NCEE 2013-4001

U. S. DEPARTMENT OF EDUCATION

Evaluation of the Content Literacy Continuum: Report on Program Impacts, Program Fidelity, and Contrast

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

Evaluation of the Content Literacy Continuum: Report on Program Impacts, Program Fidelity, and Contrast

Final Report

December 2012

Authors:
William Corrin
MDRC

James J. Lindsay
American Institutes for Research

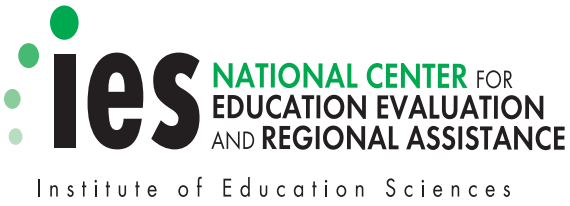
Marie-Andrée Somers and Nathan E. Myers
MDRC

Coby V. Meyers and Christopher A. Condon
American Institutes for Research

Janell K. Smith
MDRC

Project Officer:
Sandra Garcia
Institute of Education Sciences

NCEE2013–4001
U.S. Department of Education



U.S. Department of Education

Arne Duncan

Secretary

Institute of Education Sciences

John Q. Easton

Director

National Center for Education Evaluation and Regional Assistance

Ruth Curran Neild

Commissioner

December 2012

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract ED-06-C0-0019 with Regional Educational Laboratory Midwest administered by Learning Point Associates.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the report.

This report is available on the Institute of Education Sciences website at http://ncee.ed.gov and the Regional Educational Laboratory Program website at http://edlabs.ed.gov.

Alternate Formats Upon request, this report is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

Disclosure of Potential Conflict of Interest

The research team for this study consisted of Regional Educational Laboratory Midwest administered by Learning Point Associates (prime contractor) and two subcontractors, MDRC and Survey Research Management. None of these organizations or their key staff members have financial interests that could be affected by the findings of this study. No one on the Technical Working Group, convened annually by the research team to provide advice and guidance, has financial interests that could be affected by the study findings.

Acknowledgements

This study represents a collaborative effort among the authors, the staff from the participating school districts and schools, the program developers, and our colleagues at Learning Point Associates, MDRC, and Survey Research Management.

Appreciation goes to Action Designs and University of Kansas Center for Research on Learning for collaborating on this project, implementing the intervention within participating high schools and for sharing program-related details and implementation data with the evaluation staff. In particular, the authors would like to thank Don Deshler, Keith Lenz, Suzanne Robinson, Nancy Sander, Susan D'Aniello, Pam Leitzell, Jerri Neduchal, Peggy Reyes, and Sue Woodruff.

Members of REL Midwest's technical working group provided valuable feedback during the design, implementation, and reporting phases of this project. Members of this technical working group included Ellen Goldring, James Heckman, Larry Hedges, Rebecca Maynard (prior to 2009), and Brian Rowan. Russ Whitehurst joined the REL Midwest technical working group in 2010. Ellen Kisker also provided input on technical issues related to this evaluation.

At Learning Point Associates / American Institutes for Research we wish to thank Steve Cantrell and Matt Dawson for initiating this project and Chris Brandt, Julia Thomas, and Carrie Scholz for their site recruitment efforts. Danielle Carnahan, Charlene Cobb, Chris Rauscher, and Rebecca Bates developed the classroom observation protocol and trained site visitors in its use. Shaheen Khan and Jiuping Chen developed the database for the observation protocol. Thanks go to Sheila Rodriguez, Denise Powell, and Winsome Waite for coordinating site visits with participating schools. Appreciation also is extended to the 16 site visitors who conducted the classroom observations.

We have appreciated the contributions of various MDRC staff members during the life of this project. Jim Kemple played a key role in leading this project at its inception. Elizabeth Nelson helped prepare IRB and OMB materials, as well as worked with Shelley Rappaport on the development of the earliest iterations of the observation and interview protocols. Saskia Levy and Adam Wodka helped us recruit districts and schools into the study. Alix Barasch and Asya Magazinnik contributed to the collection and analysis of school records data, and also worked with the schools and districts to obtain master schedule information and created the classroom observation schedules for our site visit teams. Zeest Haider also assisted with school records data collection and analysis, as well as the processing and analysis of the GRADE reading assessment data. Susan Sepanik, Michael Weiss, and Pei Zhu all provided valuable technical advice to the project team regarding our analyses.

Survey Research Management administered the reading assessments to students within participating schools and scored the answer forms. We wish to thank Linda Kuhn for managing this aspect of the project and Tony Lavender for preparing the electronic test score data files. We also thank Ashley Bronzan, Michelle Michalski, Rachel Mudd, and Rebecca Richardson for coordinating the efforts of the assessment teams.

Finally, we would like to extend our appreciation to reviewers of this and previous project reports: Terry Salinger at American Institutes for Research, Jennifer Hamilton at Westat, and Fred Doolittle at MDRC.

Contents

Figures

Tables

Box

Executive Summary

Large numbers of adolescents enter high school lacking the necessary reading skills to be academically successful. The demand for strong reading and writing skills increases as students get promoted to high school grades. Not only do high school teachers rely more heavily on textbooks to convey critical course content to students, but the content in those textbooks also gets more challenging (Heller & Greenleaf, 2007). Moreover, by grade 9, the reading standards students are expected to meet also increase in difficulty. High school students are expected not only to remember facts but also to induce themes, processes, and concepts from material and relate those "higher order concepts" to new content (Biancarosa & Snow, 2006). Adolescent students are expected to read and to produce complex texts, whose structures and modes of presenting information vary according to genres and content (ACT, 2009). The Common Core State Standards (Common Core State Standards Initiative, 2010), which have been adopted by almost 75 percent of the states, emphasize this variation by differentiating "college and career readiness" standards in grades 6–12 according to the reading and writing skills needed in history/social studies, science, and technology subjects.

School district leaders and high school administrators not only face the challenge of providing students with additional instruction focused on improving reading skills, but they also must simultaneously help students master necessary subject area content for which they are held accountable. These leaders need information on interventions that can be integrated within the high school curriculum to help struggling adolescent readers acquire the strategies necessary to read at proficient levels. This report presents the findings of a rigorous experimental impact evaluation and implementation study of one such intervention, the Content Literacy Continuum (CLC), developed by researchers at the University of Kansas Center for Research on Learning. This evaluation of CLC was conducted by three partnering organizations: REL Midwest, MDRC, and Survey Research Management. Thirty-three high schools in nine districts across four Midwestern states agreed to participate in this evaluation, and 28 of those 33 schools continued their participation throughout the entire study period. Full implementation of this intervention began in the 2008/09 school year and continued through the 2009/10 school year.

Given that CLC was designed to address the dual needs of high schools to support both the literacy and content learning of students, the evaluation focused on program impacts on reading comprehension test scores and students' accumulation of course credits in core content areas.[1] To assess the impacts of CLC on these outcomes, the study team conducted a cluster randomized trial. That is, participating high schools within each district were randomly assigned either to implement CLC (CLC schools) or to continue with "business as usual" (non-CLC schools). Impacts were estimated by analyzing the outcomes of students at the CLC schools compared with those at the non-CLC schools. The evaluation's primary research questions focused on the impact of CLC on students' reading comprehension and course performance at the end of the second year of implementation. Secondary research questions compared the first-year impacts and second-year impacts and also investigated program impacts on other student outcomes.

[1] The four core content areas are English language arts, history/social studies, mathematics, and science. Core credit accumulation was measured as credits earned in those four subject areas as a percentage of the total credits needed to graduate.

In addition, the evaluation examined the implementation of the CLC framework within the CLC schools. This report presents findings regarding the degree to which schools assigned to implement CLC set up the necessary structures and organizational processes needed to support implementation of CLC (referred to as *structural* fidelity in this report) and the degree to which the pedagogical practices emphasized in CLC-related professional development were apparent within the instruction of core content teachers in participating schools (referred to as *instructional* fidelity). The structures and instruction at CLC schools and non-CLC schools also were compared to provide information about the contrast that CLC implementation provided compared with business as usual.

The key impact and implementation findings discussed in this report are as follows:

Findings from primary impact analyses

- There were no statistically significant differences in reading comprehension scores between CLC schools and non-CLC schools (effect size = 0.06 standard deviations for grade 9 students and 0.10 standard deviations for grade 10 students). Therefore, it cannot be concluded that the CLC framework improved students' reading comprehension scores in the second year of the study, in either grade level.
- Nor did CLC have a statistically significant impact on the students' accumulation of core credits (as a percentage of the total needed to graduate) in the second year, in either grade level. The impact estimate was negative for grade 9 students and positive for grade 10 students (effect sizes = −0.17 and 0.02 standard deviations, respectively).

Findings from secondary impact analyses

- The estimated impacts of CLC on the primary outcomes, reading comprehension test scores and core credit accumulation, in the first year do not differ statistically from the second-year impacts. CLC did not have a statistically significant impact on students' reading comprehension in the first year (effect size = 0.13 standard deviation). CLC also did not have a statistically significant impact on students' credit earning in core content classes (effect size = −0.04 standard deviation).
- In terms of secondary outcomes, the CLC program had a statistically significant, positive impact on grade 9 students' reading vocabulary in the first year of the study. However, it did not affect their grade point average (GPA) by a statistically significant amount, in either grade level or in either study year.
- The analyses that examined whether the intervention had stronger effects for some subgroups of students (for example, groups defined by grade 8 reading proficiency, being overage for grade, or eligibility for special education services) suggest similar results for the various groups. Nor do the data suggest that CLC was more effective in some school districts than in *others.*

Implementation findings

- Of the 28 schools that participated in the evaluation for two years, 15 had been randomly assigned to implement the CLC framework. It was rare, however, for the CLC schools to establish all the structural components necessary for CLC. In the first year, 11 of these 15 schools implemented five or fewer of nine structural components at an adequate level or better. Implementation of these components was somewhat less successful in the second

year, as all 15 schools implemented five or fewer of these components at an adequate level or better.

- The percentages of observed core content teachers who explicitly used CLC-specific content enhancement routines or learning strategies were 22 percent during year 1 and 11 percent during year 2. Although this percentage is less than intended by the program developer, in both year 1 and year 2, the rate was double that of the rate of use among core content teachers in non-CLC schools.
- Observations of instruction of core content teachers in CLC schools during year 1 indicated that one of the three pedagogical practices emphasized during CLC professional development was included in instruction at a level considered "adequate" by the program developers. For year 2, these scores averaged across all CLC schools were all below the program developer's cut point for "adequate."
- Use of CLC-specific content enhancement routines and strategies among teachers of the CLC-specific Fusion Reading course within CLC schools was observed to be 62 percent for both years of implementation.

The intervention: the Content Literacy Continuum

The Content Literacy Continuum (CLC) combines whole-school and targeted approaches to supporting student literacy and content learning. The intervention combines instructional routines and learning strategies that have been developed and tested by the University of Kansas Center for Research on Learning (KU-CRL). First, within the CLC framework, student weaknesses in literacy skills across all core subject areas are addressed by training core content teachers to use instructional routines and to model learning strategies that may help students of varying skill levels better comprehend critical content provided through instruction and text.[2] Second, CLC is designed to offer targeted reading support to struggling adolescent readers by training reading teachers to provide these students with intensive reading instruction. KU-CRL developed a distinct curriculum—Fusion Reading—for these struggling readers who have foundational decoding, fluency, and comprehension skills (that is, students reading at least at a grade 4 level) but who are sufficiently below grade level to warrant more intensive support (Hock, Brasseur, & Deshler, 2008).

The CLC framework is organized into levels of support for students. The three levels of CLC that were implemented as part of this project are as follows:

Level 1: content enhancement. Teachers are trained to follow instructional routines and use instructional tools that help them align instruction with standards and curricula and help students develop skills such as organizing content, drawing connections among topics, and learning essential vocabulary.

[2] Literacy skills are defined here as reading and writing skills. Although the levels of CLC evaluated in this study incorporate activities designed to improve both reading and writing, this evaluation focuses on program impacts on students' reading achievement.

Level 2: embedded strategy instruction. Teachers are trained to embed content-related learning strategies into their instruction, intended for students to learn and apply with the goal of improving their ability to master subject area content.

Level 3: intensive strategy instruction. Fusion Reading classes provide more intensive strategy instruction targeted toward students who are reading two to five years below grade level.

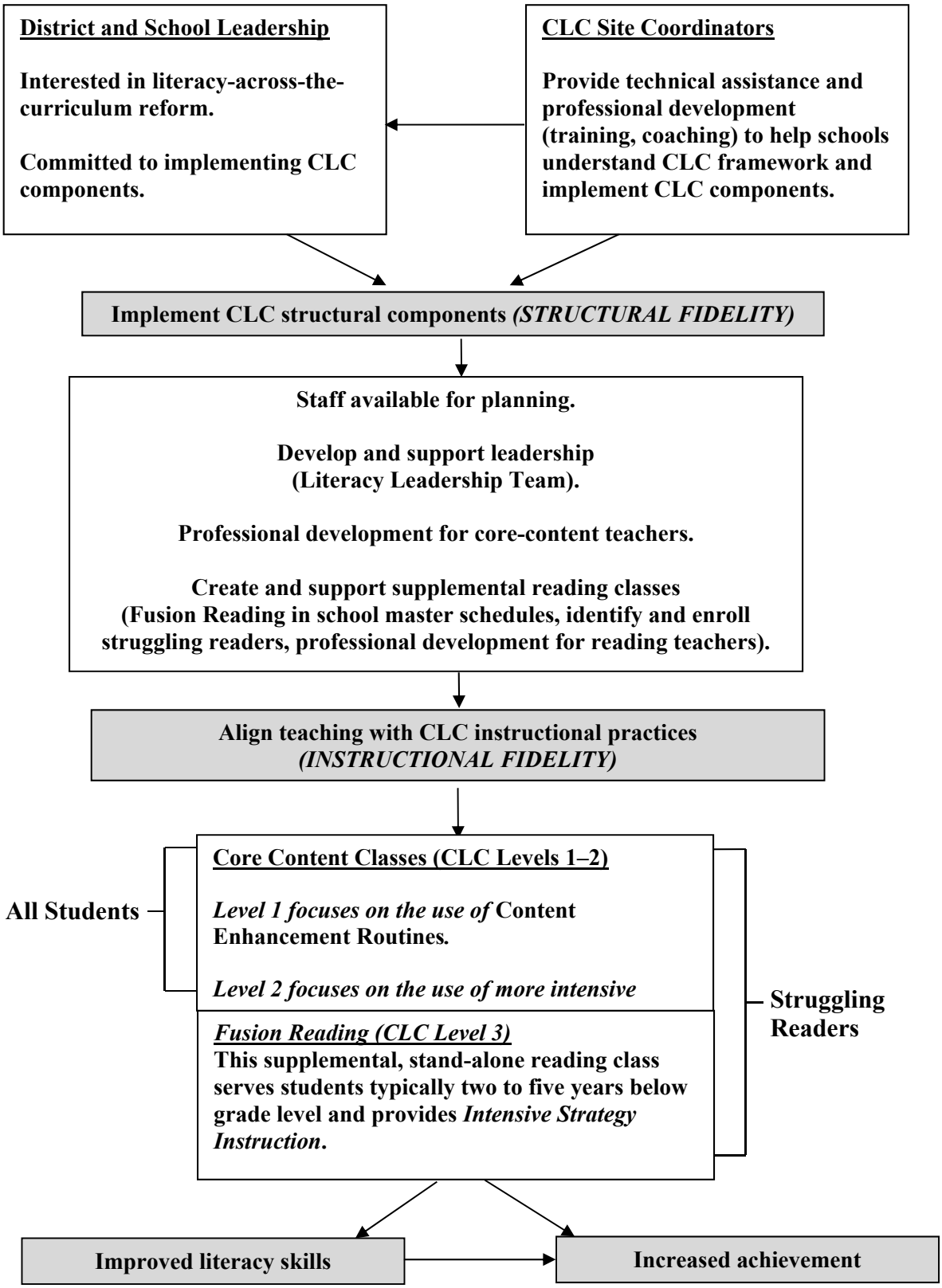
CLC can be viewed as high-school-level analog to tiered response-to-intervention (RTI) reading instruction frameworks used in elementary schools. RTI frameworks often are organized in three tiers of instructional support. The first tier focuses on support for all students—schoolwide or classroom-wide. The second tier focuses on smaller, targeted groups of students who require more intensive support than the instructional practices used more broadly for all students. The third tier provides even more intensive support, very often in the form of special education services, to students who are struggling even after receiving first- and second-tier support (Gersten et al., 2008).

The leveled CLC model also is intended to provide greater academic support for students with greater need. More specifically, CLC levels 1 and 2 correspond to the first tier of an RTI model: they are implemented in all core content classes within a school in support of all students. Level 3 of CLC is analogous to the second tier of response to intervention in that it targets support to struggling readers by supplementing that received in core content classes with support in Fusion Reading classes.[3]

Integration of the CLC framework throughout a high school involves professional development of core content and reading teachers by KU-CRL–trained implementation specialists (referred to later as "site coordinators"). These specialists also work with school leaders to establish organizational structures and processes within schools to coordinate the components of CLC. These components include establishing a team made up of faculty from all core content areas who can guide the rollout of CLC professional development, establishing Fusion Reading classes, enrolling students who are two to five years below grade level in reading into those classes, and setting aside enough time for the professional development of teachers on CLC's content enhancement routines and embedded learning strategies. According to the logic model (figure 1), establishing these structural components creates conditions that facilitate the instructional components of the program—the implementation in classrooms of the CLC instructional routines and the learning strategies with students. The logic model also theorizes that students' exposure to the routines and strategies throughout the school day *should* lead to improved literacy skills and improved general achievement.

[3] The levels of CLC that are not part of this study (levels 4 and 5) focus on providing special-needs students with even more support. To continue the RTI analogy, these students would be similar to the students in the third tier of an RTI framework.

Figure 1. CLC theory of action



The evaluation of CLC

This evaluation focused on both the implementation and impacts of the Content Literacy Continuum (CLC) intervention. The implementation study examined how successful the schools were in putting the framework in place and the contrast this implementation created between the CLC and non-CLC schools in their structures and practices. The impact study investigated the effectiveness of CLC in changing the academic outcomes of students.

Site recruitment, random assignment, and the study sample

High schools within states served by REL Midwest (Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin)—where at least one-third of students scored below proficient on state standardized reading or English language arts assessment, at least one-fourth of students were eligible for free or reduced-price lunch, and there were at least 100 grade 9 students—were recruited for this evaluation project. In addition, potential study schools had to be interested in supporting the literacy needs of their students but not already implementing a tiered, whole-school approach similar to CLC.

Schools and districts meeting these criteria were approached by the project team with the opportunity to participate in the CLC evaluation project. The program developer and the study team conducted phone conferences and in-person informational sessions with the districts and schools that expressed interest in the study. These recruitment efforts yielded 33 high schools in nine school districts within four states that met the study criteria and agreed to participate in this evaluation project.

Within their school districts, participating schools were randomly assigned to implement the CLC intervention (CLC schools) or continue with "business-as-usual" school programming (non-CLC schools).[4] Random assignment resulted in 17 CLC schools assigned to implement CLC and 16 non-CLC schools. Twenty-eight of the 33 participating high schools continued their participation in the evaluation throughout the entire study period (reasons for discontinuing were school closure and fear of conflict with state-mandated changes resulting from state sanctions). For this final sample of 28 schools, 15 were CLC schools, and 13 were non-CLC schools.[5]

The CLC high schools began full implementation of CLC during the summer leading into the 2008/09 school year.[6] Study participation continued through the end of the 2009/10 school year. During the first full year of implementation, professional development focused on teachers of grade 9 students. Year 2 of implementation focused on teachers of students in grades 9 and 10.

[4] Rural school districts with only a single high school were combined into a "consortium." These schools were randomly assigned as a single "block."

[5] This sample of 28 high schools is the primary analysis sample for the findings presented in this report. Analyses of baseline data found that the CLC and non-CLC schools were similar in terms of both school-level and student-level characteristics.

[6] Three CLC schools were able to begin implementation planning activities during the 2007/08 school year.

Research questions

This evaluation was designed to address two research questions:

- What are the impacts of the CLC program on grade 9 students' reading comprehension and accumulation of core credits in 2009/10, the second year of the study?
- What are the impacts of the CLC program on these outcomes for grade 10 students in the second year of the study?

Secondary research questions included the following:

- What are the impacts of the CLC program on grade 9 students' reading comprehension and credit earning in the first year of the study? Are these impacts different from those of the second year of the study?
- What are the impacts of the CLC program on other academic outcomes, such as vocabulary test scores and grade point average?
- Is the impact of the CLC program on grade 9 and 10 students' reading comprehension and credit earning in the second year of the study greater for some subgroups of students than for others? The study team examined impacts for subgroups of students defined by scoring above or below reading proficiency cut-offs in eighth grade, being overage for their grade level or not, and being classified as eligible for special education services or not at the start of ninth grade.

This evaluation of CLC is designed not only to address these research questions but also to examine and report on the implementation of the CLC framework. The study examined the degree to which implementing schools established the structures needed to support CLC and the degree to which teachers within CLC schools incorporated the content enhancement routines, learning strategies, and CLC-emphasized pedagogical practices within their instruction. The study also examined the extent to which implementing the CLC framework resulted in a contrast in structural components and instructional practices between the CLC schools and the non-CLC schools that continued with business as usual.

Data collection

To measure reading comprehension and vocabulary, test scores were obtained by administering the Group Reading Assessment and Diagnostic Evaluation (GRADE) to grade 9 students at the end of year 1 and to grade 9 and 10 students at the end of year 2. In addition, student transcript data were obtained from participating school districts at the end of each year of implementation. These data provided information about students' course performance, in particular their accumulation of credits in core subject areas. The districts also provided historical student records from which the study team obtained baseline data on student characteristics to describe the sample and include as covariates in the impact analyses.

