	

Table of contents

Table 12 | Post-test Results for DIBELS Next and SEL: Hispanic students.

Measure Name	Variables	Burst	Control	Significance Test
LNF GK	Unadjusted mean	49.39	42.95	t (151) = 2.30, p < 0.05, effect size = 0.22
	Adjusted mean	49.45	42.91	
	Unadjusted standard deviation	19.28	16.51	
	n	61	93	
PSF GK	Unadjusted mean	49.33	45.41	t (151) = 1.36, n.s., effect size = 0.13
	Adjusted mean	49.4	45.36	
	Unadjusted standard deviation	17.59	19.43	
	n	61	93	
NWF-CLS GK	Unadjusted mean	38.38	29.44	t (151) = 3.37, p < 0.05, effect size = 0.33
	Adjusted mean	38.42	29.41	
	Unadjusted standard deviation	17.76	15.57	
	n	61	93	
NWF-CLS G1	Unadjusted mean	52.79	47.08	t (164) = 2.26, p < 0.05, effect size = 0.21
	Adjusted mean	53.85	46.38	
	Unadjusted standard deviation	22.2	22.37	
	n	66	101	
NWF-WWR GK	Unadjusted mean	4.38	1.68	t (151) = 3.08, p < 0.05, effect size = 0.30
	Adjusted mean	4.39	1.67	
	Unadjusted standard deviation	6.65	4.54	
	n	61	93	
NWF-WWR G1	Unadjusted mean	13.83	10.56	t (164) = 2.89, p < 0.05, effect size = 0.28
	Adjusted mean	14.2	10.32	
	Unadjusted standard deviation	9.07	8.74	
	n	66	101	
DORF G1	Unadjusted mean	44.79	28.94	t (164) = 5.19, p < 0.05, effect size = 0.49
	Adjusted mean	45.88	28.23	
	Unadjusted standard deviation	25.1	21.48	
	n	66	101	
DORF G2	Unadjusted mean	67.08	50.95	t (107) = 2.73, p < 0.05, effect size = 0.16
	Adjusted mean	61.77	53.64	
	Unadjusted standard deviation	21.98	25.07	
	n	37	73	

Table of contents

Table 12 | Post-test Results for DIBELS Next and SEL: Hispanic students.

Measure Name	Variables	Burst	Control	Significance Test
DORF G3	Unadjusted mean	78.87	71.9	t (87) = 1.12, n.s., effect size = 0.07
	Adjusted mean	76.69	73.04	
	Unadjusted standard deviation	23.46	25.6	
	n	31	59	
DAZE G3	Unadjusted mean	13.61	11.1	t (87) = 1.63, n.s., effect size = 0.12
	Adjusted mean	13.12	11.36	
	Unadjusted standard deviation	6.89	6.55	
	n	31	59	

Table of contents

Table 13 | Post-test Results for DIBELS Next and SEL: English as a second language students.

Measure Name	Variables	Burst	Control	Significance Test
DIBELS Composite Score GK	Unadjusted mean	116.67	118.02	t (327) = 0.01, n.s., effect size = 0.00
	Adjusted mean	117.1	117.05	
	Unadjusted standard deviation	45.93	46.95	
	n	229	101	
DIBELS Composite Score G1	Unadjusted mean	105.38	89.3	t (325) = 2.54, p < 0.05, effect size = 0.21
	Adjusted mean	105.88	88.72	
	Unadjusted standard deviation	75.58	71.65	
	n	176	152	
DIBELS Composite Score G2	Unadjusted mean	159.51	132.08	t (229) = 0.58, n.s., effect size = 0.03
	Adjusted mean	150.36	146.53	
	Unadjusted standard deviation	86.92	83.04	
	n	142	90	
DIBELS Composite Score G3	Unadjusted mean	245.07	236.99	t (238) = –1.46, n.s., effect size = –0.08
	Adjusted mean	237.04	248.91	
	Unadjusted standard deviation	119.19	104.76	
	n	144	97	
STAR Early Literacy Scaled Score GK	Unadjusted mean	620.01	620.82	t (304) = 0.22, n.s., effect size = 0.02
	Adjusted mean	621.07	618.38	
	Unadjusted standard deviation	101.76	105.33	
	n	214	93	
STAR Early Literacy Scaled Score G1	Unadjusted mean	666.65	654.86	t (309) = 1.30, n.s., effect size = 0.11
	Adjusted mean	667.72	653.68	
	Unadjusted standard deviation	110.85	105.93	
	n	164	148	
STAR Early Literacy Scaled Score G2	Unadjusted mean	731.41	708.79	t (211) = 0.44, n.s., effect size = 0.03
	Adjusted mean	724.24	719.67	
	Unadjusted standard deviation	88.97	95.26	
	n	129	85	
STAR Early Literacy Scaled Score G3	Unadjusted mean	784.6	777.16	t (223) = –0.37, n.s., effect size = –0.03
	Adjusted mean	779.97	783.53	
	Unadjusted standard deviation	81.12	80.57	
	n	131	95	

Table of contents

Table 13 | Post-test Results for DIBELS Next and SEL: English as a second language students.

Measure Name	Variables	Burst	Control	Significance Test
LNF GK	Unadjusted mean	43.6	44.5	t (327) = –0.16, n.s., effect size = –0.01
	Adjusted mean	43.77	44.1	
	Unadjusted standard deviation	19.4	16.35	
	n	229	101	
PSF GK	Unadjusted mean	42.95	42.95	t (327) = 0.20, n.s., effect size = 0.02
	Adjusted mean	43.08	42.66	
	Unadjusted standard deviation	17.27	19.26	
	n	229	101	
NWF-CLS GK	Unadjusted mean	30.12	30.57	t (327) = –0.02, n.s., effect size = 0.00
	Adjusted mean	30.25	30.29	
	Unadjusted standard deviation	16.56	18.02	
	n	229	101	
NWF-CLS G1	Unadjusted mean	48.39	46.83	t (325) = 0.88, n.s., effect size = 0.08
	Adjusted mean	48.52	46.68	
	Unadjusted standard deviation	22.82	20.79	
	n	176	152	
NWF-WWR GK	Unadjusted mean	4.1	2.38	t (327) = 3.06, $p < 0.05$, effect size = 0.29
	Adjusted mean	4.16	2.25	
	Unadjusted standard deviation	5.66	5.44	
	n	229	101	
NWF-WWR G1	Unadjusted mean	11.72	10.93	t (325) = 0.99, n.s., effect size = 0.09
	Adjusted mean	11.77	10.87	
	Unadjusted standard deviation	9.71	8.8	
	n	176	152	
DORF G1	Unadjusted mean	33.61	26.14	t (325) = 3.76, $p < 0.05$, effect size = 0.32
	Adjusted mean	33.74	25.99	
	Unadjusted standard deviation	22.88	20.43	
	n	176	152	
DORF G2	Unadjusted mean	56.94	46.37	t (229) = 1.81, n.s., effect size = 0.10
	Adjusted mean	54.24	50.62	
	Unadjusted standard deviation	26.08	24.13	
	n	142	90	

Table 13 | Post-test Results for DIBELS Next and SEL: English as a second language students.

Measure Name	Variables	Burst	Control	Significance Test
DORF G3	Unadjusted mean	69.39	69.32	t (238) = −2.25, p < 0.05, effect size = −0.13
	Adjusted mean	67.37	72.31	
	Unadjusted standard deviation	30.27	27.46	
	n	144	97	
DAZE G3	Unadjusted mean	12.4	10.23	t (238) = 1.91, n.s., effect size = 0.15
	Adjusted mean	12.05	10.74	
	Unadjusted standard deviation	7.14	5.72	
	n	144	97	

Table of contents

Table 14 | Post-test Results for DIBELS Next and SEL: Students receiving subsidized lunch.