To examine CLC implementation, three other types of data were collected. First, the program developer's implementation staff shared their reports from their monthly visits to the CLC schools with the study team. Second, interviews were conducted with district administrators and administrators from participating high schools (both CLC and non-CLC) who were most familiar with literacy-related initiatives going on within schools and districts. Third, site visitors observed

the instruction of core content and reading teachers within all participating schools and recorded the presence or absence of pedagogical practices emphasized during professional development sessions for CLC.

Implementation

In terms of program implementation, the study team examined both fidelity (how successful the CLC schools were in implementing CLC to the degree expected by the program developer) and contrast (how different CLC implementation made CLC schools from non-CLC schools).

Fidelity. Fidelity of implementation was examined in two ways. First, the degree to which schools implemented the structural or procedural elements necessary for CLC (referred to in this report as *structural* fidelity) was examined by reviewing monthly site visit reports maintained by the program developer's school implementation staff and by interviewing district and school administrators. Second, data from classroom observations were examined to determine the degree to which CLC-specific content enhancement routines and learning strategies as well as pedagogical practices (sequenced instruction, multiple instructional modalities, and interactive and scaffolded instruction[7]) emphasized in CLC professional development were incorporated within the instruction of teachers of core content subjects and reading (referred to in this report as *instructional* fidelity).

Structural fidelity varied across the schools implementing CLC, but, overall, schools were unsuccessful in implementing all the structural and procedural aspects of the programs at the level expected by the program developer. One high school was unable to implement CLC at all because of district-initiated reconstitution during the implementation period. On average, across all 15 schools that comprised the analytic sample, adequate fidelity was achieved on 4 of the 10 structural components during year 1 of the study. In year 2 of the study, CLC schools achieved adequate fidelity on 1 of the 9 structural components.

The incorporation of CLC-specific content enhancement routines and learning strategies into core content teachers' instructional practice was limited. The percentage of observed core content teachers who explicitly used these routines or learning strategies in year 1 was 22 percent. Observed use of CLC-specific routines and strategies among core content teachers in CLC schools was lower in year 2 (11 percent) than in year 1. However, use of CLC-specific content enhancement routines and strategies was higher among teachers of the CLC-specific Fusion Reading course compared with their peers teaching core subject courses. The CLC-specific practices were observed in 62 percent of Fusion classes visited in year 1 and in 61 percent of these classes in year 2. Year 1 observations of core content teachers' instruction averaged across all CLC-implementing schools suggest that one of three CLC-emphasized pedagogical practices was apparent at a level considered adequate by the program developer. Observations of instruction of core content teachers during year 2—again averaged across

[7] These three pedagogical practices are defined as follows. Sequenced instruction: across all levels of the CLC program, teachers are trained to deliver instruction through carefully planned sequences that establish the purpose for instruction, engage students in literacy instruction, and review the content of such instruction. Multiple instructional modalities: across the CLC program, teachers are trained to present information by using a combination of verbal, graphical or visual, and written modalities. Interactive and scaffolded instruction: the CLC program encourages instruction that is highly interactive, in which teachers and students actively co-construct knowledge. Instruction is deliberately scaffolded, with instructional activities progressing through several steps: teacher-centered modeling, teacher-mediated practice, student-mediated practice, and independent student practice.

schools—indicated that none of the three CLC-emphasized pedagogical practices was implemented at levels considered adequate by the program developer. Caution is warranted in placing too much confidence in these findings regarding the three pedagogical practices since interrater reliability estimates calculated on pilot observations and observations of film clips were low.[8]

Contrast. As with implementation fidelity, the contrast between CLC schools and non-CLC schools was framed in terms of structure and instruction. In regard to structural support for CLC, the most notable contrast between CLC and non-CLC schools was that CLC schools more often sought to address students' literacy needs through literacy-across-the-curriculum support and through the combination of such support with the provision of supplemental reading classes. In end-of-year interviews, administrators at CLC schools were significantly more likely than their peers at non-CLC schools to make those claims about both years of implementation. In the interviews at the end of the second year of implementation, the CLC administrators also were more likely to indicate that professional development that year had been of a singular focus on student literacy. Otherwise, however, CLC and non-CLC administrators responded similarly to other questions regarding structures and processes important to the implementation of the CLC framework: the continuity and sustainability of professional development; the alignment of school improvement efforts; and the use of data to inform decisions about students, instruction, and school priorities. In terms of instruction, CLC-specific instructional routines and strategies were observed in use twice as often in classrooms in CLC schools as in classrooms in non-CLC schools in both years of the study.

Program impact

Impacts were analyzed using two-level hierarchical linear models. Along with schools' random assignment blocks, several student background characteristics were included in these models as covariates: (1) whether students were overage for grade at the beginning of grade 9, (2) prior achievement (grade 8 state test scores in reading and math), (3) educational indicators (English language learner and special education status), (4) socioeconomic factors (free or reduced-price lunch eligibility), and (5) demographic factors (race/ethnicity and gender). Impact estimates were based on intent-to-treat analyses (schools' status based on random assignment rather than whether CLC was actually implemented within the schools).

Primary impacts. Impacts on the two primary outcome measures after two years of implementation are presented in table 1, and the findings are discussed below:

- There were no statistically significant differences in reading comprehension scores between CLC schools and non-CLC schools (effect size = 0.06 standard deviation for grade 9 students and 0.10 standard deviation for grade 10 students). Therefore, it cannot be concluded that the CLC framework improved students' reading comprehension scores in the second year of the study, in either grade level.
- Nor did CLC have a statistically significant impact on students' accumulation of course credits needed for graduation in core subject areas in the second year, in either grade level.

[8] Average percent agreement of observers for the three pedagogical constructs was 42 percent during year 1 and 40 percent during year 2.

The impact estimate was negative for grade 9 students and positive for grade 10 students (effect sizes = −0.17 and 0.02 standard deviations, respectively).

Secondary impact questions. Additional analyses of the impact of CLC on these two outcomes were conducted to investigate whether there was variation in impacts by implementation year, student subgroups, or school district. In addition, impacts on secondary outcome measures we're examined.

First year impacts and comparing impacts by year of implementation. The estimated impacts of CLC on the primary outcomes, reading comprehension test scores and core credit accumulation, in the first year do not differ statistically from the second-year impacts. CLC did not have a statistically significant impact on students' reading comprehension in the first year (effect size = 0.13 standard deviation). CLC also did not have a statistically significant impact on students' credit earning in core content classes (effect size = −0.04 standard deviation).

Impacts on secondary outcome measures. Although reading comprehension test scores and core credit earning were identified as primary outcomes for this evaluation, the study team also analyzed the impact of CLC on secondary indicators of reading achievement and academic performance in core courses—students' test scores on the vocabulary section of the GRADE and their grade point averages (GPA).

- The CLC program had a statistically significant, positive impact on grade 9 students' reading vocabulary in the first year of the study. However, it did not affect their GPA by a statistically significant amount, in either grade level or in either study year.

Variation in impacts. The study did not find conclusive evidence that CLC was more or less effective for any particular student subgroup or in any individual school districts on the primary outcomes.

Taken together, the implementation and impact findings indicate that the CLC framework, a comprehensive instructional reform, was not adequately implemented in the CLC schools in two years, and that in general the intervention did not have an impact on student outcomes. These findings from this evaluation resonate with the conclusions drawn by Borman, Hewes, and Overman (2003) based on their meta-analysis of research on other comprehensive school reform efforts. In their report, Borman and colleagues summarized the research as indicating that it is difficult to have an impact on student outcomes in the early years of a comprehensive school reform effort, even among students who are struggling the most and may stand to benefit the most and even in the most motivated schools and districts. Based on the meta-analytic findings, Borman et al. (2003) conclude that comprehensive school reform efforts often require four or more years to make a notable difference.

Table 1. Impacts on reading comprehension and credit accumulation, year 2

Outcome	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Impacts on reading comprehension, GRADE respondent sample[a]					
Grade 9 sample					
Average standard score	91.6	90.7	0.9	0.06	0.262
Corresponding grade equivalent	*6.5*	*6.2*			
Corresponding percentile	*28*	*25*			
Sample size					
Students (total = 5,011)	2,975	2,036			
Schools (total = 28)	15	13			
Grade 10 sample					
Average standard score	96.7	95.1	1.6	0.10	0.203
Corresponding grade equivalent	*7.7*	*7.2*			
Corresponding percentile	*40*	*35*			
Sample size					
Students (total = 4,546)	2,908	1,638			
Schools (total = 28)	15	13			
Impacts on credit accumulation, school records sample[b]					
Grade 9 sample					
Credits earned in core subject areas (%)	22.2	23.8	–1.6	–0.17	0.058
Sample size					
Students (total = 7,951)	4,467	3,484			
Schools (total = 28)	15	13			
Grade 10 sample					
Credits earned in core subject areas (%)	41.3	40.9	0.4	0.02	0.726
Sample size					
Students (total = 8,514)	4,888	3,626			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note. This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, English as a second language (ESL) status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

b. The cumulative number of credits earned is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study and from the GRADE assessment administered as part of the CLC study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Chapter 1: Introduction

Findings from the National Assessment of Education Progress (NAEP) in reading indicate that 69 percent of grade 8 students lack the reading skills necessary to perform grade-level work (National Center for Education Statistics, 2009).[9] This standardized assessment is administered to representative samples of students in the United States. In 2009, the percentages of grade 8 students within states served by the Midwest Regional Education Laboratory (Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin) who scored below the proficient level on the assessment, ranged from 62 percent to 69 percent (table 1.1). NAEP 2009 state averages for states within the Midwest region for grade 8 reading scores ranged from 262 to 270, all below the 281 proficiency cut-score, as was the national average of 264 (National Center for Education Statistics, 2009). In addition, two states in the region were among 11 states nationally that participated in the 2009 NAEP pilot in reading for grade 12 students. Findings from that assessment indicate that 60 percent of Illinois students and 61 percent of Iowa students scored below the proficient level compared with a national average of 63 percent (National Center for Education Statistics, 2010).

Table 1.1. Percentage of adolescent students in REL Midwest states failing to meet proficiency levels in reading according to NAEP standards

State	2009 NAEP (grade 8) (percent)
Illinois	67
Indiana	68
Iowa	68
Michigan	69
Minnesota	62
Ohio	64
Wisconsin	66

The demand for strong reading and writing skills increases as students get promoted to high school grades. Not only do high school teachers rely more heavily on textbooks to convey critical course content to students, but the content in those textbooks also gets more challenging (Heller & Greenleaf, 2007). Moreover, by grade 9, the standards that students are expected to meet also increase in difficulty. High school students are expected not only to remember facts but also to induce themes, processes, and concepts from material and relate those "higher order

[9] Every two to three years, the National Center for Education Statistics administers common reading assessments (the National Assessment of Educational Progress, or NAEP) to representative samples of students in grades 4 and 8 in each of the 50 states, the District of Columbia, territories, and schools run by the U.S. Department of Defense. Students' performance on the assessment is categorized as "basic," "proficient," and "advanced." Students whose scores fall in the proficient or advanced categories are considered reading at a level necessary to perform grade-level work, while students performing at the "basic" level are considered as having partial mastery of the reading skills necessary at their grade level (see http://nationsreportcard.gov/glossary.asp for more information).

concepts" to new content (Biancarosa & Snow, 2006). Adolescent students are expected to read and produce complex texts, whose structures and modes of presenting information vary according to genre and content (ACT, 2009). The Common Core State Standards (Common Core State Standards Initiative, 2010), which have been adopted by more than 75 percent of the states, emphasize this increased challenge and subject area differentiation in literacy needs by specifying "college and career readiness" standards for grades 6–12 inclusive of reading and writing skills needed in history/social studies, science, and technology subjects.

The large numbers of adolescents who enter high school lacking the necessary reading skills place leaders in high schools and local education agencies in a quandary: how can high schools provide students with additional instruction on reading strategies while also providing the necessary content for which they are accountable? These leaders need information on interventions that can be integrated within the high school curriculum to help struggling adolescent readers acquire the strategies necessary to read at proficient levels.

Various targeted and whole-school interventions have been developed with the aim of improving the reading skills of adolescent students who are reading below grade level. Some of these interventions either have been subjected to rigorous efficacy studies or are presently undergoing such studies. This report presents the findings of a rigorous experimental evaluation of one such intervention, the Content Literacy Continuum (CLC), developed by researchers at the University of Kansas Center for Research on Learning (KU-CRL).

This introductory chapter begins with a brief review of recent evidence on the efficacy of interventions that are either targeted toward struggling adolescent readers or provide literacy supports schoolwide (that is, across content areas). The chapter then describes CLC, the specific intervention under evaluation, and provides an orientation to the study itself. The chapter concludes by presenting the overall structure of this report.

Recent findings from studies of interventions targeting struggling adolescent readers

Interventions designed to improve reading skills of adolescents have been developed, publicized, and subjected to rigorous evaluation. Some of the approaches to addressing the literacy needs of adolescent students take the form of *targeted interventions* (programs or approaches that provide reading strategy instruction and practice to "struggling readers" or students who are below grade level in reading). Other approaches represent *whole-school interventions* (programs or approaches that integrate literacy strategy instruction and practice within multiple courses or content areas for the purpose of increasing reading performance of all students). Furthermore, other interventions take a *tiered* or *hybrid* approach by providing differing amounts of strategy instruction to students at different levels of reading ability. A brief review of some of these interventions is provided in the following sections.

Research on targeted interventions for adolescent literacy

One recently published report examined the efficacy of two interventions designed to provide additional instruction on reading strategies to struggling readers (Somers et al., 2010). The two interventions examined in this study—a rigorous student-level randomized control trial—are

stand-alone courses for struggling adolescent readers taught by teachers trained on each intervention's curriculum. Findings within this report show that these interventions produce modest and limited impacts on reading-related skills of adolescent students.[10] Statistically significant impacts of 0.09 and 0.06 standard deviation units were found on reading comprehension test scores (compared with a "business as usual" control group taking a variety of typical high school elective courses in areas such as visual and performing arts, career and technical education, and health) and state achievement tests on English language arts, respectively. However, no impact was found on students' vocabulary test scores. Despite these impacts on students' reading comprehension, the study found that 77 percent of the students who participated in the reading classes were still more than two years below grade level in reading. Moreover, impacts that students exhibited after one year of involvement in the intervention disappeared during the following year when they were no longer enrolled in the supplemental reading class.

Findings from an additional nine randomized control trials of interventions targeted at struggling readers at the middle school and high school levels have been published recently. These studies, funded through the U.S. Department of Education's Office of Elementary and Secondary Education Striving Readers Grants, provide findings based on four years of program implementation (U.S. Department of Education Office of Elementary and Secondary Education, 2010). The targeted interventions, developers, study settings, and impacts are summarized in table 1.2. To date, the evaluation findings on these interventions range from −0.06 to +0.29 standard deviations, with 16 of the 18 published impact estimates falling below 0.20 standard deviations.

Research on whole-school interventions for adolescent literacy

An alternative approach to addressing deficits in adolescents' literacy skills involves incorporating reading strategy instruction within core content classes (for example, English language arts, mathematics, science, social studies) throughout the school day (Biancarosa & Snow, 2006; Billmeyer, 1996; Greenleaf & Schoenbach, 2004; Heller & Greenleaf, 2007; Kamil et al., 2008; National Association of Secondary School Principals, 2005). These whole-school approaches generally involve providing core content teachers in middle schools and high schools with professional development on reading strategies and methods for integrating those strategies within their content instruction.

[10] The study also reported positive and statistically significant impacts on students' grade point averages and credits earned across the four core content areas of English language arts, mathematics, social studies, and science. These impacts were significant only for the grade 9 students in their grade 9 year, which was the year they were enrolled in the supplemental reading classes.

Table 1.2. Initial impact estimates for adolescent literacy interventions being examined as part of Striving Readers grants

Intervention	Developer	Setting		Outcome measures	Impacts (standard deviations)
		Grade level	Location		
Chicago Striving Readers (SR) English Language Arts Curriculum and Achieving Maximum Potential (AMP) afterschool program	Chicago Public Schools with researcher from National Louis University	SR replaces ELA for tiers 2 and 3 in grades 6–8[a]; AMP is supplemental for tier 3 in grade 6	Select schools in Illinois	ISAT: vocabulary, reading strategies, comprehension, and literature	Tier 2 students: –0.06 Tier 3 students: 0.00 (3 years exposure to intervention)
Learning Strategies Curriculum	University of Kansas Center for Research on Learning	An elective for students in grades 6 and 9	Select schools in Kentucky	GRADE: vocabulary, comprehension, oral language subtests	Grade 6 sample: 0.08 Grade 9 sample: 0.12*
Read 180 Enterprise Edition	Scholastic, Inc.	An elective for students in grades 6–8	Select schools in Tennessee	ITBS: total reading; TCAP: reading and language arts	ITBS : 0.01; TCAP: 0.05 (2 years exposure to intervention)
Read 180 Enterprise Edition	Scholastic, Inc.	Replaces ELA in grades 6–8	Select schools in New Jersey	SAT: language arts, comprehension, vocabulary	Language Arts: 0.12; Comprehension: 0.09; Vocabulary: 0.02 (3 years exposure to intervention)
Read 180 Enterprise Edition	Scholastic, Inc.	Replaces ELA, high school grades	Select juvenile detention centers in Ohio	Scholastic Reading Inventory Assessment	0.22** (2 quarters exposure to intervention)
Xtreme Reading Strategic Instruction Model	University of Kansas Center for Research on Learning	Replaces ELA for grades 7–8; an elective for grades 9–10	Select schools in Oregon	GRADE: vocabulary, comprehension, oral language; OAKS: reading and literature	**Middle school** • GRADE: 0.29** • OSAT: 0.12* **High school** • GRADE: 0.12* • OSAT: 0.02 (1 year exposure to intervention)

Intervention	Developer	Setting		Outcome measures	Impacts (standard deviations)
		Grade level	Location		
Strategies for Literacy Independence across the Curriculum (SLIC)	T. McDonald & C. Thornley, Education Associates, New Zealand	An elective for students in grades 7 and 9 (year 1) and 7–10 (year 2)	Select schools in California	CST: English language arts; Degrees of Reading Power: comprehension	• CST: -0.03 • DRP: 0.03 (2 years exposure to intervention)
Read 180 Enterprise Edition	Scholastic, Inc.	An elective for students in grade 9	Select schools in Massachusetts	SDRT: comprehension	0.11* (1 year exposure to intervention)
Xtreme Reading Strategic Instruction Model	University of Kansas Center for Research on Learning	An elective for students in grade 9	Select schools in Massachusetts	SDRT: comprehension	0.02 (1 year exposure to intervention)

* Significant at 95 percent confidence level; **Significant at 99 percent confidence level.

Note: CST = California Standards Test; DRP = Degrees of Reading Power; ELA = English language arts; GRADE = Group Reading Assessment Diagnostic Evaluation; ISAT = Illinois Standard Achievement Test; ITBS = Iowa Test of Basic Skills; na = not applicable; OAKS = Oregon Assessment of Knowledge and Skills; SAT = Stanford Achievement Test; SDRT = Stanford Diagnostic Reading Test; SPED = special education; TCAP = Tennessee Comprehensive Assessment Program.

a. The Chicago SR program administers a reading assessment early in the school year and divides students into three tiers, with the strongest students identified as tier 1 and the weakest students identified as tier 3.

Source: Striving Readers Implementation and Impact Studies on school years 2006-07 through 2009-10. (http://www2.ed.gov/programs/strivingreaders/performance.html)

The What Works Clearinghouse (WWC) provides reviews of evaluative studies of interventions that use designs capable of detecting causal relationships between interventions and outcomes. Although WWC does have reviews of three whole-school literacy interventions suitable for middle schools and high schools (schools serving grades 6–12), only one of the reviews includes studies of whole-school approaches that actually tested for impacts on adolescent students' outcomes. Thus, few rigorous evaluation studies exist that have examined whether whole-school approaches have an impact on adolescent literacy.[11]

That research base is expected to build dramatically over the next two to three years as additional findings from Striving Readers grant projects become public. (Currently available findings appear in table 1.2.) In addition to the targeted interventions that are being sponsored and evaluated as part of this grant program, the program also is supporting whole-school approaches. Several of these interventions are being tested using cluster-randomized control trials, although most are being evaluated using time series analysis (that is, comparison of reading performance in schools prior to implementation with performance after implementation). Regardless of evaluation design, implementation of the whole-school interventions was slated to take either four or five years (most ending in 2011), making evaluation findings unavailable until 2011 at the earliest.

An example of a tiered, whole-school adolescent literacy intervention: the Content Literacy Continuum

A third approach to providing adolescent students with reading strategy instruction includes elements of both targeted and whole-school approaches. That is, literacy strategies can be provided to students across content area classes while students needing extra reading support can be given the opportunity to participate in supplemental reading classes. In their review of relevant rigorous research, Slavin, Cheung, Groff, and Lake (2008) classify this type of intervention with "instructional process" models that, when applied to entire schools, represent a whole-school or comprehensive school reform model.

The Content Literacy Continuum (CLC), developed by researchers at KU-CRL, is an example of this type of combined approach. First, within the CLC framework, student weaknesses in literacy skills across all core subject areas are addressed by training core content teachers to use instructional routines and to model learning strategies that may help students of varying skill levels better comprehend critical content from instruction and text.[12] These instructional routines and learning strategies have been developed, tested, and refined repeatedly by KU-CRL researchers over the past 30 years; however, they have not been tested using research designs that control for other potential confounding factors (Deshler & Schumaker, 2006; Schumaker & Deshler, 2003).[13]

[11] The What Works Clearinghouse's intervention reports in the area of adolescent literacy can be found at http://ies.ed.gov/ncee/wwc/reports/topicarea.aspx?tid=15.

[12] Literacy skills are defined here as reading and writing skills. Although the levels of CLC evaluated in this study incorporate activities designed to improve both reading and writing, this evaluation focuses on program impacts on students' reading achievement.

[13] The instructional routines and learning strategies comprise the cross-content component of the CLC framework referred to as KU-CRL's Strategic Instruction Model (SIM) (Lenz, Ehren, & Deshler, 2005). SIM is marketed for classroom use.

Second, CLC is designed to offer targeted reading support to struggling adolescent readers by training reading teachers to provide these students with focused instruction on learning strategies and to show students how to apply the strategies using a series of short, engaging novels. KU-CRL developed a distinct curriculum for classes for these struggling readers who have foundational decoding, fluency, and comprehension skills (that is, at least at a grade 4 level), but who are sufficiently below grade level to warrant more intensive support. This reading curriculum has undergone several iterations, each of which was refined through testing and evaluation. KU-CRL calls the current version Fusion Reading (Hock, Brasseur, & Deshler, 2008).

Integration of the CLC framework throughout a high school involves professional development of core content and reading teachers by KU-CRL–trained implementation specialists (referred to later as "site coordinators"). These specialists also work with school leaders to establish organizational structures and processes within schools to coordinate the components of CLC. According to KU-CRL, the coordination is critical for helping all students develop the necessary literacy skills and apply those skills throughout the school day.

Fundamental to the CLC approach is the assumption that students need to have the literacy skills that allow them to access, interpret, and express critical course content in order to master that content. Providing students with explicit literacy instruction within content area courses also allows students to develop and apply their literacy skills within relevant school contexts. According to KU-CRL, the combination of literacy instruction incorporated within core content areas, intensive literacy instruction support for students who are struggling with reading, and coordination of teacher professional development and student placement within a school results in a framework that undergirds schoolwide instructional change with the intent of increasing students' literacy skills and their mastery of course content. Specific details on the components of the CLC approach and implementation of the CLC framework within schools are presented in the following sections.

The multilevel CLC framework

The CLC framework can be viewed not only as a hybrid of whole-school and targeted approaches to improving literacy skills of adolescent students, but also as an approach that provides increasing literacy support for students with greater learning needs. The project described in this report implemented and tested three of the CLC "levels of support":[14]

- ***Level 1: enhanced content instruction.*** Teachers of core subjects (that is, English language arts, mathematics, science, social studies) develop and use instructional routines designed to help students at all literacy levels master critical content, acquire vocabulary and background knowledge to improve comprehension and communication skills, and be better able to organize complex content. These instructional routines are referred to as Content Enhancement Routines.

[14] The full CLC framework involves five levels, with levels 4 and 5 focused on providing even more intensive supports to students with greater learning needs. Students being served at these levels often receive special educational services. At the outset of the project, the REL and the developer decided that focusing implementation on levels 1–3 would be more reasonable given the expectation of working in approximately 10 districts across multiple states within only a two-year implementation time frame.

- ***Level 2: embedded strategy instruction.*** Core content teachers work with students to develop one or two learning strategies (such as paraphrasing or self-questioning) that align with the specific demands in their courses.
- ***Level 3: intensive strategy instruction.*** Teachers in stand-alone, supplemental reading classes (Fusion Reading) provide more intensive instruction to students who are reading two to five years below grade level (that is, students who need to develop comprehension strategies).

Given these three levels of support, CLC can be viewed as high school–level analog to tiered response-to-intervention (RTI) reading instruction frameworks used in elementary schools. RTI frameworks often are organized in three tiers of instructional support. The first tier focuses on support for all students—schoolwide or classroom-wide. The second tier focuses on smaller, targeted groups of students who require more intensive support than the instructional practices used more broadly for all students. The third tier provides even more intensive support, very often in the form of special educational services, to students who are struggling even after receiving first- and second-tier support (Gersten et al., 2008).

CLC levels 1 and 2 correspond to the first tier of an RTI model: they are implemented in all core content classes within a school in support of all students. Level 3 of CLC is analogous to the second tier of response to intervention in that it targets support to struggling readers by supplementing the support received in core content classes with support in Fusion Reading classes.[15]

CLC theory of action

Full implementation of the multilevel CLC framework follows a stepwise theory of action, with its ultimate goal being improved student academic achievement. The theory of action assumes that commitment to implementation by district and schools leaders and collaborative planning for implementation by school leaders and CLC site coordinators (that is, CLC-trained professional developers external to the district) will result in *structural* changes in schools, such as the formation of a school literacy team to lead framework implementation, the addition of supplemental reading classes to support struggling readers, and the provision of CLC-specific professional development for school staff. These changes are designed to create conditions under which *instructional* change aligned with the CLC model can take place. The program developers believe that CLC-aligned instructional practices ultimately improve the quality of teaching experienced by students and, as a result, improve students' literacy skills and overall achievement. Moreover, improvement in overall achievement may occur both indirectly (as a result of strengthened literacy skills, allowing students to access critical content more readily) and directly (as a result of more effectively presented critical content). The theory of action presented in figure 1.1 illustrates this progression from initial district and school commitment and communication with CLC site coordinators to structural and instructional changes to improvement in student outcomes.

[15] The levels of CLC that are not part of this study (levels 4 and 5) focus on providing special-needs students with even more support. To continue the RTI analogy, these students would be similar to the students in the third tier of an RTI framework.

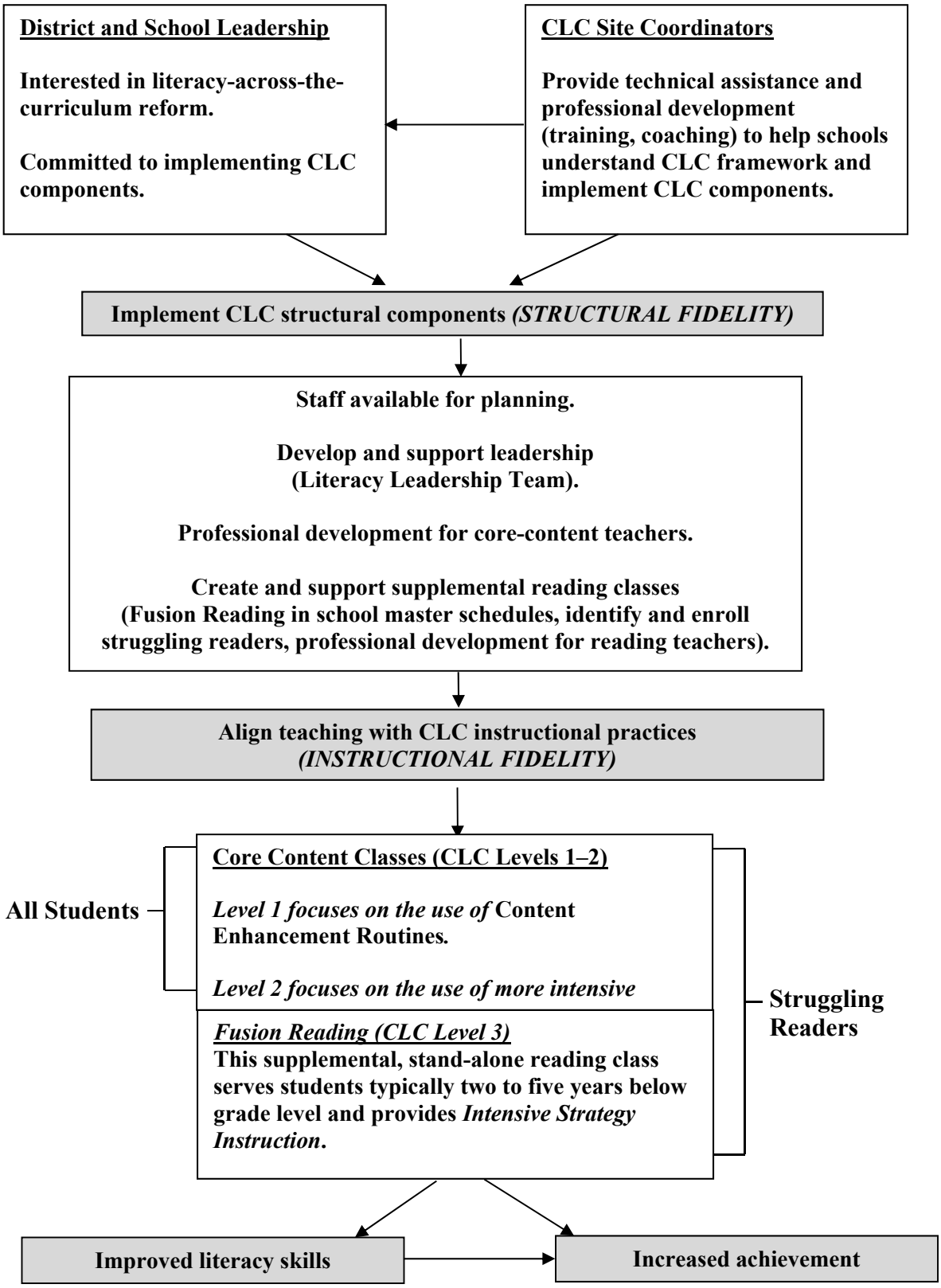
Given the multiple components of the CLC framework, its staged implementation from initial district and school commitment to structural change to instructional change, and its intent to change teaching and learning across a school, CLC implementation should be understood as a challenging and time-consuming undertaking. Prior research findings provide context for understanding the kind of time and effort that are necessary to implement a comprehensive instructional reform such as CLC and realize its full impact. Desimone's reviews of research on comprehensive school reform models cite numerous studies conducted in late 1980s and 1990s suggesting that it takes time (more than three years) to implement a school reform model (Desimone, 2000, 2002). Aladjem et al. (2006) found that comprehensive school reform models (defined as schoolwide initiatives that require coordination and changes to multiple school processes) can produce impacts during the third, fourth, or fifth year of implementation assuming high fidelity across all model components. Moreover, a meta-analysis of comprehensive school reform models by Borman, Hewes, Overman, and Brown (2003) suggests that years of implementation matter. Comprehensive school reform models implemented between one and four years showed standardized mean differences (effect sizes) ranging from 0.13 to 0.15. Models implemented for five years or more showed impacts of 0.23 to 0.50 (Borman et al., 2003). Implementing the kind of schoolwide change that the CLC framework entails requires significant effort and time for full implementation of the model and full realization of its potential impacts on teacher practices and student outcomes.

Structural components of CLC

CLC site coordinators initiate schoolwide literacy reforms by securing the support of district and school leadership. This support includes working with site coordinators to implement the CLC program's key structural components, which focus on planning, leadership, professional development for core content teachers, and the creation and support of supplemental reading courses for struggling readers.

Planning. To get CLC implementation started, school staff need to participate in planning activities that provide information to CLC site coordinators about the school, its staff, and its students. Staff and the site coordinator also cocreate a plan for implementing the CLC framework such that it aligns to school goals and practices as well as student needs. District support often is necessary to help obtain school-level data and to ensure that time is made available for school staff to participate in planning.

Figure 1.1. CLC theory of action



This planning phase is intended to last a minimum of one semester. During that time, CLC site coordinators conduct five activities designed to introduce the CLC framework to school and district staff, to outline the necessary steps for implementation in the schools, and to gather information about the schools. These activities are (1) discussing the CLC framework and implementation with district and school liaisons, (2) meeting with the school leadership team, (3) providing an overview of CLC to school faculty, (4) administering a school climate survey to school staff, and (5) interviewing teachers. These activities involve an information exchange intended to facilitate the ongoing collaboration during CLC implementation between the site coordinator and the school with which she is working.

By design, the CLC program is intended to be phased into high schools, focusing first on all grade 9 students and providing the necessary curricular materials and professional development to their teachers. Implementation then expands to grade 10 teachers and students in the next year.[16] This stepwise approach was designed to support stronger initial implementation among a smaller group of teachers and to increase the chances of successful adoption, implementation, and sustainability over time.

Leadership. During initial meetings with school leaders, site coordinators emphasize the need to establish the Literacy Leadership Team consisting of teachers representing all core content areas and school leadership. This team meets monthly and its purpose is to guide the implementation of CLC practices and make sure that teachers have the resources necessary for doing CLC-aligned instruction. Specifically the Literacy Leadership Team can help develop a schedule for the site coordinator's work within the school; prioritize specific instructional components of CLC for implementation (Content Enhancement Routines and Learning Strategies, to be described later in this chapter) based on student performance data, perceived areas of instructional weakness, and other ongoing initiatives; support the identification and enrollment of students in need of supplemental reading classes (Fusion Reading, described below); and encourage core content teachers to participate in CLC professional development and use the routines and strategies in their classrooms.

Professional development for core content teachers. To help establish a schoolwide emphasis on content literacy, site coordinators' first meetings with core content teachers involve providing the overview of CLC that highlights how the concerted schoolwide, multitiered approach can improve instruction and student learning. The meetings also allow teachers the opportunity to ask questions regarding the CLC framework and the rollout of the professional development. Then the professional development focuses on supporting these teachers in their implementation of levels 1 and 2 of the CLC framework in their classrooms: enhanced content instruction and embedded strategy instruction. The use of CLC instructional practices by core content teachers is the method by which all students are exposed to content literacy instruction.

Through discussions with each school's Literacy Leadership Team, site coordinators request a monthly site visit of two to three days (totaling 18 to 27 days total across the year) during which they provide on-site professional development, modeling, and coaching of core content teachers on the teaching routines and learning strategies that are key aspects of the instructional side of CLC (that is, Content Enhancement Routines and Learning Strategies). During the first year of

[16] Implementation in a third year would typically extend to the rest of the high school staff, expanding CLC fully across grades 9–12. This study includes only two years of CLC implementation—thus, to teachers and students in grades 9 and 10.

implementation, site coordinators typically introduce a small number of routines and strategies to core content teachers. This approach emphasizes depth of mastery over breadth of training and provides site coordinators with time to gauge teachers' specific instructional needs before introducing additional strategies, both of which aim to encourage teacher buy-in.

Most professional development with teachers takes place during these monthly two- or three-day site visits to the school. During these visits, site coordinators—often with a partner—work with teachers to develop strategies and routines, demonstrate the strategies and routines in teachers' classrooms, and observe teachers as they use the routines or strategies with their students. Some of this work is conducted in large-group training sessions, but much of it takes place with small groups or individuals. Site coordinators also are accessible via email or telephone to teachers and school leaders who might have questions about routines or learning strategies. In addition, teachers within implementing schools are given access to Web-based tools that provide guidance about how to develop routines and to video clips showing routines and strategies being used in classroom instruction.[17]

Create and support supplemental reading classes. Some students require literacy support in addition to that provided by their core content teachers through the use of CLC instructional practices in English language arts, mathematics, science, and social studies classes. Thus site coordinators' initial meetings with school principals, instructional leaders, and the Literacy Leadership Team also involve establishing the level 3 Fusion Reading classes to support students assessed as being about two to five years below grade level in reading or who have foundational reading skills but need additional support with comprehension. Setting up these courses often requires preliminary work with district leaders to create a course number and course description. The site coordinator then can work with the school leaders to use existing data to identify the students requiring the level 3 support (or administer a reading test to students for this purpose) and determine the necessary number of course sections within the school's master schedule and the number of teachers needed to teach them. The school leaders and site coordinator also identify which teachers could teach the level 3 classes (usually experienced English, reading, or special education teachers).

Teachers identified by schools to teach sections of the level 3 Fusion Reading course receive training on the Fusion curriculum and recommended instructional techniques during a three-day summer training session, followed up by a two-day training in the fall. These training sessions are led jointly by site coordinators and the KU-CRL researchers who developed the Fusion curriculum. Additional training is scheduled with teachers as needed. The three-day summer Fusion training sessions involve a course overview, materials and texts used in the course, recommended timeline for the course, discussion of the high-interest adolescent literature and other reading materials used for the course, and discussion of several of the learning strategies that can be shared with students during the course's first few months. Teachers are taught how to administer the Gates-MacGinitie[18] reading test to students at the beginning of the course (in the

[17] These tools were developed by TeachScape in collaboration with KU-CRL.

[18] The Gates-MacGinitie reading test is a group-administered assessment of student reading achievement (MacGinitie, MacGinitie, Maria, Dreyer, and Hughes, 2011). Additional details can be found at the publisher's website: http://www.riversidepublishing.com/products/gmrt/details.html.

fall) and at the end of the course (in the spring).[19] During the fall follow-up training sessions, reading teachers are given a refresher of strategies covered during the summer and are provided additional training on the remaining learning strategies and how to incorporate them into the Fusion class.

Additional support is provided to reading teachers during the two to three days per month that site coordinators spend in each school. Site coordinators connect with reading teachers to observe their sections of the Fusion class; answer any questions; and provide suggestions on engaging students, refining their presentation of strategies, or handling other concerns.

Instructional features of CLC

The implementation of these four types of structural supports is expected to create conditions under which CLC-aligned instruction can occur in classrooms. This instruction falls into two categories. First are instructional practices used in core content classes intended to support the content learning and literacy needs of *all* students (CLC levels 1–2). Second are instruction and support provided in supplemental reading classes to a *targeted subgroup* of students who are struggling readers (CLC level 3). Instruction in core content classes and instruction in the supplemental reading classes need to be coordinated so that implementation of the CLC framework represents connected support across levels. Therefore some shared pedagogical practices of CLC-aligned instruction are employed by teachers at levels 1, 2, and 3.

Content Enhancement Routines and Learning Strategies (CLC levels 1 and 2). The implementation of levels 1 and 2 of the CLC framework is focused on content-area classes (as opposed to the reading classes, which represent level 3 of CLC). Professional development focuses on training teachers in routines and strategies designed to help all students learn critical content within the core curriculum. This professional development for core subject teachers is critical if schools are to successfully implement CLC. Level 1 focuses on the use of Content Enhancement Routines; level 2 focuses on the use of more intensive Learning Strategies that are embedded in the school's existing content area curricula.

Content Enhancement Routines (CLC level 1). Professional development for level 1 of the CLC program involves training core content teachers to use instructional routines designed to make critical content more accessible to students. According to KU-CRL, these Content Enhancement Routines "help teachers of core subjects select the critical content, learn how to enhance that content for mastery, and then implement these enhancements through the use of explicit and sustained teaching routines" (Lenz, Ehren, & Deshler, 2005, p. 5).

KU-CRL organizes CLC's Content Enhancement Routines into four strands:

- **Routines for planning and leading learning.** These routines are used to introduce and emphasize the main ideas in courses (Course Organizer Routine), units (Unit Organizer Routine), or lessons (Lesson Organizer Routine). These routines also help students understand how the main ideas in courses, units, and lessons are interrelated.
- **Routines for teaching concepts.** These routines are used to introduce new concepts and situate those concepts within students' prior knowledge. These routines might help students

[19] This is the assessment that KU-CRL typically recommends Fusion teachers use as a pretest and a posttest of the students in their classes. These data are used by the site coordinators and the schools for program purposes and are not part of this evaluation's research.

learn new concepts by linking them to familiar ones (Concept Anchoring Routine), comparing and contrasting key concepts (Concept Comparison Routine), or defining concepts within a larger framework of knowledge and analyzing their key characteristics and examples (Concept Mastery Routine).

- **Routines for exploring text, topics, and details.** These routines are used to highlight the main ideas and related essential details within key texts or topics. These routines might help students construct a concrete representation of relationships between main ideas and essential details (Framing Routine) or explore words, phrases, and details related to a key topic's importance (Clarifying Routine).
- **Routines for increasing performance.** These routines are used to improve students' understanding, recall, and application of new information. These routines might help students use visual and auditory mnemonics to recall essential vocabulary (Vocabulary LINCing Routine) or pose and answer a critical question to explore a body of content information (Question Exploration Routine).

Learning Strategies (CLC level 2). Level 2 of the CLC program involves training core content teachers to provide more intensive instruction in selected learning strategies to their students. Core content teachers select one or two strategies based on their courses' critical content and student need. Then they explicitly teach those strategies, embedding them within their content instruction. These strategies often overlap with strategies taught in the level 3 Fusion Reading classes. The significance of this overlap is that it offers opportunities for struggling readers enrolled in Fusion Reading to get additional reinforcement of particular Learning Strategies outside of their reading classes.

As with the Content Enhancement Routines, the Learning Strategies are organized into strands (University of Kansas Center for Research on Learning, 2008):

- **Strategies for acquisition.** These strategies help students acquire information from written material by strengthening their decoding and comprehension skills. For example, the Word Mapping Strategy trains students to identify common morphemes (word prefixes, suffixes, roots) when decoding unfamiliar words. The Inference Strategy trains students to make inferences about texts and answer inferential questions about those texts.
- **Strategies for storage.** These strategies assist students with identifying, organizing, and storing important information. For instance, the LINCS Vocabulary Strategy, a more intensive version of the Vocabulary LINCing Routine, teaches students to use visual and auditory mnemonics to recall essential vocabulary.
- **Strategies for expression and demonstration.** These strategies focus on students' writing and how they demonstrate their understanding and mastery of material. The Paragraph Writing Strategy (in which students are taught to organize ideas and plan the sequence of topic, detail, and clincher sentences within a paragraph) and the Sentence Writing Strategy (in which students are taught to identify and create a variety of sentence types) are two examples of these strategies. The Essay Test Taking Strategy teaches students to analyze essay questions and organize, write, and edit structured responses to those questions.

Intensive strategy instruction (CLC level 3): Fusion Reading. As part of the implementation of the CLC framework, reading teachers, special education teachers, and other support personnel are trained to serve students who need more intensive support than that provided in levels 1 and

2. Within the CLC program, this intensive instruction (level 3) is delivered through a supplemental, stand-alone reading class called Fusion Reading, a two-year, multicomponent program designed to improve the literacy skills of struggling high school students (Hock et al., 2008).

These supplemental reading classes generally serve students whose decoding skills and fluency levels are two to five years below grade level and who need to develop the comprehension strategies to successfully meet the reading demands of the high school curriculum. Fusion Reading classes can meet either daily or every other day, and classes can be adjusted to fit either 90-minute or shorter periods (that is, 50 to 60 minutes).

KU-CRL provides Fusion Reading teachers with a prescribed, two-part instructional scope and sequence that can be covered in a single year of daily, 90-minute classes or two years of daily, shorter classes. If schools wish to implement daily Fusion Reading classes shorter than 90 minutes, each part of the Fusion Reading scope and sequence can stand alone as a coherent, yearlong reading course for students in grades 9 and/or 10.