Measure Name	Variables	Burst	Control	Significance Test
DIBELS Composite Score GK	Unadjusted mean	119.19	113.49	t (676) = 1.91, n.s., effect size = 0.18
	Adjusted mean	119.51	113.12	
	Unadjusted standard deviation	48.01	44.35	
	n	360	319	
DIBELS Composite Score G1	Unadjusted mean	102.17	91.99	t (650) = 2.61, p < 0.05, effect size = 0.23
	Adjusted mean	103.47	90.75	
	Unadjusted standard deviation	74.05	69.04	
	n	317	336	
DIBELS Composite Score G2	Unadjusted mean	159.45	143.53	t (543) = 2.55, p < 0.05, effect size = 0.17
	Adjusted mean	158.14	145.19	
	Unadjusted standard deviation	87.81	87.06	
	n	305	241	
DIBELS Composite Score G3	Unadjusted mean	250.01	262.68	t (441) = –1.75, n.s., effect size = –0.10
	Adjusted mean	250.77	261.72	
	Unadjusted standard deviation	114.08	104.6	
	n	247	197	
STAR Early Literacy Scaled Score GK	Unadjusted mean	614.3	606.9	t (635) = 1.16, n.s., effect size = 0.11
	Adjusted mean	615.04	606.03	
	Unadjusted standard deviation	98.21	105.2	
	n	344	294	
STAR Early Literacy Scaled Score G1	Unadjusted mean	686.16	673.95	t (613) = 1.99, p < 0.05, effect size = 0.18
	Adjusted mean	687.55	672.6	
	Unadjusted standard deviation	103.65	104.74	
	n	304	312	
STAR Early Literacy Scaled Score G2	Unadjusted mean	737.99	731.95	t (514) = 0.97, n.s., effect size = 0.08
	Adjusted mean	738.29	731.58	
	Unadjusted standard deviation	91.47	86.6	
	n	288	229	
STAR Early Literacy Scaled Score G3	Unadjusted mean	789	789.45	t (425) = –0.30, n.s., effect size = –0.03
	Adjusted mean	788.25	790.27	
	Unadjusted standard deviation	78.1	76.64	
	n	223	205	

Table 14 | Post-test Results for DIBELS Next and SEL: Students receiving subsidized lunch.

Measure Name	Variables	Burst	Control	Significance Test
LNF GK	Unadjusted mean	43.84	42.15	t (676) = 1.49, n.s., effect size = 0.14
	Adjusted mean	43.96	42.02	
	Unadjusted standard deviation	18.86	16.88	
	n	360	319	
PSF GK	Unadjusted mean	44.2	43.01	t (676) = 1.00, n.s., effect size = 0.10
	Adjusted mean	44.31	42.88	
	Unadjusted standard deviation	19.3	19.29	
	n	360	319	
NWF-CLS GK	Unadjusted mean	31.15	28.33	t (676) = 2.51, p < 0.05, effect size = 0.24
	Adjusted mean	31.24	28.22	
	Unadjusted standard deviation	16.96	15.55	
	n	360	319	
NWF-CLS G1	Unadjusted mean	49.19	47.67	t (650) = 1.39, n.s., effect size = 0.13
	Adjusted mean	49.54	47.34	
	Unadjusted standard deviation	23.1	21.69	
	n	317	336	
NWF-WWR GK	Unadjusted mean	3.42	1.64	t (676) = 4.91, p < 0.05, effect size = 0.47
	Adjusted mean	3.45	1.61	
	Unadjusted standard deviation	5.82	3.96	
	n	360	319	
NWF-WWR G1	Unadjusted mean	11.88	10.57	t (650) = 2.33, p < 0.05, effect size = 0.21
	Adjusted mean	12.02	10.44	
	Unadjusted standard deviation	9.55	9.19	
	n	317	336	
DORF G1	Unadjusted mean	32.47	28.62	t (650) = 3.09, p < 0.05, effect size = 0.27
	Adjusted mean	32.86	28.25	
	Unadjusted standard deviation	22.63	20.9	
	n	317	336	
DORF G2	Unadjusted mean	55.62	50.32	t (543) = 3.21, p < 0.05, effect size = 0.21
	Adjusted mean	55.24	50.8	
	Unadjusted standard deviation	24.94	24	
	n	305	241	

Table of contents

Table 14 | Post-test Results for DIBELS Next and SEL: Students receiving subsidized lunch.

Measure Name	Variables	Burst	Control	Significance Test
DORF G3	Unadjusted mean	70.92	74.54	t (441) = –1.99, p < 0.05, effect size = –0.12
	Adjusted mean	71.12	74.3	
	Unadjusted standard deviation	28.33	27.98	
	n	247	197	
DAZE G3	Unadjusted mean	12.9	13.16	t (441) = –0.33, n.s., effect size = –0.03
	Adjusted mean	12.94	13.11	
	Unadjusted standard deviation	7.27	7.17	
	n	247	197	

Notes

Notes

To learn more about mCLASS Intervention or speak to a product expert visit **amplify.com**.

Amplify.