Fusion classes include three major elements: motivation, classroom management, and reading instruction. Motivation is cultivated by providing students with engaging and accessible adolescent literature; teaching them learning strategies for successfully reading such literature; and encouraging them to create goal-oriented, motivating visions of their future selves. Classroom management creates a positive learning environment by blending whole-group, small-group, and individual instruction and establishing high expectations and clear routines linked to reading instruction and classroom behavior. Fusion Reading's core element—reading instruction—emphasizes two primary components of reading: word recognition and language comprehension (Hock et al., 2008):

- **Word recognition** instruction helps students respond to word-level demands within narrative and expository texts. Instruction in word recognition is primarily provided through the Bridging Strategy in which students are explicitly taught advanced phonics, decoding, word recognition, and fluency skills through a series of short but intense activities.
- **Language comprehension** instruction helps students improve their ability to derive meaning from text. Instruction in language comprehension includes (1) the Prediction Strategy, which helps students use clues and prior knowledge to predict a text's content and to confirm predictions by examining the text and (2) the Summarization Strategy, which helps students use their prior knowledge to identify and paraphrase a text's main ideas and important details. The language comprehension component also includes the Seven-Step Vocabulary Process, in which students are trained in a deliberate process to acquire and retain new vocabulary.

Finally, Fusion Reading embeds strategy instruction within a cognitive and metacognitive process known as "Thinking Reading." During Thinking Reading, teachers engage students in discussions of text while modeling and reviewing word recognition and language comprehension strategies. Through scaffolded instruction, teachers guide students toward applying such strategies with increasing independence.

During a typical Fusion Reading class, a teacher might begin instruction with a brief warm-up, followed by 20 minutes of Thinking Reading. The teacher then might devote the remaining class

time to vocabulary instruction, explicit instruction in one or more Learning Strategies, and/or opportunities for independent reading, depending on curricular progress and student need.

Those teachers trained to teach Fusion Reading receive written instructional materials that provide guidance on how to establish the course, what the Learning Strategies are, and how to teach those Learning Strategies.

Shared pedagogical practices of CLC levels 1, 2, and 3. Although representing literacy instruction of varying intensity, CLC levels 1, 2, and 3 are designed with intentional cross-level continuity (Lenz et al., 2005). In other words, the CLC framework is designed so that instructional approaches are connected across levels. Thus, in addition to maintaining a consistent focus on literacy-oriented routines and strategies, CLC levels 1, 2, and 3 share several other key pedagogical practices: sequenced instruction, multiple instructional modalities, and interactive and scaffolded discourse. This shared pedagogy across levels is viewed by KU-CRL as being particularly beneficial to struggling readers, who will experience similarities in instructional approaches and see connections across their content area classes and their reading classes.

Sequenced instruction. Across all levels of the CLC program, teachers are trained to deliver instruction through carefully planned sequences that establish the purpose for instruction, engage students in literacy instruction, and review the content of such instruction.

Content Enhancement Routines (CLC level 1) are presented through a three-phase instructional sequence known as Cue-Do-Review. In the Cue phase, teachers establish the purpose for instruction by cuing students that a specific routine will be used, explaining how it will help students learn, and specifying what students need to do to participate in the routine. In the Do phase, teachers engage students in the specified routine (typically anchored by a visual device known as a graphic organizer, described in more detail below) and co-construct new content knowledge through a series of interactive tasks. In the Review phase, teachers review students' understanding of new content knowledge and the routine used to co-construct that knowledge. Typically the Cue-Do-Review sequence is completed within a single class period.

The embedded Learning Strategies (CLC level 2) are presented through an eight-stage instructional sequence that represents a more intensive version of Cue-Do-Review. The first three stages (Pretest and Make Commitments, Describe, and Model) establish the purpose for learning a new strategy, present the strategy's uses and processes, and explicitly demonstrate how the strategy should be used. During the next three stages (Verbal Practice, Controlled Practice and Feedback, and Advanced Practice), teachers engage students in multiple forms of practice: explaining and using the new strategy; applying the strategy to materials designed for practice; and using the strategy in increasingly independent, less structured tasks. The final two stages (Posttest and Make Commitments and Generalization) review students' mastery of the new strategy and encourage the strategy's application to new contexts. This eight-stage instructional sequence extends over multiple class periods, often across multiple weeks.

Fusion Reading's intensive strategy instruction (CLC level 3) follows a similar sequence to that occurring within CLC level 2. Teachers first establish the purpose for learning a new strategy and present or demonstrate the strategy. Next teachers engage students in multiple forms of practice. Finally teachers review students' mastery of that strategy. For instance a Fusion Reading teacher might begin instruction related to the Prediction Strategy by preparing students to learn a new strategy, introducing the new strategy, and modeling how a proficient reader

makes predictions. Next the teacher explains each step of the Prediction Strategy: checking for clues within a text, linking those clues to prior knowledge, using clues and prior knowledge to make predictions, and examining whether the text confirms predictions. After explaining each step, the teacher engages students in guided, partnered, and individual practice of the step. Finally the teacher concludes each part of the strategy lesson by reviewing students' work with the strategy and clarifying any questions they might have about the strategy or its applications. As in level 2 this instructional sequence usually extends over multiple class periods.

Multiple instructional modalities. Across the CLC program, teachers are trained to present information by using a combination of verbal, graphical or visual, and written modalities. Content Enhancement Routines and Learning Strategies (CLC levels 1 and 2 for content area teachers) typically use visual devices known as graphic organizers, which are visual schema that help students learn, organize, and apply important information related to targeted content knowledge. CLC trains teachers to embed graphic organizers within classroom instruction through a blend of oral and written discourse. Finally these graphic organizers—and the discursive modalities associated with them—are explicitly linked with a named routine or strategy, which helps students recognize specific routines and strategies, recall their steps, and apply them to new situations.

Fusion Reading (CLC level 3) also uses graphic organizers to help students organize information during literacy instruction, although these graphic organizers are somewhat simpler and less abstract than those used in CLC levels 1 and 2. For example teachers provide students with a Blank Steno Pad Template during the Seven-Step Vocabulary Process, which the teacher and students use to record, define, and analyze new vocabulary words. Fusion Reading also trains teachers to embed these graphic organizers within a blend of oral and written discourse.

Interactive and scaffolded discourse. The CLC program encourages instruction that is highly interactive, in which teachers and students actively co-construct knowledge while using the various Content Enhancement Routines and Learning Strategies (CLC levels 1 and 2 for content area teachers). As indicated previously instruction in these routines and strategies is deliberately scaffolded, with instructional activities progressing through several steps: teacher-centered modeling, teacher-mediated practice, student-mediated practice, and independent student practice.

During Fusion Reading (CLC level 3), teachers also present instruction in reading and learning strategies in an interactive manner, aiming for a balance of teacher-student discourse. In addition Fusion Reading instruction reflects the same scaffolded approach emphasized within CLC levels 1 and 2, progressing from teacher-mediated practice to student-mediated practice. Relative to CLC levels 1 and 2, however, Fusion Reading might progress more slowly through these phases or return more often to earlier phases, depending on student need.

Implementation support

With support from the program developer, district and school staff members are expected to implement both the structural and instructional components of CLC to make this schoolwide approach to literacy work. For this project the implementation of CLC in schools was supported by Action Designs, an organization led by researchers affiliated with KU-CRL. Action Designs follows KU-CRL's implementation model for CLC, whereby trained site coordinators work with school leaders and staff to establish the necessary structures for CLC (for example, implementing

the Fusion Reading classes, creating a Literacy Leadership Team, and inserting CLC-specific professional development activities into school calendars) and to plan for demonstrating and teaching instructional features to teachers.

Action Designs directed a cadre of site coordinators who are former teachers who have successfully implemented CLC or the Strategic Instruction Model (levels 1 and 2 of the CLC) within their own classrooms and received further training in how to guide other teachers in the implementation of these approaches. Each school implementing CLC was assigned a site coordinator who worked with the school's leadership to establish the necessary schoolwide structures and provided the requisite professional development to teachers on the program's instructional routines and strategies. Throughout the course of their work with a school, the CLC site coordinators solicited feedback from the school administrators and staff involved in leading CLC implementation on the successes or challenges related to putting CLC structures and practices in place. The site coordinators then could make adjustments as necessary to strengthen the implementation at the school.

The site coordinators also received peer and supervisory support in their work. Site coordinators working in the same or nearby states often traveled in pairs to one another's assigned schools to assist with site visit tasks and activities. Face-to-face meetings that included all site coordinators and Action Designs' central leadership team were conducted two to three times a year. At these meetings the professional development plans that were created by each school's Literacy Leadership Team in collaboration with its site coordinator were reviewed and fine-tuned based on formative data from the schools. Furthermore one of Action Designs' leadership team members was responsible for oversight of the work of the site coordinators. This team member spoke on the phone weekly and exchanged email daily with each site coordinator. These communications allowed for collaborative problem solving regarding any implementation challenges site coordinators were facing. These calls and emails also provided ongoing updates on implementation progress to the Action Designs leadership team. The site coordinators also participated in annual meetings hosted by KU-CRL to benefit from ongoing training for CLC professional developers, to stay current on KU-CRL–supported practices, and to connect with other professional developers and school staff members in the Strategic Instruction Model (SIM) national network.

The study: a rigorous impact evaluation

This study is a rigorous test of CLC conducted by REL Midwest, MDRC, and Survey Research Management (SRM).[20] These organizations have conducted this study to build knowledge about the cumulative effects of the CLC components when used together throughout a school. In particular this study uses a school-level randomized experiment to assess the causal impact of CLC on students' achievement across content areas and reading skills.

REL Midwest and MDRC evaluated the CLC framework by conducting a two-year (2008/10) cluster randomized trial in which high schools within school districts were randomly assigned to implement the CLC intervention (the CLC group) or not implement CLC and continue with

[20] SRM played a data collection role by overseeing the administration and scoring of a reading test in both years of the study.

business as usual (the non-CLC group). Outcomes for three groups of students were analyzed: students in grade 9 in 2008/09 and students in grades 9 and 10 in 2009/10.

Thirty-three high schools in nine school districts participated in the first year of the study, and 28 schools in eight school districts also participated in the second year.[21] Participating schools are diverse in terms of location and context. The participating districts represent large city, midsize city, and rural locales predominantly in Michigan and Ohio (22 schools), but also in Indiana and Wisconsin (6 schools). For this study the research team recruited sites with characteristics often associated with greater need for support: larger high schools with relatively high percentages of students eligible for free or reduced-price lunch and relatively large numbers of students reading below proficient on standardized tests of reading. The CLC cross-curriculum and multilevel approach also had to be sufficiently different from other literacy and reform efforts underway in eligible schools. Finally school and district leaders had to demonstrate a commitment to implement the intervention and facilitate the research team's data collection efforts.

This study seeks to determine whether the CLC framework improves the academic outcomes of students. The primary research questions of the study focus on the impact of the CLC program on students' academic outcomes in the second year of the study:

- What are the impacts of the CLC program on grade 9 students' reading comprehension and accumulation of core credits in 2009/10, the second year of the study?
- What are the impacts of the CLC program on these outcomes for grade 10 students in the second year of the study?

The primary analysis focuses on the second year of the study because at that point in time, grade 9 teachers would be more experienced with the elements and delivery of the CLC program (thereby potentially producing stronger impacts on grade 9 students' academic outcomes). In addition, many grade 10 students would have received two years of CLC-related instruction and services, thereby allowing the research team to examine issues related to greater dosage of the CLC program. Moreover, based on its prior experience supporting the implementation of the CLC framework, KU-CRL expects that it takes at least two years for the intervention to mature within a school and that effects most likely will appear only after two years of implementation. In addition, as noted earlier in this chapter, research on comprehensive school reform has indicated that it often takes time for reforms to take hold fully in a school, often three or more years. Thus even this study may not follow the implementation of CLC in participating schools long enough to realize the framework's full impact; rather it may reflect initial impacts based on an initial foothold within participating schools.

For the exploratory (or secondary) impact analysis, the evaluation team examines questions that may contribute to the interpretation of the primary impact results and generate hypotheses for future research. These questions examine whether the CLC program is more effective for some subgroups of students than others and whether it has greater impacts over time:[22]

[21] The reasons for attrition from the study include school closure due to declining enrollment and fear that CLC's approach would conflict with state-mandated changes due to state sanctions.

[22] A subset of the teachers in the second year of the study will have gained experience in CLC instructional techniques. (Because of teacher mobility, some teachers in the CLC schools will be new to the program.) At the same time, a subset of grade 10 students in the second year of the study will have been exposed to two years of CLC instruction. (Because of

- What are the impacts of the CLC program on other academic outcomes, such as vocabulary test scores and grade point average?
- Is the impact of the CLC program on reading comprehension and credit earning for students in grades 9 and 10 in the second year of the study greater for some subgroups of students than for others?
- What are the impacts of the CLC program on grade 9 students' reading comprehension and credit earning in the first year of the study? Are these impacts different from those occurring in the second year of the study?[23]

This evaluation of CLC is designed not only to address these research questions, but also to examine the implementation of the CLC framework. The study examines the degree to which implementing schools established the structures needed to support CLC and the degree to which teachers within CLC schools incorporated the Content Enhancement Routines, Learning Strategies, and CLC-emphasized pedagogical practices within their instruction. The study also examines the extent to which implementing the CLC framework resulted in a contrast in structural components and instructional practices between the CLC schools and the non-CLC schools that continued with business as usual. This report provides initial information about these implementation issues.

Structure of this report

The remaining chapters of this report provide more detailed information about the evaluation's research design, the implementation of structural and instructional aspects of CLC, and impacts on student outcomes. Chapter 2 discusses methodological issues such as the experimental design, the sample of schools and students, data collection, measures, and response rates. Chapter 3 presents information regarding implementation of the *structural* features of CLC. Chapter 4 then focuses on the degree to which teachers demonstrated the *instructional* features of CLC within the classroom. The fifth chapter of this report presents the impacts of CLC on student outcomes, in particular on students' reading ability and course performance. Chapter 6 includes findings from exploratory analyses that attempt to provide additional insight into the major impact findings from chapter 5. Finally chapter 7 presents an overall summary and limitations of the findings.

student mobility, some grade 10 students will be new to a CLC school and thus CLC instruction.) Given these conditions, it is relevant to investigate differences in impacts by study year.

[23] Impacts on the academic outcomes of grade 10 students cannot be estimated for the first year of the study because the CLC program was implemented only in grade 9 in the first year.

Chapter 2: Study Design, Data Collection, and Impact Analysis

This chapter provides an overview of the research design for the Content Literacy Continuum (CLC) study. It begins by describing the schools in the study and provides a discussion of the data that were collected to measure student outcomes and to assess implementation of the CLC framework. This is followed by a description of the samples of students that are used to evaluate impacts on student outcomes and the samples of classrooms used to evaluate implementation fidelity and the service contrast in instructional practice. The chapter concludes with a discussion of the analytic methods used to assess the impact of CLC in this report. The following key points are discussed in this chapter:

- The target population of schools for the CLC study consists of high schools in the Midwest region serving high proportions of low-income and low-achieving students.
- In the first year of the study, 33 eligible high schools from nine school districts were recruited into the study and participated in the first year of the study. Of these schools, 17 were randomly selected to implement the CLC intervention (the CLC group), and the remaining 16 schools agreed not to implement the intervention during the study period (the non-CLC group). Of these 33 high schools, 28 schools from nine school districts also participated in a second year of implementation (15 CLC schools and 13 non-CLC schools). The impact analysis in this report focuses on the latter group of 28 schools because, for these schools, student outcomes can be tracked over two school years and for two grade levels.
- In the school year before random assignment, CLC and non-CLC schools did not systematically differ with respect to their characteristics. This confirms that random assignment resulted in two groups of schools whose characteristics are statistically equivalent in expectation at the start of the study, and that any differences in outcomes between CLC and non-CLC schools can be attributed to the impact of the CLC framework.
- The target population of students for this study includes all grade 9 students enrolled in the study schools in the spring of year 1, and all grade 9 and grade 10 students enrolled in the spring of year 2.
- Students' reading achievement was measured using the Group Reading Assessment Diagnostic Evaluation (GRADE) reading assessment, which was administered in the spring of each study year. Data on students' credit accumulation and grade point average (GPA) were obtained from school records provided by the participating school districts.
- Impacts on academic outcomes (credit accumulation and GPA) are based on students enrolled at the study schools on the last day of the school year, and impacts on reading achievement are based on students who completed the GRADE assessment in the spring of each school year. Students in CLC and non-CLC schools are not systematically different with respect to their background characteristics and prior achievement. Thus these samples preserve the balance that was achieved with random assignment, which means that differences in student outcomes between the two groups of schools reflect the impact of the CLC framework.

- The primary indicators of effectiveness in this report are impacts on students' reading comprehension scores and their accumulation of core credits in the second year of the study. Statistical power calculations indicate that the study can detect an impact of 0.14 to 0.25 in effect size on these primary outcomes, depending on the grade level and outcome of interest.

Schools in the study

Because CLC is a whole-school intervention, the impact evaluation is based on a cluster (school-level) random assignment design. Among the schools that agreed to be part of the study, a subset was randomly selected to implement the CLC intervention (CLC group), and the remaining schools represent the non-CLC control group. The random assignment of schools to the CLC intervention makes it is possible to draw valid inferences about the impact of the framework on teacher and student outcomes. The non-CLC schools serve as a benchmark or counterfactual for how schools assigned to the CLC group (and their students) would have performed had they not had the opportunity to implement the intervention. Therefore differences in outcomes between the CLC and the non-CLC groups represent the impact that the CLC intervention had on student outcomes in these schools. The remainder of this section describes the recruitment of schools into the study, the random assignment process, and the characteristics of schools that participated in the evaluation.

Recruitment and random assignment of schools

The target population for the CLC study consists of high schools in the Midwest region serving high proportions of low-income and low-achieving students. As a first step in the site-recruitment process, the study team identified school districts in the Midwest region with at least two high-needs high schools, as defined by the following two criteria: at least one-third of students in the high school are struggling readers (that is, they are below proficient on the state or district reading assessment), and at least one-fourth of students come from low-income families (that is, they are eligible for free or reduced-price lunch).

Within this target group of schools, eligibility for the study was further limited to schools in which it would be possible to provide a fair test of the CLC framework—that is, sites that were presumed to be capable of implementing CLC with a reasonable level of fidelity and where there would be a clear service contrast between the CLC intervention and "business as usual." Schools already implementing a combination of "literacy-across-the-curriculum" programming and supplemental reading courses for struggling readers were not eligible for the study. There also needed to be a commitment from the superintendent and district leadership to the CLC intervention and its evaluation—they had to be willing to make school records data available, cooperate with research and data requirements, and facilitate Institutional Review Board approval within the district. To ensure that there would be diversity in terms of context and location, the study team made a concerted effort to approach school districts in several states and to recruit school districts of varying sizes (large city, midsize city, and rural). Based on power calculations, the goal was to identify and recruit approximately 30 to 40 high schools into the study.

Site recruitment unfolded during the 2007/08 school year and the summer of 2008. The recruitment process began with the identification of 57 eligible school districts. REL Midwest

sent a one-page recruitment letter to the superintendent or assistant superintendent for curriculum/instruction in each of these districts. A week later these individuals were called to see if they had received the letter and whether they would like to participate in a conference call regarding the project. This call was conducted with district staff, and in some instances high school principals, as well as program development staff from Action Designs, REL Midwest, and MDRC. Of the 57 eligible districts, 22 districts (39 percent) participated in an informational conference call. If a school district wanted to pursue the opportunity further, an in-person "mini-conference" was then held with district staff and school leaders. At this conference, the CLC intervention and the study design were described, and school staff were offered the opportunity to ask questions. Of the 57 eligible districts, 13 districts (23 percent) participated in a mini-conference. After the mini-conference, most schools needed to have internal discussions to determine whether faculty would support the project, regardless of whether they were assigned to the treatment or control group. In order for a district to participate in the study, a verbal commitment was needed from at least two eligible high schools in each district, given the experimental study design. A Memo of Understanding was sent to the district liaison and to school principals. The district superintendent, assistant superintendents, and all school leaders at interested high schools had to sign the document prior to random assignment.

Of the 57 eligible districts, nine districts (16 percent) were ultimately recruited into the study, representing four states (Indiana, Michigan, Ohio, and Wisconsin). These districts decided to participate because they had clear literacy goals and/or were interested in the additional professional development resources offered by the program. Districts that declined to participate did so for several reasons. Of the 13 districts that participated in the mini-conference, those that chose not to participate in the study gave various reasons for their decision: they were already implementing literacy initiatives, they were unable to implement the program (for example, because they could not get their schools or teachers to support the program), or they declined because timing of the study was inconvenient (for example, because of changes in leadership at the district level). A few districts were deemed ineligible by the study team after further information about the districts' students and literacy programs became available.[24]

In the nine districts that were recruited for the study, a total of 33 eligible high schools agreed to participate. Among these 33 schools, 17 schools were randomly selected to implement the CLC framework (CLC group), and the remaining 16 schools continued with business as usual (non-CLC group). Randomization occurred immediately after a district's participation in the study was confirmed, so that the developers could start planning the rollout of the intervention in the selected schools as soon as possible. Schools in the CLC group received the professional development and coaching support (paid for by REL Midwest); schools in the non-CLC group received $1,000 to cover the costs associated with data collection.[25]

[24] The number of districts deemed ineligible was less than five.

[25] Originally 37 schools were randomized to CLC and non-CLC conditions. Four schools were small high schools that shared space with other small schools on large campuses. These schools were blocked for randomization separately from any others in the sample (two CLC schools and two non-CLC schools). After randomization, but before the start of program implementation in CLC schools, the small schools were reconsolidated into large high schools. Given that this post-random assignment reconsolidation resulted in the merging together of schools in both experimental conditions, it would not have been possible to determine which students would have gone to which small school were those schools still in existence. The technical advisory group for this project recommended that these schools be considered "non-existent" rather than "no shows" or schools lost through attrition, given that maintaining the schools in the sample would have compromised the internal validity of the study.

The random assignment of schools was conducted separately for different groups of schools (in other words, random assignment was "blocked"). Random assignment blocks are usually school districts, but there are exceptions. In two districts the random assignment of schools was conducted separately for subgroups of schools with similar levels of achievement within each district—yielding two random assignment blocks per district. Another random assignment block is a consortium of two schools in neighboring rural districts. Thus there are 10 random assignment blocks in total (five school districts, four subgroups of schools within two school districts, and one consortium of school districts). The blocking of random assignment improves the power of the study to detect impacts and ensures a better balance in the characteristics of schools in the CLC and the non-CLC groups. In most blocks the study team randomly selected half of the participating high schools to implement the CLC intervention. In any district in which an odd number of schools were recruited into the study, random assignment resulted in one more CLC school than non-CLC school.

It is important to mention two issues that affect the pool of participating schools and students. The first is that during the site recruitment process, a decision was made to relax the eligibility criteria so that districts with some higher performing schools could include all of their schools in the study. This relaxation of the eligibility criteria was done to accommodate district administrators' wishes that more of their schools participate in the study. These higher performing schools were allowed to participate in the study as separate random assignment blocks within the same district, and schools within blocks were randomly assigned to the CLC group and the non-CLC group. Because of this change in the eligibility criteria, however, 12 of the 33 schools recruited into the study did not meet the initially set criterion that at least one-third of their students be "struggling readers." Among the 12 schools that did not meet this criterion, the percentage of nonproficient readers (based on the state's high school ELA test) ranged from 7 percent to 32 percent; among the 21 schools that did meet this criterion, the percentage of nonproficient readers ranged from 37 percent to 87 percent.

The second issue is that although 33 schools participated in the first year of the study (as part of either the CLC or the non-CLC group), not all 33 schools were able to continue for a second year. Five schools withdrew from the study after the first year because of competing school reform priorities or as a result of district restructuring. Therefore, in the second year of the study, the study sample was reduced to 28 schools: 15 schools in the CLC group and 13 schools in the non-CLC group.

In this report all implementation and impact results are based on the 28 schools that participated in both years of the study. As discussed later in this chapter, impacts in the second year of the study provide primary evidence of effectiveness because CLC implementation is expected to be stronger in the second year. By extension the 28 schools that participated in the second year of the study are the "primary school sample" in this evaluation. In order to draw inferences about changes in implementation and impacts over time for this group of schools, findings for the first year of the study are based on these 28 schools and not on all 33 schools that participated in the first year. Readers interested in the first-year impact findings for all 33 schools that participated in the first year of the study are referred to appendix M. Conclusions about CLC's effectiveness in the first year of the study are the same for this larger group of schools.

Characteristics of schools in the study

The characteristics of the study schools prior to their recruitment into the study are compared against the characteristics of other U.S. high schools in table 2.1, which focuses on the characteristics of the 28 study schools that participated in the study for two school years.[26] Characteristics are shown for the 2006/07 school year, because recruitment and random assignment occurred during the course of the 2007/08 school year and summer 2008.

As shown in the table, the study schools enroll a high proportion of students with characteristics associated with low academic performance and educational disadvantage, which is to be expected given the eligibility criteria for the study. In the school year before recruitment, 61 percent of students in the study schools were eligible for Title I services, and 57 percent of students were eligible for free or reduced-price lunch. Fifty-four percent of students were Hispanic or Black.[27] The promoting power of the study schools—a proxy of their graduation rate—was 62 percent.[28] These schools are located predominantly in large or midsize cities. Also, within this group of 28 schools, there is no systematic difference in the characteristics of CLC and non-CLC schools prior to random assignment (table 2.2). This confirms that the two groups of schools were statistically equivalent in expectation before the start of the study, and that the non-CLC group provides a counterfactual for what would have happened in the CLC schools had they not implemented the intervention.[29]

For comparative purposes, information about high schools in the REL Midwest region and the United States that meet the same eligibility criteria as the study schools also is included in table 2.1.[30] Relative to other high-needs schools in the region and country, schools in the CLC study include a higher proportion of students with characteristics associated with low performance or economic disadvantage, as shown in table 2.1. In particular, these schools have lower levels of student promotion, higher percentages of students eligible for free or reduced-price lunch, and higher eligibility rates for Title I funding. In addition schools in the CLC study enroll higher percentages of minority students than other high-needs schools in the Midwest and United States.

[26] See appendix M for the characteristics of the 33 schools that participated in the first year of the study.
[27] Black includes African American, Hispanic includes Latino.
[28] Balfanz and Legters's (2004) measure of promoting power approximates a school's graduation rate. It is calculated as the ratio of the number of grade 12 students in a given school year to the number of grade 9 students from three years prior.
[29] Omnibus tests indicate that for each of the school samples (the 28 schools that participated in both years of the study as well as the 33 schools that participated in year 1), the CLC and the non-CLC groups of schools are not systematically different in terms of their characteristics prior to random assignment. See also appendix M for tables showing the characteristics of the 33 first-year study schools by intervention group (CLC schools compared with non-CLC schools).
[30] The eligibility criteria are the following: at least 25 percent of students qualified for free or reduced-price lunch; the schools enrolled at least 100 students in grade 9, were not charter schools, were defined as regular schools by the Common Core of Data (CCD), and were operational at the time of the CCD report. The criterion that one-third of students be struggling readers was not included in the restrictions for this table because the CCD does not report student achievement data or proficiency information. The 971 eligible Midwest schools represent 709 schools districts in the region (table 2.1).

Table 2.1. Characteristics of CLC study schools and other U.S. high schools in the school year prior to random assignment (2006/07)

Characteristic	Study schools	Eligible Midwest schools[a]	Eligible U.S. schools[b]	All U.S. schools[c]
Title I status (% of schools)	60.7	40.6	48.6	35.3
School average free/reduced-price eligible students	56.9	45.1	48.6	34.5
Race/ethnicity (school average % of students)				
Hispanic	4.3	7.3	20.1	13.9
Black, non-Hispanic	49.2	22.8	24.9	17.5
White, non-Hispanic	42.2	65.8	49.0	62.8
Other	3.7	2.9	5.1	5.0
Male (school average % of students)	50.9	50.2	50.1	50.3
Average total school enrollment	1,387	1,062	1,249	1,252
Percentage of students in grade 9	31.2	28.2	29.4	28.2
Percentage of students in grade 10	25.4	25.1	25.4	25.2
Percentage of students in grade 11	21.6	22.9	22.3	23.0
Percentage of students in grade 12	18.0	20.7	19.9	20.9
Average promoting power (%)[d]	62.0	75.4	71.1	78.4
Average number of full-time teachers[e]	80	59	72	73
School setting (% of schools)				
Large or midsize city	92.9	33.4	32.2	23.9
Urban fringe of a large or midsize city	0.0	20.2	20.8	29.7
Large or small town	0.0	23.2	21.2	18.6
Rural area	7.1	23.3	25.8	27.9
Number of schools	28	971	5,018	9,668

Note: In the "Study schools" column, values are for the 28 schools that participated in both years of the study. Rounding may cause slight discrepancies in calculating sums and differences.

a. Includes schools enrolling grades 9–12 located in states served by the REL Midwest (Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin) that had at least 100 students in grade 9 during 2006/07, had at least 25 percent of students qualifying for free/reduced-price lunch, were not charter schools, were defined as "regular" schools by the Common Core of Data, and were operational at the time of the CCD report.

b. Includes all schools meeting the same eligibility criteria for Eligible Midwest Schools but were located in the 50 states and the District of Columbia (DC).

c. Includes all schools enrolling grades 9–12 in the 50 states and DC that had at least 100 students in grade 9 during 2006/07, were defined as "regular" schools by the CCD, and were operational at the time of the CCD report.

d. Promoting power is calculated as the ratio of grade 12 students in 2006/07 to grade 9 students in 2003/04. The resulting ratio, expressed as a percentage, is a proxy for approximate graduation rate.

e. Data on the number of full-time teachers are available for not quite all of the study schools.

Source: U.S. Department of Education, National Center for Education Statistics (NCES) Common Core of Data (CCD), Public Elementary/Secondary School Universe Survey Data, 2006/07 and 2003/04.

Table 2.2. Characteristics of study schools before random assignment (2006/07), by treatment status

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Title I status (% of schools)	66.7	60.0	6.7	0.272
Free/reduced-price eligible students (school average % of students)	57.5	57.7	−0.2	0.949
Race/ethnicity (school average % of students)				
Hispanic	4.9	3.2	1.7	0.222
Black, non-Hispanic	51.0	52.5	−1.4	0.734
White, non-Hispanic	40.2	39.9	0.3	0.940
Other	3.3	3.8	−0.5	0.614
Male (school average % of students)	50.4	51.6	−1.2	0.129
Average total school enrollment	1,476	1,279	196	0.099
Percentage of students in grade 9	29.7	33.9	−4.2	0.120
Percentage of students in grade 10	26.1	24.6	1.6	0.220
Percentage of students in grade 11	22.2	20.9	1.3	0.211
Percentage of students in grade 12	18.6	16.7	1.9	0.231
Average promoting power (%)	68.4	52.0	16.4	0.085
Average number of full-time teachers[a]	81.2	78.0	3.2	0.562
Test of systematic difference between groups[b] ($\chi 2 = 13.4$)				0.343
Number of schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study, in the school year prior to random assignment (2006/07). Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. Data on the number of full-time teachers are available for not quite all of the study schools.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between schools in the CLC and non-CLC groups, with respect to the characteristics included in this table.

Source: U.S. Department of Education, National Center for Education Statistics (NCES) Common Core of Data (CCD), Public Elementary/Secondary School Universe Survey Data, 2006/07 and 2003/04.

Data sources and measures

The CLC study utilized a variety of data sources to measure the outcomes of students enrolled in the study schools and the background characteristics of these students. Information also was collected to assess the fidelity and quality of CLC implementation, and to measure the nature and quantity of literacy services offered in non-CLC schools as part of business as usual. Following is an overview of the data sources and measures utilized in this study. The data sources in this study and their purpose are summarized in table 2.3.

Table 2.3. Data sources for the CLC study

Data source	Purpose	Measures	Sample
Student outcomes and background characteristics			
GRADE assessment	Impacts on student outcomes	Reading achievement scores (reading comprehension and reading vocabulary subtests)	Students enrolled in the study schools in the spring at the time of testing[a]
Follow-up school records	Impacts on student outcomes	Grades in core courses, credit accumulation in core courses, state test scores, and disciplinary outcomes	Students enrolled in the study schools on the last day of the school year[a]
Historical school records	Descriptive analyses and covariates in the impact analysis	Background characteristics and scores on grade 8 state tests in English and mathematics	Students enrolled in the study schools on the last day of the school year[a]
Implementation fidelity and service contrast			
Implementation data from Action Designs (organization that supported CLC implementation)	Fidelity to the structural elements of the CLC program	Implementation of the structural elements of the CLC program (e.g., receipt and focus of professional development, class scheduling, presence of materials, support from school leadership)	CLC schools

Data source	Purpose	Measures	Sample
Interviews with district staff	Structural fidelity and local context	District-level literacy supports in the study districts (e.g., literacy leadership teams, literacy as a reform goal, literacy-related professional development)	All study districts
Interviews with school staff	Service contrast in structural literacy supports	School-level supports and emphasis on literacy in the study schools (e.g., existence of supplemental reading classes, literacy taught across content areas)	All study schools
Classroom observations	Fidelity to the instructional elements of the CLC program	Constructs measuring the use of CLC instructional strategies and routines by classroom teachers (sequenced instruction, multiple instructional modalities, coherent instruction)	CLC schools[b] (core content[c] and Fusion Reading classrooms)
	Service contrast in classroom instruction	Fidelity constructs (above) and constructs measuring the use of high-quality instructional practices by classroom teachers (comprehension strategies, activating knowledge, and consolidating knowledge)	All study schools[b] (core content classrooms[c] only)

a. Grade 9 classrooms in year 1 of the study, and grade 9 and grade 10 classrooms in year 2 of the study.

b. Grade 9 students in year 1 of the study, and students in grades 9 and 10 in year 2 of the study.

c. The four core content areas are English language arts, social studies, science, and mathematics.

Student outcomes and background characteristics

This study evaluates the CLC intervention's impact on the outcomes of students enrolled in the study schools in the spring of each study year. In the first year of the study (2008/09), the target population includes grade 9 students, and in the second year (2009/10), the target population includes both grade 9 and grade 10 students. Information on the outcomes and background characteristics of these students comes from the three sources: the GRADE reading assessment, follow-up school records, and historical school records.

As noted, the sources of the information about the students' outcomes and background characteristics are as follows:

- **Reading assessment.** The Group Reading Assessment and Diagnostic Examination (GRADE) assessment was used to measure students' reading achievement. The GRADE is a norm-referenced, research-based reading assessment that can be administered to groups. It is meant to be a diagnostic tool to assess the reading skills individuals have and which skills need to be taught.[31] It is used widely to measure performance and to track the growth of an individual student and groups of students from fall to spring and from year to year. It also has been used in numerous U.S. Department of Education evaluations of literacy interventions, including Striving Readers Grants and the Enhanced Reading Opportunities study. The GRADE contains multiple subtests, including two reading comprehension subtests (sentence comprehension and passage comprehension), a listening comprehension subtest, and a vocabulary subtest.[32] The GRADE was administered to students in the spring of each study year (April and May). To ease testing burden, schools were given the option of either testing all of their students or testing a random sample of classrooms. Four schools chose the latter option in the first study year, and five schools chose it in year 2 of the study.[33]
- **Follow-up and historical school records.** Participating school districts were asked to provide end-of-year school records for all students enrolled in the study schools on the last day of each school year (students in grade 9 in year 1, and students in grades 9 and 10 in year 2). These records include course transcript data (credits attempted, credits earned, course marks) and information on attendance, to be used as outcome measures in the impact evaluation. Districts also provided historical school records for these students, such as

[31] *Group Reading Assessment and Diagnostic Evaluation* provides technical information about the GRADE (American Guidance Service, 2001a, 2001b). In this study, Level H (Form A) was administered to grade 9 students, and Form B was administered to grade 10 students. Alternate-form reliabilities are 0.90 for grade 9 students and 0.84 for grade 10 students. The test-retest reliability of the GRADE assessment (Level H) is 0.92 for grade 9 students and 0.90 for grade 10 students. For grade 9 students, the split-half reliability is 0.95 for the reading comprehension subtest and 0.92 for the vocabulary subtest (Level H, Form A). For grade 10 students, the split-half reliability is 0.96 for the reading comprehension subtest and 0.91 for the vocabulary subtest (Level H, Form B) (American Guidance Service, 2001b, pp. 85–87).

[32] In addition to the raw score (total number of items answered correctly), the GRADE also provides standardized scale scores, normal curve equivalent scores, grade equivalent scores, percentile scores, and stanine scores. Standardized scale scores are used for the impact analyses presented in this report.

[33] In these schools, MDRC randomly sampled classrooms of students for GRADE testing from a master schedule of grade 9 and grade 10 classrooms listed during the class period(s) in which testing was scheduled. The number of sampled classrooms was chosen on the basis of the goal of selecting 125–150 students for testing, to ensure that even with absences, at least 100 students per grade would write the exam on the day of testing. See appendix A for details on the procedure for sampling classrooms for GRADE testing.

students' demographic characteristics and their test scores on grade 8 state assessments.[34] These data provide information on students' academic achievement before their participation in the study, which makes it possible to describe the sample to confirm that students enrolled in CLC and non-CLC schools at the end of each implementation year are similar with respect to their background characteristics and prior achievement. These characteristics also are used as covariates in the impact analysis to improve the precision of the estimated impact on student outcomes.

In some of the study districts, passive parental consent was necessary for GRADE testing and the release of school records data; in other districts, parental consent was not necessary (either active or passive).[35] Assent from students was not required in any district. In this report, "consenting students" refers to students whose parents did not object to data collection.

These data make it possible to evaluate CLC's impact on two types of student outcomes that it aims to improve: students' reading achievement and their academic performance in core courses. The outcomes in these two domains are described below, followed by a description of the background characteristics and prior achievement measures that are available from historical student records for describing students in the study.

Reading achievement. A central objective of the CLC framework is to provide students with strategies that expert readers use to understand written texts, which in turn should improve students' reading comprehension. The CLC intervention also provides strategies aimed at helping students break down word meanings through advanced decoding skills and strategies for recognizing word structures (root words, prefixes, and suffixes). Hence, CLC also may improve students' reading vocabulary. Accordingly the following measures were used to examine the intervention's impact on students' reading skills in the spring of each implementation year:

- **Reading comprehension.** A student's average score on the two reading comprehension subtests included in the GRADE (passage comprehension and sentence comprehension) is used to assess the student's reading comprehension skills.
- **Reading vocabulary.** The vocabulary subtest in the GRADE is used to assess whether CLC increases the breadth of words that a student can leverage to understand the subject matter taught in his or her core content classes.

Reading comprehension and vocabulary scores are provided in standard score units by the American Guidance Service, which publishes the GRADE.[36]

Performance in core courses. Another goal of the CLC intervention is to improve students' performance in core content areas (English language arts, social studies, science, and math). The

[34] Historical records were not requested for grade 10 students in year 2 because, for most of these students, this information was available from the collection of historical records for grade 9 students in the first year of the study.

[35] In five school districts in year 1 and four districts in year 2, passive parental consent was required prior to sharing students' school records and administering the GRADE assessment. In these districts, parental notification letters were sent to students' families, alerting them to the research activities and asking them to notify the school if they did not want their child to participate in the study.

[36] Specifically, each student's raw scores on the GRADE subtests and composite scores were converted to standard scores based on national norms for Level H, Grade 9, Spring Testing (American Guidance Service, 2001b, pp. 30–33). Based on these norms, a standard score of 100 on the GRADE reading comprehension or vocabulary test is average for a representative group of students at the end of their year in grade 9. The standard deviation of the standard score for both subtests is 15. A standard score of 85 corresponds, approximately, to the 4.9 grade equivalent.

expectation is that by providing students with the reading strategies that they need to understand complex texts in different subject areas, the CLC framework will help students learn more of the subject matter taught in their core high school courses. In turn students' course grades will improve, which will enable them to earn the credits they need to graduate.[37] The CLC framework also may have an effect on students' course performance independent of its effect on literacy skills because some CLC pedagogical routines and learning strategies are more general and not necessarily literacy focused. In this study the impact of CLC on course performance is assessed using the following two indicators, both of which are derived from students' course transcripts:

- **Credit accumulation in core subject areas.** This measure gauges whether students' performance in their core courses is such that they are making progress toward graduation. The measure is defined as the number of core credits earned cumulatively by the end of the school year as a percentage of the number of core credits required for graduation in a student's district.[38] Thus this measure captures the extent to which students have progressed toward satisfying the graduation requirements in their district.
- **Grade point average (GPA) in core subject areas.** This measure gauges the extent to which students are learning the subject matter taught in their core courses. The measure is defined as students' average grade in core subject areas during the school year and is based on a four-point scale.[39]

Background characteristics and prior achievement. Information also was collected on students' background characteristics and prior achievement, for the purpose of describing the sample and for use as covariates in the impact model. These data come from the historical records provided by the participating school districts, as follows:

- **Background characteristics.** Information was made available on students' demographic characteristics (race, gender, age) and their educational classification in prior school years (special education, eligibility for English as a second language [ESL] services, and eligibility for free or reduced-price lunch).
- **Grade 8 test scores.** As part of the school records request, districts in the study provided students' scores on the grade 8 state assessments in English language arts (or reading) and mathematics. These scores provide information on students' academic achievement before they entered the study schools.
- **Prior attendance rate.** Also available are data on the number of days that students were absent, present, and enrolled in grade 8.

[37] This report focuses on core courses—rather than elective courses—because performance in core courses is an especially important determinant of high school graduation and college readiness.

[38] Course catalogs and student handbooks were used to determine districts' graduation requirements.

[39] Course grades were converted to a common four-point scale in order to make it possible to pool these data across schools. The following conversion was used to convert letter grades: A+/A/A− = 4.0; B+/B/B− = 3.0; C+/C/C− = 2.0; D+/D/D− = 1.0; F = 0.0. School districts using "percentage" grades were asked to provide a method to convert these percentages to letter grades, which made it possible to apply the above conversion in these schools as well. Students' GPA in core subject areas is defined as their average numerical grade across the four core subject areas for which students have a grade (as students did not always enroll in and, therefore, receive a mark in all four core subjects in a given year).

Implementation fidelity and service contrast

In order to contextualize the impact findings, data were collected on two aspects of implementation that can affect the magnitude of impacts on student outcomes: (1) the *fidelity* with which CLC schools implemented the CLC framework's elements as specified by the developers and (2) the *service contrast*, or the extent to which the CLC intervention—as implemented in the CLC schools—differs from the literacy supports and instruction provided in the non-CLC schools.[40]

As described in chapter 1, the CLC framework includes a set of both structural and instructional components. Therefore the fidelity with which schools implemented the CLC intervention is assessed for each of these two types of components, as is the service contrast between CLC and non-CLC schools.

The remainder of this section describes the data used in this study to measure these dimensions of CLC implementation. The measures derived from these data sources are described in greater detail in chapters 3 and 4, prior to the presentation of the related findings. The sample sizes reported in this section are based on the 28 schools that participated in both years of the study.

Structural components of the CLC framework. As described in chapter 1, the CLC intervention includes several key structural components, such as the establishment of Literacy Leadership Teams, the provision of professional development for core content teachers, and the scheduling of supplemental reading classes for struggling readers. Several types of data were used to measure schools' adherence to these CLC structural components, and to gauge how the structural components as implemented differed from what was provided as part of business as usual in the non-CLC schools:

- **Implementation data from Action Designs (structural fidelity).** The adherence of schools to the CLC structural components was assessed using data made available by the developers. Several types of data were systematically collected by Action Designs (the organization that supported CLC implementation) and its six site coordinators.[41] Action Designs created a system of record-keeping tools to monitor and track the implementation of the structural components in CLC schools. These tools provide information about the planning activities undertaken by schools, the number and timing of professional development activities, dates when implementation began and ended at each school, the scheduling of level 3 Fusion Reading classes, and the formation of Literacy Leadership Teams. Site coordinators also maintained records of their monthly professional development visits to schools. Site coordinators documented their visits using a uniform debrief report template created by Action Designs. Each report was completed and returned to Action Designs within a few weeks following each visit to a school.[42] Site coordinators answered standardized open-ended questions and also included additional details, depending on their visit. Uniform data gathered from the reports included frequency of site visits, CLC routines and strategies

[40] Although non-CLC schools probably were not involved in literacy-based school reform efforts, it is likely that they made some efforts at comprehensive school change or literacy support for students as part of business as usual.

[41] Site coordinators acted as both CLC instructors and liaisons between the developers and the schools. They were paired with schools based primarily on geographical proximity. On average, each of the six site coordinators was responsible for 2.7 schools.

[42] Records for completed visits are available for all study schools in both study years.

taught to staff, meetings held with district-level and school-level staff, coaching activities, classroom observations, and supporting and challenging factors to implementation.[43]

- **Interviews with district administrators (implementation context and structural fidelity).** District-level priorities and supports may have facilitated or impeded successful implementation of the CLC framework. Of particular relevance is the extent to which literacy reform is a district-level priority and, if so, the types of support provided to schools to implement and sustain literacy reform. This information is important for understanding the context in which CLC was implemented in the study schools and thus for interpreting impacts on student outcomes. To understand the local context, semi-structured interviews were conducted with district administrators. These interviews provide descriptive information about the same topics as the interviews with school staff but at the district level. Interviews were conducted with the district-level administrator most knowledgeable about these topics.[44] In year 1, project staff interviewed personnel in six of the nine districts that participated in the study; in year 2 of the study, interviews were conducted in seven of these districts.[45] The average length of the interviews was 57 minutes in year 1 and 36 minutes in year 2.[46] The interviews were later transcribed.[47]
- **Interviews with school staff (structural contrast).** Semi-structured interviews with school staff were used to gauge the extent to which the literacy supports in CLC schools differed from those in the non-CLC schools.[48] The interviews focused on three topics: the school's overall instructional priorities, its literacy-focused priorities and initiatives, and its professional development activities during the year. The same interview protocols were used for administrators at both CLC and non-CLC schools, thus making it possible to descriptively compare the findings from interviews at CLC and non-CLC schools.[49] Interviews were

[43] These qualitative data were categorized and quantified using NVivo analytic software.

[44] Positions of district-level personnel interviewed include director of English language arts, director of professional development, associate superintendent of secondary education, and associate superintendent of curriculum and instruction.

[45] In some districts, district-level interviews were not conducted because changes in district personnel prevented the project team from interviewing an administrator knowledgeable of district policy and initiatives related to literacy instruction. In other districts, district-level interviews were not conducted because the district included only one school (in which case, only a school-level interview was conducted).

[46] The length of the interviews ranged from 47 minutes to 66 minutes in year 1, and from 17 minutes to 50 minutes in year 2.

[47] Interviews were first coded qualitatively using NVivo software and then coded again, quantifying the interview data where applicable, in Excel.

[48] Specific questions were provided for interviewers to ask. They asked follow-up questions or provided prompts if respondents did not sufficiently address the questions asked. If parts of subsequent questions were answered previously in the interview, the interviewer either asked for confirmation of the administrator's response to the question or proceeded to the next unanswered question. Because the availability of administrators for this interview often was limited, interviewers frequently were forced to omit nonessential interview questions.

[49] The average length of the interviews was 49 minutes in year 1 (range = 24–77 minutes) and 36 minutes in year 2 (range = 15–55 minutes). These interviews were transcribed, and two members of the research team then coded the interviews by placing interview data in categories with the assistance of NVivo and quantifying the interview data in Excel spreadsheets. The categories used when quantifying the data had been established either through the development of the interview protocol or through the development of categories in NVivo. In order to measure the interrater reliability (IRR) of the categories, both researchers coded the same randomly selected interviews. Two rounds of sampling and coding were conducted. In year 1, four school interviews were sampled and coded in each round (IRR = 0.78 in the first round and 0.91 in the second). In year 2, two school interviews were sampled and coded in the first round (IRR = 0.93), and another five interviews were sampled and coded in the second round (IRR = 0.94). All coding inconsistencies from the two rounds were discussed at length (1) to determine the appropriate code and (2) to resolve enduring inconsistencies. The remaining interviews were split evenly between the two coders and coded separately.

conducted in the spring of each study year with the administrator at each school most knowledgeable about literacy instruction and schoolwide reforms. In year 1 of the study, project staff conducted semi-structured interviews with administrators in 23 of the 28 schools that stayed in the study for both years (11 of 15 CLC schools and 12 of 13 non-CLC schools). In year 2, interviews were conducted in 27 of these schools (all 15 CLC schools and 12 of 13 non-CLC schools). In both study years, the majority of interviewees were either the school principal or the associate principal.[50]

Instructional components of the CLC framework. As described in chapter 1, the CLC intervention also includes a series of instructional components to be used by teachers in the classroom, such as content enhancement routines and embedded strategy instruction. Classroom observations were conducted in both CLC and non-CLC schools to gauge (1) the fidelity with which these instructional elements were implemented in CLC schools (instructional fidelity) and (2) to measure the extent to which instructional practice in content area classrooms in CLC schools differed from instruction in non-CLC schools (instructional contrast).

When observing these classrooms, trained observers used a structured protocol to rate teachers' instructional practices and the characteristics of their classroom's learning environment. The observation protocol developed for this study is called the ACE protocol—for *Activating* knowledge, *Constructing* knowledge, and *Extending* knowledge. The protocol was developed by the CLC evaluation team in collaboration with literacy experts.[51] The instructional practices included in the protocol have been identified in the reading literature as high-quality practices for content instruction and literacy; thus the protocol is general enough to be used in both CLC and non-CLC study schools and provides a tool for comparing literacy instruction in CLC and non-CLC classrooms (instructional contrast). In addition, a subset of items in the protocol is specific to the CLC intervention; these items are used to evaluate the extent to which CLC-specific strategies and routines are used in classrooms in the CLC schools (instructional fidelity). The instructional measures derived from the classroom observations are described in chapter 4.

In order to obtain a representative portrait of core content instruction in the study schools, classrooms were selected randomly for observation from a sampling frame that included mainstream classes and special education inclusion classes in the four core subject areas (English language arts, mathematics, science, and social studies).[52] In CLC schools, Fusion Reading classes also were randomly sampled for observation to assess instructional fidelity in these supplemental reading classes, which are a core component of level 3 services in the CLC

[50] In year 1, the 23 interviewees included 14 principals and 9 assistant/associate principals or curriculum personnel (that is, coaches). In year 2, the 27 interviewees included 16 principals, 7 assistant or associate principals, and 4 other personnel (that is, coaches, librarian, CLC liaison).

[51] Technical information on the development of the ACE protocol and the foundational basis for the protocol is provided in appendix B.

[52] In order to ensure that classrooms from each core content area would be observed, the random sampling of classrooms was stratified by subject area. The number of classrooms sampled per subject area in each school depended on the number of periods in the school's schedule, but in general, a higher number of English language arts and social studies classrooms were selected for observation than other subject areas. Thus, in the analyses presented in this report, observations are weighted to adjust for unequal sampling probabilities across subject areas. The process by which classrooms were sampled is described in greater detail in appendix A.

framework.[53] The classrooms used in the analyses presented in this report are described in a later section of this chapter.

In both study years classroom observers were either former educators who lived near the study sites or research staff knowledgeable about the vocabulary of instruction. All observers had previous experience in conducting observations for purposes of evaluating other teachers or for conducting research. The observers participated in a training session led by the developers of the protocol. A more detailed description of the selection and training of observers is provided in appendix B.

The ACE protocol was piloted in February and March of year 1. Classroom observations in the pilot round were conducted by nine observers in eight study schools (four CLC schools, four non-CLC schools). To determine interrater reliability, 23 core content and reading classrooms were observed by two observers, who made independent ratings while seated in the classrooms. Percentage agreement across all ratings among the observers was 88 percent. However, when the classroom ratings were aggregated into the three broader constructs referred to as "CLC-emphasized pedagogical practices" (see chapter 4 for a description of these measures), the average reliability for those aggregates was 42 percent, (see appendix H and appendix I for more information on fidelity features and reliability). Following the pilot, the research team attempted to provide more clarity to observers regarding definitions of various behaviors in the protocol. However, interrater reliability was not assessed again during year 1. Reliability was retested in the fall of year 2 by having all observers watch a video of classroom instruction and record their observations using the ACE protocol. Average reliability on the aggregate constructs was still only 40 percent.

Student analysis samples

This section describes the samples of students used to estimate impacts on student outcomes in this report. As noted earlier in this chapter the target population for this study consists of all grade 9 students enrolled in study schools during year 1 and grade 9 and grade 10 students enrolled in the study schools during year 2. However, outcome data are not available for all students in the target population for several reasons. First, in districts where passive consent was required, some parents did not consent to the release of their child's records. Second, reading achievement data are missing for some students because of absences on the day of GRADE testing and other factors. This means that impacts must be estimated on the basis of the sample of students for whom data are available on the outcome of interest.[54]

Before describing the analysis samples used in this report, it is important to define what is meant by *grade level* in the impact analysis. As noted earlier this study looks at the impact of the CLC intervention by grade level—that is, on grade 9 students enrolled in the study schools in the spring of year 1, and on grade 9 and grade 10 students enrolled in the spring of year 2. In the

[53] The decision to use classrooms—rather than teachers—as the sampling unit was based on the rationale that teachers' instructional practices can vary depending on the group of students they are teaching. As discussed in the next section, this has implications for the analysis of the contrast in instruction between CLC and non-CLC schools.

[54] See appendix C for a more detailed analysis and discussion of the reasons why outcome data are not available for all students enrolled in the study schools.

impact evaluation, grade levels are based not on students' official grade, but on the number of years that students have been in high school, as follows:

- **Grade 10 sample.** This includes grade 10 students in the analysis sample who are enrolled in the study schools in the spring, as well as all grade 9 students retained from the prior school year.
- **Grade 9 sample.** This includes students in the analysis sample who are new to grade 9 in a given study year.

Thus the grade 9 sample includes students who are in their first year of high school, whereas the grade 10 sample includes students who are beyond their first year of high school.

There are two reasons for defining students' grade level on the basis of time in high school rather than their official grade level. The first is that a student's official grade level does not necessarily reflect the kind of courses that he or she is taking. At the high school level, students who are retained do not retake *all* of their classes, only the ones that they failed. This means that students who are retained in grade 9 will be enrolled in some grade 10 classes. This provides a strong rationale for including retained students with the remainder of their original cohort. The second reason for defining grade levels on the basis of time in high school is that CLC could potentially affect a student's official grade level. In particular, if the CLC intervention increases the number of credits earned by students, it also may affect whether or not students are promoted to the next grade and, by extension, their official grade level. If such an impact were to occur in the first year of the study, then in the second year of the study, the characteristics of students in a given grade level would differ for CLC schools compared with non-CLC schools.[55] In this scenario the impact of the intervention for students in a given grade level (for example, grade 9 students) would be biased because grade 9 students in non-CLC schools would no longer provide the right counterfactual for grade 9 students in CLC schools. A solution to this problem is to define a student's grade level on the basis of his or her time in high school rather than his or her official grade. Because the former is unlikely to be affected by CLC, the intervention's estimated impact on student outcomes is no longer biased when this definition is used.[56]

The remainder of this section describes the student samples used in the evaluation in greater detail, for each outcome domain.[57] The sample sizes reported in this section are for the 28 schools that remained in the study for two school years.[58] It is important that there is no systematic difference between students in CLC and non-CLC schools for any of the analysis

[55] For example, one might expect students officially designated as grade 10 students to be lower achieving in CLC schools than in non-CLC schools because in the former group, the grade 10 sample would include struggling students who were promoted as a result of the intervention. By the same logic, grade 9 students in CLC schools would be higher achieving than grade 9 students in non-CLC schools.

[56] One way in which the CLC intervention could affect time in high school is if it had an impact on dropping out of high school or, in other words, if there were intervention-induced attrition from the sample. In this scenario the characteristics of students still enrolled in schools would differ for CLC schools compared with non-CLC schools, which means that the estimated impact of the intervention on student achievement would be biased. This bias can be assessed by examining whether attrition rates from grade 9 to grade 10 are lower in CLC schools than in non-CLC schools. Based on school records data, attrition does not appear to differ across the two groups of schools: the number of grade 9 students in year 1 who were still enrolled in the school in year 2 does not differ by a statistically significant amount between CLC and non-CLC schools. See appendix C for these results.

[57] See appendix C for further detail on the creation of the analysis samples in this report.

[58] See appendix M for the characteristics of students (grade 9) in all 33 schools that participated in the first year of the study.

samples described below.[59] Thus these samples preserve the balance that was achieved with random assignment, which means that differences in student outcomes between the two groups of schools provide an unbiased estimate of the impact of CLC.[60]

School records sample

The CLC intervention's impact on course performance outcomes (credit accumulation and GPA) is based on the school records sample. This sample includes all students enrolled in the study schools on the last day of the school year, for whom course transcript data were provided by the school district. In year 1, this sample includes 7,365 grade 9 students; in year 2, the school records sample includes 7,951 grade 9 students and 8,514 grade 10 students.[61] Relative to official enrollment numbers reported by schools to the state, the school records sample represents 84 percent of enrolled grade 9 students in year 1 and 89 percent of students enrolled in grades 9 and 10 in year 2.[62]

The characteristics of students in the school records sample by study year and grade level, for the 28 schools that participated in both study years, appear in table 2.4. As seen in this table, the school records sample includes a high proportion of students who have poor academic performance. For example, in the average study school in year 2, 58 percent of grade 9 students achieved proficiency on the state English language arts or reading assessment in grade 8 (or conversely, 42 percent of students did not achieve proficiency on these state assessments). In addition, 29 percent of students were overage for grade when they first started grade 9, suggesting that they had been retained in a prior school year before entering high school. In terms of their demographic characteristics, 66 percent of students in the average study school were eligible for free or reduced-price lunch, and 56 percent were Black or Hispanic.

It is important to note that no systematic difference exists between students in CLC and non-CLC schools with respect to the background characteristics and prior achievement of students in the school records sample. This is true for both study years, and for both grade levels.[63] This indicates that the balance achieved by the experimental design is preserved in the school records

[59] Omnibus tests indicate that for each of the analysis samples, students in CLC and non-CLC schools are not systematically different in terms of their background characteristics and prior achievement. See appendix C for detailed tables of the characteristics of students in the analysis samples for each intervention group (CLC and non-CLC groups).

[60] As described earlier, random assignment ensures that the characteristics of students in CLC and non-CLC schools are statistically equivalent in expectation at the beginning of the study. However, random assignment does not guarantee that this equivalence will be maintained at follow-up in the spring of each study year, after the intervention has been implemented. This equivalence would be compromised, for example, if the CLC intervention were to have an impact on high school dropout rates. Thus it is important to test whether there is a difference between students in CLC and non-CLC schools in the analysis samples.

[61] Impacts on credit accumulation are based on the entire school records sample. Impacts on GPA are based on the subset of students in the school records sample for whom course marks are known. For some students, credit accumulation information is available, but course marks were either not provided or they were coded on the basis of an atypical scheme (not using letter grades A–F or numerical marks). These students' marks are treated as missing. If a student's course marks are missing in all four core subject areas, then this student is not included in the analysis of impacts on GPA.

[62] The percentage of official enrollees included in the school records sample rates does not differ by a statistically significant amount between CLC and non-CLC schools, in either study year or grade level. In school districts where parental consent was not required, the school records sample includes 95 percent of enrolled students in year 1 and 96 percent in year 2; in districts where passive consent was required, the school records sample includes 74 percent of enrolled students in year 1 and 82 percent of students in year 2. See appendix C for a more detailed discussion.

[63] See appendix C for detailed tables of the characteristics of students in the school records sample, by intervention group (CLC and non-CLC group).

sample.[64] Thus differences in course performance between students in CLC and non-CLC schools can be interpreted as the effect of CLC on these outcomes and are not due to preexisting differences in achievement between the two groups of students.

GRADE respondent sample

Information on students' reading achievement comes from the GRADE assessment, which was administered to students in the spring of each study year. However, not all students took the GRADE reading assessment. Therefore the impact of CLC on reading achievement is based on the GRADE respondent sample, which is composed of students for whom a GRADE score is available. In year 1 of the study, the GRADE respondent sample includes 4,786 grade 9 students; in year 2, the GRADE respondent includes 5,011 grade 9 students and 4,546 grade 10 students.

The GRADE respondent sample includes fewer students than the school records sample, primarily due to absences on the day of testing.[65] In the first year of the study, 56 percent of all grade 9 students in the school records sample have a GRADE test score. In year 2, 62 percent of all grade 9 students in the school records sample—and 52 percent of all grade 10 students in the school records sample—have a GRADE test score.[66] These response rates do not differ by a statistically significant amount between CLC and non-CLC schools.[67]

[64] In an experimental design, an important concern is program-induced student attrition. The CLC intervention could potentially affect attrition from the sample (from year 1 to year 2) by reducing the percentage of students who drop out of school. If so, this would compromise the internal validity of the impact findings because grade 10 students enrolled in CLC schools would no longer be comparable to students in non-CLC schools. However, attrition from the sample does not differ between CLC and non-CLC schools. See appendix C for further details.

[65] There are two other causes of GRADE nonresponse. First, many ESL and special education students were not tested because the GRADE cannot be administered to students who require certain types of testing accommodations (for example, alternate language or large-print test forms, audio recordings, an aide, or a translator). Second, in order to ease testing burden, two of the study districts chose the option of testing random samples of their students rather than all students. However, these two factors explain only a small portion of nonresponse. See appendix C for a more detailed discussion.

[66] These response rates are not exactly equal to the ratio of the GRADE respondent sample to the school records sample because some of the former sample is not a perfect subset of the latter. Some students who took the GRADE assessment in the spring (April, May) were no longer enrolled in the district on the last day of the school year. Thus the GRADE respondent sample also includes a small number of students who moved out of the school district during the spring and who are not in the school records sample. See appendix C for details.

[67] See appendix C for a comparison of GRADE response rates in CLC and non-CLC schools.

Table 2.4. Characteristics of students in the school records sample, by study year and grade level

Characteristic	Year 1	Year 2	
	Grade 9	Grade 9	Grade 10
Average age (years)	14.7	14.8	14.8
Overage for grade before entering high school (%)[a]	27.8	29.4	34.0
Free/reduced-price lunch eligible (%)	66.6	66.4	66.2
Race/ethnicity (%)			
Hispanic	5.2	5.6	5.0
Black, non-Hispanic	49.2	50.6	46.7
White, non-Hispanic	40.5	38.2	43.4
Other	5.1	5.6	4.9
Male (%)	51.9	52.2	53.1
English language learner (%)	9.6	9.9	10.7
Special education (%)	19.3	18.0	19.4
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)			
Reading/English language arts	60.5	58.2	61.0
Mathematics	53.6	55.7	54.3
Attendance rate in grade 8 (%)	89.1	89.3	88.8
Sample size			
Number of students[b]	7,365	7,951	8,514
Number of schools	28	28	28

Note: This table is based on the 28 study schools that participated in both years of the study. Values in the table are the observed mean characteristics for students in the school records sample (CLC group and non-CLC group). Values in the Grade 9 columns are based on students who are in grade 9 for the first time. Values in the Grade 10 column are based on grade 10 students and retained grade 9 students. Background characteristics are measured in grade 8 or at the start of grade 9.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9 (high school entry).

b. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full sample of students in the school records sample. The percentage of missing data is highest for grade 8 test scores: 23.2 percent (math) and 24.0 percent (reading) for grade 9 students in the first year; 31.0 percent (math) and 31.7 percent (reading) for grade 9 students in the second year; and 50.3 percent (math) and 51.5 percent (reading) for grade 10 students in the second year. The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.0 percent (age) to 1.6 percent (English language learner status) for grade 9 students in the first year, from 0.4 percent (age) to 2.2 percent (special education status) for grade 9 students in the second year, and from 0.7 percent (age) to 23.2 percent (English language learner status) for grade 10 students in the second year.

Source: Calculations from historical school records data obtained for the Content Literacy Continuum study.

Because a substantial number of students do not have a GRADE test score, it is important to examine whether the GRADE respondent sample is representative of students enrolled in the study schools (as measured by students in the school records sample) because this affects the extent to which the impacts on the GRADE are generalizable to all students in the study schools. Accordingly the characteristics of students in the GRADE respondent sample, by study year and grade level, are presented in table 2.5. As seen in this table, the GRADE sample includes a substantial proportion of students with poor academic achievement; for example, more than one-third (38 percent) of grade 9 students did not achieve proficiency on the state reading assessment in grade 8. However, students in the GRADE respondent sample are higher performing on average than students in the school records sample (see table 2.4). This difference arises because students who did not take the GRADE—absentees and students requiring special testing accommodations—are likely to be lower achieving.[68] Therefore estimated impacts on reading achievement in this report may not be generalizable to all students enrolled in the study schools in the spring, nor to students who did *not* take the GRADE (absentees, ESL, and special education students).

However, it is important also to note that there is no systematic difference between students in CLC and non-CLC schools with respect to the background characteristics and prior achievement of students in the GRADE respondent sample. This is true for both study years and for both grade levels.[69] This means that even though the GRADE respondent sample represents only a subset of the target population, one may still be confident that the difference in test scores between respondents in CLC and non-CLC schools represents the impact of CLC on reading achievement rather than preexisting differences in achievement between the two groups of students.

[68] Statistical tests confirm that in terms of their background characteristics, students who wrote the GRADE are systematically different from students who did not take the test. See appendix C.

[69] See appendix C for detailed tables of the characteristics of students in the GRADE respondent sample, by intervention group (CLC and non-CLC group).

Table 2.5. Characteristics of students in the GRADE respondent sample, by study year and grade level

Characteristic	Year 1	Year 2	
	Grade 9	Grade 9	Grade 10
Average age (years)	14.7	14.7	14.6
Overage for grade before entering high school (%)[a]	22.7	24.6	23.1
Free/reduced-price lunch eligible (%)	65.6	66.0	64.1
Race/ethnicity (%)			
Hispanic	4.8	4.6	4.0
Black, non-Hispanic	45.5	51.7	46.1
White, non-Hispanic	45.9	38.8	46.0
Other	3.9	4.9	3.9
Male (%)	49.9	49.8	49.6
English language learner (%)	10.9	7.9	9.4
Special education (%)	11.3	10.8	12.0
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)			
Reading/English language arts	67.2	62.5	69.3
Mathematics	59.8	60.5	62.4
Attendance rate in grade 8 (%)	91.1	90.7	91.3
Sample size			
Number of students[b]	4,786	5,011	4,546
Number of schools	28	28	28

Note: This table is based on the 28 study schools that participated in both years of the study. Values in the table are the observed mean characteristics for students in the GRADE respondent sample (CLC group and non-CLC group). Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Values in the grade 9 columns are based on students who are in grade 9 for the first time. Values in the grade 10 column are based on grade 10 students and retained grade 9 students. Background characteristics are measured in grade 8 or at the start of grade 9.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9 (high school entry).

b. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full sample of students in the GRADE respondent sample. The percentage of missing data is highest for grade 8 test scores: 21.9 percent (math) and 22.1 percent (reading) for grade 9 students in the first year, 26.4 percent (math) and 26.7 percent (reading) for grade 9 students in the second year, and 35.5 percent (math) and 35.8 percent (reading) for grade 10 students in the second year. The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.9 percent (age) to 2.0 percent (free lunch status) for grade 9 students in the first year; from 0.1 percent (age) to 2.6 percent (special education status) for grade 9students in the second year, and from 0.1 percent (age) to 15.1 percent (English language learner status) for grade 10students in the second year.

Source: Calculations from historical school records data obtained for the Content Literacy Continuum study.

Classroom observation sample

As explained previously in this chapter, classrooms in the study schools were randomly selected for observation as part of the implementation research for this study. This data collection was conducted to provide information regarding the fidelity of implementation of instructional aspects of CLC and to provide some information regarding the service contrast in instruction between CLC and non-CLC schools.

In the first year of the study, grade 9 classrooms were sampled for observation, and in the second year of the study, classes in both grade 9 and grade 10 were sampled because the CLC framework was implemented in both grade levels in its second year. In the spring of year 1, classroom observations were conducted in 27 of the 28 schools that participated in both years of the study.[70] Across these schools, 299 grade 9 core content classrooms were selected for observation (11 classrooms per school on average). In CLC schools, 30 Fusion Reading classes also were sampled (two reading classes per CLC school on average). In the spring of year 2, 312 grade 9 and grade 10 core content classrooms were sampled for observation (11 per school on average). In CLC schools, 46 Fusion Reading classes were sampled (3.1 reading classes per CLC school on average).

The distribution of sampled core content classrooms across subject areas and grade levels is presented in table 2.6.[71] As shown, mathematics sections (representing 28 percent of grade 9 classrooms) are more numerous than other subject areas. Also, in year 2, grade 9 class sections (representing 58 percent of sampled classrooms) are more numerous than grade 10 sections in the schools' schedules.

During the spring site visits, however, not all sampled classrooms were observed. In some cases observers would sit in on a classroom different from the one that they had been intended to see in a particular class period.[72] In these situations the classroom that the observer did see was dropped from the analysis because it was not *randomly* selected for observation; similarly the classroom that the observer was supposed to see is missing (unrated) and cannot be included in the analysis. Thus findings on instructional fidelity and contrast presented in this report are based on sampled classrooms that were "correctly" observed, which are referred to as the "instructional sample."

The instructional sample

The instructional sample is used to measure the fidelity with which teachers in CLC schools implement the instructional routines and strategies of the intervention, and the extent to which

[70] Among these schools, those in the CLC group and the non-CLC group do not differ by a statistically significant amount with respect to their characteristics prior to random assignment. This suggests that even in this restricted sample, one may have a high degree of confidence that differences in instructional practice between the two groups are an unbiased estimate of the impact of the CLC intervention.

[71] These distributions are weighted to adjust for different sampling probabilities across subject areas. Thus they represent the distribution of classrooms across subject areas and grades in the study schools.

[72] In situations in which the regular classroom teacher was absent on the day of observation, observers were instructed to view the substitute teacher instead. Therefore any nonrandom deviations from the intended observation schedule are due to observers having viewed an entirely different class section, rather than to their having observed a substitute teacher.

implementation of these framework components create an instructional contrast between CLC and non-CLC schools. This sample includes all content area classrooms observed in the study schools, plus the Fusion Reading classes observed in CLC schools. In year 1, the instructional sample includes 213 grade 9 core content classrooms across the 28 schools that participated in both study years (representing 71 percent of sampled core content classrooms in year 1) as well as 25 grade 9 Fusion Reading classes observed in the 15 CLC schools (representing 83 percent of sampled reading classrooms in CLC schools in year 1). In year 2, the sample includes 295 grade 9 and grade 10 classrooms across all schools (representing 95 percent of sampled core content classrooms in year 2) and 26 Fusion Reading classes observed in CLC schools (representing 57 percent of sampled reading classrooms in CLC schools in year 2).[73]

Table 2.6. Characteristics of sampled, observed, and unobserved core content classrooms, study schools

Characteristic	Sampled		Observed		Not observed	
	Spring yr 1	Spring yr 2	Spring yr 1	Spring yr 2	Spring yr 1	Spring yr 2
Subject area (%)						
English language arts	23.4	22.2	22.8	23.3	24.7	4.5
Social studies	24.1	23.9	24.6	23.6	22.8	27.8
Science	24.9	25.9	26.4	25.4	21.2	34.4
Mathematics	27.7	28.0	26.2	27.7	31.3	33.3
Grade level (%)						
Grade 9	100.0	57.8	100.0	56.5	100.0	78.5
Grade 10		42.2		43.5		21.5
Sample size						
Classrooms	299	312	213	295	86	17
Schools	27	28	27	28	27	28

Note: This table is based on the 28 schools that participated in both years of the study. In year 1, classroom observations were conducted in nearly all (i.e., 25 or more) of these schools; in year 2, observations were conducted in all schools. Observations are weighted to account for different sampling probabilities across subject areas. Sampled classrooms are those that were randomly chosen for classroom observations. Among these classrooms, most were observed (71% in year 1 and 95% in year 2), but some were not observed (29% in year 1 and 5% in year 2).

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

[73] Observation rates are lower in year 1 because some observers did not upload their classroom observation data into the central observation data system, and these data were lost.

The last four columns of table 2.6 present the distribution of observed and unobserved core content classrooms in the instructional sample across subject areas and grade levels, for schools that participated in both years of the study.[74] In year 1, there is no systematic difference between the distribution of observed and unobserved classrooms across subject areas. In year 2 of the study, however, the distribution of observed and unobserved classrooms differs by a statistically significant amount across subject areas.[75] These results suggest that the instructional findings for year 2 may not be generalizable to all classrooms in the study schools.[76]

In addition to having limited generalizability, the instructional findings in year 2 may not be internally valid. As table 2.7 shows, in year 2, a greater percentage of sampled classrooms was observed in CLC schools than in non-CLC schools (98 percent in CLC schools versus 91 percent in non-CLC schools). This difference of 7 percentage points is statistically significant.[77]

Table 2.7. Classroom observation rates, by study year

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Year 1 (grade 9)				
Sampled classrooms that were observed (%)	68.3	75.5	−7.2	0.514
Sample size				
Sampled classrooms (total = 299)	164	135		
Schools (total = 27)	15	12		
Year 2 (grades 9 and 10)				
Sampled classrooms that were observed (%)	97.7	90.9	6.8	0.016*
Sample size				
Sampled classrooms (total = 312)	168	144		
Schools (total = 28)	15	13		

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. In year 1, classroom observations were conducted in 27 of these schools; in year 2, observations were conducted in all schools. Rounding may cause slight discrepancies in calculating sums and differences. Rates and difference estimates are regression-adjusted for the blocking of random assignment.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

[74] These distributions are weighted to adjust for different sampling probabilities across subject areas.

[75] In Table 2.6 for year 2, observed and sampled classrooms have a more similar grade/subject distribution than not observed and sampled classrooms. This happens because the majority of sampled classrooms were observed (95 percent in year 2). Thus the grade/subject distribution of not observed classrooms does not substantially affect the overall grade/subject distribution of sampled classrooms.

[76] See appendix D for a more detailed analysis of differences between classrooms that were observed and those that were not observed. An omnibus chi-squared test was used to determine whether there is a systematic difference between classrooms that were observed and not observed, with respect to the characteristics included in Table 2.6.

[77] The difference in observation rates in year 1 (7 percentage points) is similar in magnitude to the difference in year 2, but it is not statistically significant because the average observation rate is lower in year 1 (which increases the standard error of the estimate).

This differential response rate raises concerns about whether classrooms observed in non-CLC schools provide an accurate depiction of what instructional practice in the CLC schools would have been in the absence of the intervention, and by extension, whether estimated impacts in year 2 of the study reflect the true effect of the CLC intervention on instruction. For example, consider a scenario in which ineffective teachers were more likely to refuse to be observed in non-CLC schools than in CLC schools (thus giving rise to a lower observation rate in these schools). In this situation the difference in instructional quality between CLC and non-CLC schools would be an underestimate of the true effect of the CLC framework on instruction. Data were not collected on the characteristics of teachers, so it is not possible to determine whether the observed teachers in CLC schools differ from teachers in non-CLC schools with respect to their characteristics. As shown in table 2.8, the distribution of observed classrooms across subject areas (and grade levels in year 2) does not differ by a statistically significant amount between CLC schools and non-CLC schools. However, this does not guarantee that the characteristics of teachers do not differ between the two groups of schools. Thus the difference in instructional practice between these two groups of schools may not provide an accurate estimate of the instructional contrast between CLC and non-CLC schools.

Finally, note that in the second year of the study, all instructional analyses are based on grade 9 and grade 10 classrooms *together*. There are two reasons for pooling across grades in year 2. First, the CLC framework was implemented in both grade 9 and grade 10, but the number of classroom observations could not be doubled and therefore observers had to share their time across two grade levels. This means that conducting the analysis by grade level in year 2 would substantially reduce statistical power.[78] Second, although classrooms are officially categorized as either grade 9 or grade 10 in a school's master schedule, these classes enroll a mix of grade 9 and grade 10 students, so pooling across grade levels provides a better depiction of the average instructional experience of students in a given school day.

Analysis of student impacts and the instructional contrast

This section of the chapter discusses several technical issues that lie at the heart of the evaluation's capacity to produce valid and reliable estimates of the CLC intervention's impact on student outcomes, as well as estimates of the instructional contrast between CLC and non-CLC schools (that is, the effect of the intervention on instructional practice). The section begins by describing the key components of the statistical model used to estimate impacts in this study, and it then reviews the study's sample sizes and the implications for statistical power. The section concludes by discussing the approach taken in this report with regards to multiple hypothesis testing.

Statistical models

Since schools were randomly assigned to implement the CLC framework, impacts on student outcomes can be estimated by comparing the average outcomes of students in CLC and non-CLC schools. Similarly the instructional contrast can be estimated by comparing instructional practice

[78] With the full sample of classrooms (grades 9 and 10) in year 2, the study can detect an impact of 0.29 to 0.39 in effect size, depending on the outcome measure. By splitting the sample by grade level, the minimum effect size that could be detected would increase by a factor of about 1.4 (the square root of 2), to an effect size of 0.41 to 0.55.

outcomes in CLC and non-CLC schools. In practice, a multilevel regression model is used to estimate differences in outcomes between these two groups of schools. This makes it possible to account for the blocking of random assignment by school district and for the fact that the data are clustered.

Table 2.8. Characteristics of observed core content classrooms, by program group

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Year 1 (grade 9)				
Subject area (%)				
English language arts	25.0	20.6	4.4	0.494
Social studies	26.0	24.6	1.4	0.827
Science	25.4	26.7	−1.3	0.844
Mathematics	23.6	28.1	−4.6	0.505
Test of systematic difference between groups ($\chi 2 = 0.5$)				0.919
Sample size				
Classrooms (total = 213)	111	102		
Schools (total = 27)	15	12		
Year 2 (grades 9 and 10)				
Subject area (%)				
English language arts	23.9	24.7	−0.8	0.887
Social studies	23.4	24.9	−1.5	0.781
Science	24.6	24.6	0.1	0.991
Mathematics	28.1	25.8	2.2	0.699
Grade level (%)				
Grade 9	53.3	60.9	−7.6	0.339
Grade 10	46.7	39.1	7.6	0.339
Test of systematic difference between groups [a] ($\chi 2 = 4.85$)				0.303
Sample size				
Classrooms (total = 295)	164	131		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. In year 1, classroom observations were conducted in 27 of these schools; in year 2, observations were conducted in all schools. Observations are weighted to account for different sampling probabilities across subject areas. Rounding may cause slight discrepancies in calculating sums and differences.

a. An omnibus chi-squared test was used to determine whether there is a systematic difference between schools in the CLC group and the non-CLC group with respect to the characteristics included in this table.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

Following are some of the key features of the statistical models used to estimate on student outcomes as well as on instructional practice (that is, the instructional contrast):[79]

- **Block fixed effects.** Because the study is based on school-level random assignment, the impact analysis treats schools as random effects. However, random assignment blocks (school districts or groups of schools within districts) are treated as fixed effects because districts in the CLC study cannot be considered representative of the entire Midwest region due to the purposeful nature of site recruitment. This means that the findings presented in this report should be interpreted as the estimated impact of the CLC intervention for the districts and schools in which these interventions were implemented. In other words, conclusions about CLC's effectiveness cannot be generalized to other schools or districts.
- **Pooled impact.** To estimate the overall impact of CLC in the study schools, the impact of the intervention is first estimated for each random assignment block. These estimates are then averaged together, weighting by the number of CLC schools in each block. This weighting ensures that the pooled result represents the estimated impact of CLC for the average school in the study.[80]
- **Multilevel models.** To properly account for clustering, impacts are estimated using multilevel models. For impacts on student outcomes, a two-level model is used with students nested within schools.[81] When estimating the instructional contrast, a three-level model is used, with classrooms nested within teachers, who are nested within schools.[82]
- **Covariates.** In order to improve the precision of the impact estimates, the analysis controls for random differences between students in CLC and non-CLC schools with respect to the following background characteristics:[83] whether students were overage for grade at the start of grade 9 (and likely to have been retained in a prior grade), students' grade 8 state test scores in reading and math, educational indicators and socioeconomic indicators from prior school years (eligibility for free or reduced-price lunch, ESL status, and special education status), and demographic characteristics (racial/ethnic group and gender). The choice of these covariates was made prior to estimating impacts, based on the fact that they are highly predictive of high school outcomes.[84]

---

[79] See appendix E for further details on the model.

[80] The main impact findings in this report are robust to alternate ways of weighting the pooled impact estimate. See appendix J.

[81] In school districts that chose to administer the GRADE to random samples of classrooms (rather than all students), a three-level model is used when estimating the impact of CLC on reading scores, with students nested within classrooms, nested within schools.

[82] As noted previously, classrooms, not teachers, are the sampling unit. One consequence of this decision is that a teacher could be, and frequently was, sampled and observed more than once (although with a different group of students). In the 28 schools that participated in both years of the study, 29 percent of observed teachers were seen with more than one group of students in year 1 (on average, each of these teachers was observed 2.2 times). In year 2 of the study, 27 percent of observed teachers were seen more than once (2.2 times on average).

[83] In theory it is not strictly necessary to control for these baseline characteristics because random assignment should ensure that students in CLC and non-CLC schools are similar in expectation. However, controlling for students' background characteristics can improve the precision of the impact estimates. As a sensitivity test the statistical model also was specified without controlling for students' background characteristics and prior achievement. See appendix J for unadjusted impact estimates.

[84] Missing information on these characteristics was imputed using a dummy variable approach. This approach consists of (1) imputing a value of "zero" for the missing values in each of the covariates, (2) creating a dichotomous indicator of missingness for each covariate, and (3) including these indicators alongside the imputed covariates in the statistical model

- **Sampling weights.** Both classrooms and students were sampled in this study. For the implementation analysis, classrooms were randomly sampled by subject area in order to ensure that a sufficient number of classrooms in each subject would be observed. Thus, when estimating the instructional contrast, sampling weights are used to account for the fact that some subject areas were oversampled. Also, in five of the study schools, stratified random sampling was used to select classrooms of students for GRADE testing in the second year of the study. Therefore when estimating impacts on GRADE scores in these districts, sampling weights are used to adjust for the fact that some students had a higher probability of being selected.[85]
- **Intent to treat.** As chapter 3 discusses, one school in the CLC group did not implement the intervention. Thus the findings in this report represent the estimated impact of offering schools the opportunity to implement the CLC framework ("intent to treat") rather than the estimated impact of the intervention on schools that did implement the intervention ("treatment on the treated").[86]

Estimated impacts in this report are presented both in their original metric and as an effect size. The effect size is a metric that is widely used for measuring the impact of education programs. It is defined as the impact (effect) of a program divided by the standard deviation of the outcome of interest. In this report effect sizes are calculated by dividing the estimated impact on student outcomes (or the instructional contrast) by the standard deviation of that outcome among students (or classrooms) in non-CLC schools. The standard deviation for the non-CLC group reflects the expected variability of the outcome that one would find in the absence of the CLC intervention. The impact effect size therefore provides an indication of how much the CLC framework improved students' outcomes along the expected variability in those outcomes. For example, an effect size of 0.20 represents an improvement in student outcomes that is equal to 20 percent of the standard deviation of that outcome in non-CLC schools. (A detailed explanation of how outcome levels for the CLC and non-CLC groups are presented in this report appears in box 2.1.)[87]

(Puma, Olsen, Bell, & Price, 2009). The percentage of missing data is highest for grade 8 test scores: 23.2 percent (math) and 24.0 percent (reading) for grade 9 students in the first year, 31.0 percent (math) and 31.7 percent (reading) for grade 9 students in the second year, and 50.3 percent (math) and 51.5 percent (reading) for grade 10 students in the second year. The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.0 percent (age) to 1.6 percent (English language learner status) for grade 9 students in the first year, from 0.4 percent (age) to 2.2 percent (special education status) for grade 9 students in the second year, and from 0.7 percent (age) to 23.2 percent (English language learner status) for grade 10 students in the second year. There is more missing data for grade 10 students because historical school records information is not available for students who were new to the school in grade 10.

[85] Specifically, classroom sampling was stratified by grade level and class size, in order to ensure that a sufficient number of students would be sampled. These schools were spread across fewer than four school districts. See appendix B for details.

[86] The estimated effect of the treatment on the treated can be obtained by dividing the intent-to-treat impact estimates in this report by the percentage of schools in the CLC group that actually implemented the intervention (14 of 15 CLC schools, or 93 percent).

[87] Standard deviations for the student outcome measures—by program group and overall—are presented in appendix L.

Box 2.1 Understanding the impact tables in this report

Many of the tables in this report show the difference between CLC and non-CLC schools with respect to their characteristics, the instructional practices of their teachers, and the outcomes of their students. These tables also present information about the average characteristics and outcomes of students or classrooms in the study schools, to provide context for interpreting the magnitude of the impact findings. The values presented in these tables are the following:

Estimated impact or estimated difference column: this column shows the estimated difference between CLC and non-CLC schools with respect to the characteristics of their students or classrooms (estimated difference) or the outcomes of their students (estimated impact). The values in these columns are obtained by estimating the difference in characteristics or outcomes for each random assignment block, and then weighting each of these estimates by the number of CLC schools in the block. Because of this weighting scheme—which is based on the distribution of CLC schools across blocks—the average estimate presented in the table represents the estimated impact (or difference) *for the average school in the CLC group*. When estimating impacts and differences, a hierarchical linear regression model is used to account for clustering in the data. The statistical significance of the estimated difference is indicated (*) when the *p*-value is less than or equal to 5 percent, based on a two-tailed test. Also note that all estimated impacts are intent-to-treat estimates because one school in the CLC group did not implement the intervention.

CLC group column: this column shows the observed mean outcome for schools in the CLC group. These mean-outcome levels are obtained by weighting the mean-outcome level in each random assignment block by the number of CLC schools in that block. Thus, just as in the impact or difference estimates in the table, the mean outcome reflects the observed distribution of CLC schools across random assignment blocks and represents the mean outcome for the average school in the CLC group.

Non-CLC group column: this column shows the counterfactual—that is, it provides an estimate of what the mean outcome for the average CLC school would have been had it not been randomly assigned to implement the CLC intervention. The values in this column are the mean outcomes for schools in the non-CLC group, which are regression-adjusted to reflect the observed distribution of CLC schools across random assignment blocks.

Effect size: this column shows the estimated impact (or difference) scaled as an effect size. Effect sizes are based on the standard deviation of the outcome for observations (students or schools) in the non-CLC group. This column is included because some outcome measures are based on a scale that is difficult to interpret (for example, test scores).

Statistical significance and multiple hypothesis testing

Statistical significance is a measure of the degree of certainty that one may have that a program's impact is actually non-zero. If an impact estimate is statistically significant, then one may conclude with some confidence that the program really had an effect on the outcome being assessed. If an impact estimate is not statistically significant, then the non-zero estimate is more likely to be a product of chance. In this report, statistical significance is indicated in the tables by an asterisk (*) when the *p*-value of the impact estimate is less than or equal to 5 percent when using a two-tailed test.

When making judgments about statistical significance, however, it is important to recognize the potential problems associated with conducting multiple hypothesis tests. Specifically, conducting hypothesis tests for estimated impacts on several different outcomes increases the likelihood of concluding that a given impact estimate is statistically significant, when in fact the program has no impact (this is known as a *type I error* or a *false positive*).[88] Although it is important to avoid making conclusions based on such errors, the analysis should not be so conservative with respect to producing false positive results that it unduly increases the likelihood of missing true impacts when they exist (that is, relying on *false negative* results, or a *type II error*).

When evaluating the effect of the CLC intervention in this report, two sets of safeguards are used to attenuate the risk of drawing inappropriate conclusions about effectiveness on the basis of statistically significant results that may have occurred by chance.[89] The first safeguard is to confine the analysis to a parsimonious list of student outcomes and to identify a set of "primary" outcomes and subgroups before beginning the impact analysis. All other outcomes and subgroups are considered "secondary" and used either to contextualize the primary impact findings or to generate hypotheses about impacts. The primary outcomes and key subgroups are as follows:

- **Primary outcomes.** Primary evidence of CLC's effectiveness is based on the intervention's estimated impact on two key outcomes: GRADE reading comprehension test scores (a measure of reading achievement) and credit accumulation in core subject areas (a measure of students' performance in their core courses). Students' GRADE vocabulary scores are considered a secondary indicator of effectiveness because vocabulary is not a focal point of the intervention, and students' GPA is a secondary outcome because GPA is not a standardized achievement measure.[90] Subject-specific impacts on course performance also are considered secondary. Although the estimated impact of CLC on credit accumulation and GPA is presented for each core content area separately (English language arts, math, social studies, and science), this information is included only to contextualize the impact of the intervention on all core subject areas combined.
- **Primary study year.** Confirmatory evidence of CLC's effectiveness is based on its impact in the second year of the study, because one would expect impacts (if they exist) to be larger in the second year due to greater program "maturity".

[88] In particular one would expect to see one false positive for every 20 hypothesis tests conducted when $p < .05$ is selected as the criterion for statistical significance.

[89] This approach is based on the recommendations developed by Schochet (2008).

[90] Vocabulary scores are considered a secondary outcome because a fewer number of CLC routines and strategies are focused on vocabulary as compared with reading comprehension. GPA is a secondary measure because it depends on teachers' grading standards and therefore is not a standardized measure across schools.

- **Primary subgroups.** In year 2, estimated impacts for each of the two grade levels (grade 9 and grade 10) are primary because these two groups of students may benefit from the CLC framework for different reasons. Grade 9 students in year 2 received instruction from teachers who were potentially more experienced with the delivery of the CLC strategies, and grade 10 students benefited from having received two consecutive years of CLC services. Impacts for other subgroups of students and schools are examined as part of the exploratory analysis, but they are considered secondary indicators of effectiveness in this study.

This means that there are four primary indicators of effectiveness in this report: estimated impacts on two primary outcomes (reading comprehension scores and core credit accumulation) for two grade levels (grade 9 and grade 10 students) in the second year of the study. As a further safeguard against false positives, *p*-values smaller than 0.05 for any of these four primary impact estimates are adjusted for multiple hypothesis testing based on the procedure described in Hochberg and Benjamini (1990).[91]

Statistical power

An important goal for the design of the CLC study was to ensure that sample sizes would be sufficient to detect program effects (if they exist) that are meaningful to students' lives and relevant to policy debates about the efficacy of literacy interventions. A common way to convey a study's statistical power is through the minimum detectable effect (MDE) or the minimum detectable effect size (MDES). Formally the MDE is the smallest true program impact that can be detected with a reasonable degree of power (in this case, 80 percent) for a given level of statistical significance (in this case, 5 percent for a two-tailed test). The MDES is the minimum detectable effect scaled as an effect size—in other words, it is the MDE divided by the standard deviation of the outcome of interest. In a school-level randomized experiment, the number of schools is a crucial factor that determines the extent to which the impacts on instructional practice and student outcomes can be estimated with enough precision to reject with confidence the hypothesis that the program had no effect. In general, larger sample sizes provide more precise impact estimates.[92]

The MDE and the MDES for the student outcomes in this study, given a sample of 28 schools, appear in tables 2.9 and 2.10. The minimum detectable effects in these tables are based on the standard errors of impact estimates. Hence the values in these tables represent the actual

[91] Based on recommendations by Schochet (2008), adjusted *p*-values were not calculated for impacts on the secondary outcomes (vocabulary scores, attendance, and so on) or for impacts on student or school subgroups—because these analyses are exploratory, as noted earlier.

[92] There are no universally agreed-upon standards for what constitutes small versus large impacts. A meta-analysis of treatment effectiveness studies sheds some light on this issue (Lipsey, 1990). This study found that of 102 studies, most of which were from education research, the bottom third of the distribution of impacts ranged from about 0.0 to 0.32 effect size, the middle third of impacts ranged from 0.33 to 0.55, and the top third of impacts ranged from 0.56 to 1.20. Under these rules of thumb, an MDES of 0.32 would be considered small. More recent work by Hill, Bloom, Black, and Lipsey (2008) suggests that a 0.32 MDES would be considered quite large when placed in the context of the growth in test scores expected over the course of a full year of schooling. Using data from many of the most widely used standardized reading tests, they find that the expected growth for grade 9 students ranges from an effect size of 0.15 to 0.23 for a full year of school; for grade 10 students, the expected growth ranges from 0.03 to 0.35. Documentation for the GRADE assessment indicates that the expected annual growth on the test (as an effect size) for grade 9 students is about 0.07 (American Guidance Service, 2001a, 2001b).

precision of the analyses in this report.[93] Focusing in particular on the four primary outcomes, these tables show the following:

- **Reading comprehension.** In year 2, the study is able to detect an impact on reading comprehension (in effect size) of 0.14 for grade 9 students and an effect size of 0.22 for grade 10 students.[94]
- **Credit accumulation.** In year 2, the study is able to detect an impact on core credit accumulation (in effect size) of 0.25 for grade 9 students and 0.16 for grade 10 students.[95]

Given the short time horizon of this evaluation and the fact that five schools left the study after the first year, it may not be possible to detect impacts on the primary student outcomes of magnitudes that might be anticipated on the basis of previous research on comprehensive school reform. In their meta-analysis, Borman et al. (2003) find that comprehensive school reform programs improve student outcomes by an effect size of 0.23 to 0.50 after five years of implementation. However, because of the complexity of whole-school reforms, impacts in the first few years of implementation are smaller in magnitude—ranging from 0.13 to 0.17 in effect size.[96] The CLC study is able to detect impacts of this magnitude for two of four primary outcomes.

[93] See appendix F for a more detailed discussion of how these values were calculated. Sample sizes are smaller for GPA than credit accumulation because course marks were not available for all students.

[94] The variability in the MDES across grades and outcomes is due to differences in the explanatory power of the model covariates and the size of the intraclass correlation. See appendix K for information on model fit.

[95] As seen in table 2.10, in year 2, the MDE for impacts on credit accumulation is smaller for grade 9 students than for grade 10 students, yet the MDES is larger for grade 9 students. The reason is that the standard deviation of the credit accumulation measure—which is used to convert the effect into an effect size—is smaller among grade 9 students (SD = 9.5 in non-CLC schools) than among grade 10 students (SD = 20.6 in non-CLC schools).

[96] In addition, Borman et al. (2003) conclude that evaluations based on strong study designs—including experiments—show smaller impacts than studies based on weaker designs; effect sizes for strong designs range from 0.18 to 0.21 (see Table 5 in the paper).

Table 2.9. Minimum detectable effect (MDE) and effect size (MDES) for impacts on reading achievement, GRADE respondent sample

Outcome	Number of schools	Number of students	MDE	MDES
Year 1				
Grade 9 sample				
Reading achievement (standard score)				
Reading comprehension	28	4,786	4.05	0.26
Reading vocabulary	28	4,786	2.05	0.13
Year 2				
Grade 9 sample				
Reading achievement (standard score)				
Reading comprehension	28	5,011	2.33	0.14
Reading vocabulary	28	5,011	2.43	0.15
Grade 10 sample				
Reading achievement (standard score)				
Reading comprehension	28	4,546	3.63	0.22
Reading vocabulary	28	4,546	3.13	0.20

Note: This table is based on the 28 schools that participated in both years of the study. The MDE and MDES in this table are calculated from the relevant sample size (number of schools and students) and the standard error of the impact estimate. The minimum detectable effect size is based on the standard deviation of the outcome for students in the non-CLC group of schools. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. The national average for standard scores is 100, and its standard deviation is 15.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table 2.10. Minimum detectable effect (MDE) and effect size (MDES) for impacts on course performance, school records sample

Outcome	Number of schools	Number of students	MDE	MDES
Year 1				
Grade 9 sample				
Credits earned in core subject areas (%)[a]	28	7,365	2.61	0.28
GPA in core subject areas[b]	28	7,315	0.24	0.22
Year 2				
Grade 9 sample				
Credits earned in core subject areas (%)	28	7,951	2.32	0.25
GPA in core subject areas	28	7,917	0.23	0.21
Grade 10 sample				
Credits earned in core subject areas (%)	28	8,514	3.34	0.16
GPA in core subject areas	28	8,209	0.24	0.21

Note: This table is based on the 28 schools that participated in both years of the study. The MDE and MDES in this table are calculated from the relevant sample size (number of schools and students) and the standard error of the impact estimate. The minimum detectable effect size is based on the standard deviation of the outcome for students in the non-CLC group of schools. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students.

a. The cumulative number of credits earned is scaled as a percentage of the number of credits (core or subject-specific) required for graduation in a student's district.

b. GPA in core subject areas is based on a four-point scale: A+/A/A− = 4.0, B+/B/B− = 3.0, C+/C/C− = 2.0, D+/D/D− = 1.0, F = 0.0.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Chapter 3: Implementing the Structural Components of CLC

For successful implementation of the Content Literacy Continuum (CLC), the framework developers at the University of Kansas Center for Research on Learning (KU-CRL) expect schools and districts to participate in CLC planning, establish leadership support for CLC implementation, support professional development for core content area teachers, and establish and support supplemental reading classes for struggling readers. In this report these planning, leadership, professional development, and curricular aspects of implementation are referred to as *structural* components. (Refer to chapter 1 for details about these components.) Fidelity of implementation can be examined by checking for the presence or level of implementation of these components within the high schools that were randomly assigned to implement CLC. As noted in chapter 1, Action Designs, an organization with connections to KU-CRL, provided direct support to the schools and districts to help them with their implementation of CLC.

This chapter has two main purposes. First, it addresses the extent to which these structural components were implemented within the CLC schools (that is, structural fidelity). Second, the chapter examines the degree to which these structural components were found more often in CLC schools than in non-CLC schools (that is, structural contrast). KU-CRL sees the implementation of the structural aspects of the CLC framework as facilitating the implementation of CLC instructional practices by teachers in their classrooms. Instructional fidelity (the extent to which teachers' classroom instruction adheres to principles and practices covered in CLC professional development) are discussed in the next chapter.

This chapter presents several key points:

- For the CLC framework to be implemented with fidelity, the developers of the intervention believe that certain structural components are essential.
- Data gathered through interviews of school leaders and through site coordinators' monthly site visit reports show that it was rare for CLC schools to implement all of the CLC structural components. In the first year 11 of the 15 schools implemented five or fewer of nine structural components at an adequate level or better. Implementation of these components was less successful in the second year, as all 15 schools implemented five or fewer of these components at an adequate level or better.
- Numerous factors at the district and school levels made CLC implementation challenging; these included insufficient resources to secure adequate substitute teacher coverage during professional development and changes in school or district leadership.
- During the first year and second year of full implementation, district-level and school-level support for CLC was evident in most districts but was not universal.
- During both years of the study, each school's professional development on CLC levels 1 and 2 generally focused training on a relatively small number of Content Enhancement Routines and Learning Strategies.

- Not all CLC schools were able to schedule enough sections of Fusion Reading to serve all struggling readers identified as eligible for the course.[97] In the first year, seven of the ten CLC schools for which Fusion eligibility and scheduling data were available scheduled enough Fusion Reading sections to serve their eligible, low-performing grade 9 students (an average of 100 students per school). In the second year three of the nine schools for which the relevant data were available scheduled sufficient Fusion Reading sections to serve these students (an average of 142 students per school).
- Not all schools that created enough Fusion capacity to serve their struggling readers fully utilized that capacity. In year 1, five of the seven schools with adequate Fusion capacity actually enrolled and served the number of Fusion-eligible students in the scheduled sections of the course. In year 2, fewer schools had adequate Fusion capacity (≤ four schools) and full Fusion utilization occurred in fewer than half of these schools.
- The most notable difference between CLC schools and non-CLC schools regarding structural implementation was that CLC school administrators reported more often than their non-CLC counterparts that their schools were supporting student literacy through both content area classes for all students and supplemental reading classes for struggling readers. For other structural components, there appeared to be less contrast between the two groups of schools.

Structural fidelity

The fidelity with which the structural components of the CLC program were implemented was examined for project start-up and two years of full implementation. To describe the fidelity with which CLC-implementing schools put into place the expected structural components of the program, data are presented on district- and school-level support, planning, and leadership activities; the scheduling and implementation of CLC professional development for levels 1–3; the scheduling and implementation of level 3 reading classes; and factors that either supported or did not support CLC implementation.

Implementation of structural components

As discussed in chapter 1, districts and schools new to the CLC framework and its structural components undergo a planning phase. Prior to implementation, Action Designs seeks to understand the literacy needs of a district and its participating schools. District and school personnel learn what CLC can offer to teachers and students; they also learn about their own distinct roles in facilitating implementation. District leaders are responsible for committing the district in writing to support the implementation of the framework in the participating schools. More specifically school districts are expected to provide staff with amenable instructional conditions, including adequate space and other supports; to increase school leader and teacher capacities to develop the program; and to advocate for data-based decision making. Although district-level support is necessary, actual implementation of the program occurs at the school level.

[97] Level 3 of the CLC framework requires that struggling readers receive more intensive literacy instruction through supplemental reading classes: Fusion Reading.

The CLC theory of action (see figure 1.1 in chapter 1) posits that schools need to implement particular structural components to lay the foundation for CLC-aligned instructional practices. CLC structural components central to the integration of the CLC framework into schools, as well as implementation thresholds identified by Action Designs and observed levels of implementation in each study year, are listed in table 3.1. For each component, Action Designs identified both "expected" and "adequate" thresholds for schools to achieve in their implementation of the structural components of CLC.[98] The intention is for the components to be implemented by schools at the expected thresholds under full CLC implementation, which would represent fidelity to the model. However, Action Designs also indicated to the research team that meeting the adequate thresholds could set enough of a foundation (that is, adequate fidelity to the model) for progress to be possible in the implementation of the instructional features of the intervention. Data from the site coordinators' monthly school visit debrief reports and Action Designs's implementation data was used to compute the means across CLC schools for each component. On average, schools were more successful in their implementation of the structural components in the first year compared with the second year.

Planning. Action Designs helped schools plan by conducting five planning activities: conducting a launch call with the district and schools, meeting with the school leadership team, providing a CLC overview session for school faculty, administering a School Profile Survey,[99] and conducting teacher interviews. Of these activities Action Designs identified three as most essential: conducting a launch call, meeting with school leadership, and providing a CLC overview session. On average, schools completed 3.8 planning-phase activities out of the five recommended activities.

The span of time in which schools completed their planning activities varied. A few schools (less than five) initiated their planning phases with Action Designs earlier than the others. They started in October 2007 by holding initial conference calls and hosting school visits. All of these schools experienced approximately 10.1 months of planning between their initial conference calls and the first student day of the 2008/09 school year.[100] Most schools initiated their planning phases later, between December 2007 and May 2008, but did not begin professional development activities with grade 9 teachers until 2008/09. For these schools the initial conference calls ranged from December 2007 to May 2008, and the dates of first school visits ranged from January 2008 to August 2008. These schools averaged approximately 5.0 months of planning between their initial conference calls and the first day of school of the 2008/09 school year, ranging from 2.7 to 8.1 months of planning. Overall, schools' planning phases lasted an average of 6.1 months.

[98] Many of the "expected" thresholds were identified for schools early in the study in the Roles and Responsibilities agreement signed by the principals of the CLC schools when they committed to participation in the project. The "adequate" thresholds were identified by Action Designs after initial implementation began in CLC schools.

[99] The School Profile Survey (SPS) was adapted from an Internal Coherence survey developed by Richard Elmore, professor of educational leadership, Harvard Graduate School of Education, and the Strategic Education Research Partnership. The SPS was designed to capture information about teaching, leadership, and student learning.

[100] For the early planning schools, planning coincided with some initial implementation of professional development for grade 9 teachers in 2007/08.

Table 3.1. Structural fidelity of key CLC components: comparing expected, adequate, and average implementation rates in CLC schools

CLC components	Expected rates recommended by Action Designs	Adequate rates recommended by Action Designs	Average rates across CLC Schools	
			Year 1	Year 2
Planning: completed planning activities[a]	Complete 5 activities	Complete 3 of 5 activities	3.8 activities	na
Leadership: actual number of schools that established or continued Literacy Leadership Teams by year	All CLC schools	All CLC schools	6 schools	9 schools
Leadership: number of meetings between site coordinators and school leadership	10 meetings/year	8 meetings/year	5.3 meetings	3.3 meetings
All levels: professional development (PD) days scheduled per month within the 10-month school year calendar	2–3 days/month	2 days/month	1.9 days/month	1.6 days/month
All levels: total site visit days of PD within the 10-month school year calendar	20–27 days	18 days	18.7 days	15.7 days
All levels: duration of PD months within the 10-month school year calendar	10 months	8 months	7.4 months	6.7 months
Levels 1 and 2: percentage of core content teachers targeted for training, grade 9 teachers in year 1 and grade 9 and grade 10 teachers in year 2	100	80	89.8	75.6
Levels 1 and 2: number of Content Enhancement Routines and Learning Strategies presented to core content teachers	Varies, according to school need	2–3 routines/year	3.1 routines and 0.4 strategies	2.5 routines and 0.6 strategies
Level 3: total days of formal PD received by Fusion teachers	5 days in yr 1; 3 days in yr 2	5 days in yr 1; 3 days in yr 2	3.8 days	1.7 days
Level 3: number of schools that scheduled sufficient Fusion sections to serve all eligible students (capacity)[b]	All CLC schools	All CLC schools	6 schools	$\leq$ 4 schools
Total Schools			15	15

Note: Year 1 refers to school year 2008/09, and year 2 refers to school year 2009/10. The sample of 15 schools in both years represents 14 CLC-implementing schools and one no-show school that did not implement CLC. na = not applicable.

a. The five ideal planning activities include conducting a launch call with the district and schools, meeting with the school leadership team, providing a CLC overview session for school faculty, administering a School Profile Survey, and conducting teacher interviews. The three adequate planning activities are conducting a launch call, meeting with the school leadership team, and providing a CLC overview session.

b. Data for Fusion Reading sections are based on available data provided by Action Designs from 10 schools in year 1 and 9 schools in year 2.

Source: Implementation data and site coordinators' monthly professional development debrief reports provided by Action Designs.

One additional school initiated its planning phase in 2007/08 by engaging in preliminary conference calls and visits, but it did not conduct any further planning or any professional development in that year or during 2008/09 or 2009/10. This school began reconstituting its entire staff during spring 2008. The school district and Action Designs mutually agreed to discontinue professional development efforts at the start of the 2008/09 school year when it became clear that the school would be unable to establish necessary structures. This school is considered a "no-show"; although it did not implement the program, it is still represented in the results presented in this chapter.

Despite variation in the timing of the planning phases across the CLC schools, full implementation of all three levels occurred in all schools during the same two-year period. An overview of the implementation timeline for both years of CLC program implementation is provided in figure 3.1. The timeline includes the formation of Literacy Leadership Teams and the duration of professional development sessions for levels 1–3.

Leadership. Literacy Leadership Teams are one of the fundamental components of the CLC framework. The development and support of Literacy Leadership Teams are a key part of CLC implementation. Schools varied in how they formed these teams. Some schools formed new stand-alone Literacy Leadership Teams consisting of both school administrators and teachers. Others added literacy reform as an objective to school leadership teams already in existence. Six schools formed Literacy Leadership Teams in the first year. In some of these schools, existing school leadership teams also agreed to serve as Literacy Leadership Teams, and the other schools formed new teams. During the second year, 9 schools had Literacy Leadership Teams. In all, 9 of the 15 CLC schools (60 percent) formed Literacy Leadership Teams. In addition to participating in some of the meetings of these Literacy Leadership Teams, site coordinators also met separately with school administrators throughout the school year. The goal of these separate meetings was to keep administrators continuously informed of the implementation process and to address any issues. Meetings often were scheduled on the basis of administrators' availability. On average, site coordinators held 5.3 meetings during year 1 and 3.3 meetings during year 2.

Professional development for core content teachers. During the study the Action Designs site coordinators scheduled monthly visits to provide on-site professional development, modeling, and coaching of core content teachers on Content Enhancement Routines (CLC level 1) and Learning Strategies (CLC level 2). Site coordinators worked with teachers one-on-one and in groups both during and after school hours. Although primarily devoted to training in levels 1 and 2, these visits also included professional development related to Fusion Reading (CLC level 3).

The CLC intervention is phased in at one grade level per year; therefore only grade 9 teachers were offered professional development in the first year. In the second year, both grade 9 and grade 10 teachers were offered professional development.

In the first year site coordinators spent an average of 1.9 visit days per month at each school from September 2008 to June 2009, with an average of 18.7 visit days per school across the 10-month school year (ranging from 0 to 30 days total). In the second year site coordinators spent less time visiting the CLC sites. On average they spent 1.6 days per month from September 2009 to June 2010, with an average of 15.7 visit days per school across the 10-month school year (ranging from 0 to 27 visit days total). CLC professional development training was designed to span the entire school year to facilitate continuous support for teachers. On average professional development occurred over 7.4 months of the first year and 6.7 months of the second year.

Site visits were planned to include a variety of activities to reinforce CLC routines and strategies for teachers. Although site coordinators devoted substantial portions of these visits to training and coaching core content teachers on CLC levels 1 and 2, portions of these visits also were spent supporting CLC level 3 (Fusion Reading), conducting classroom observations, and meeting with school leadership. Thus the average number of site visit days reported in table 3.1 represents multiple activities and not just direct professional development training for core content teachers.

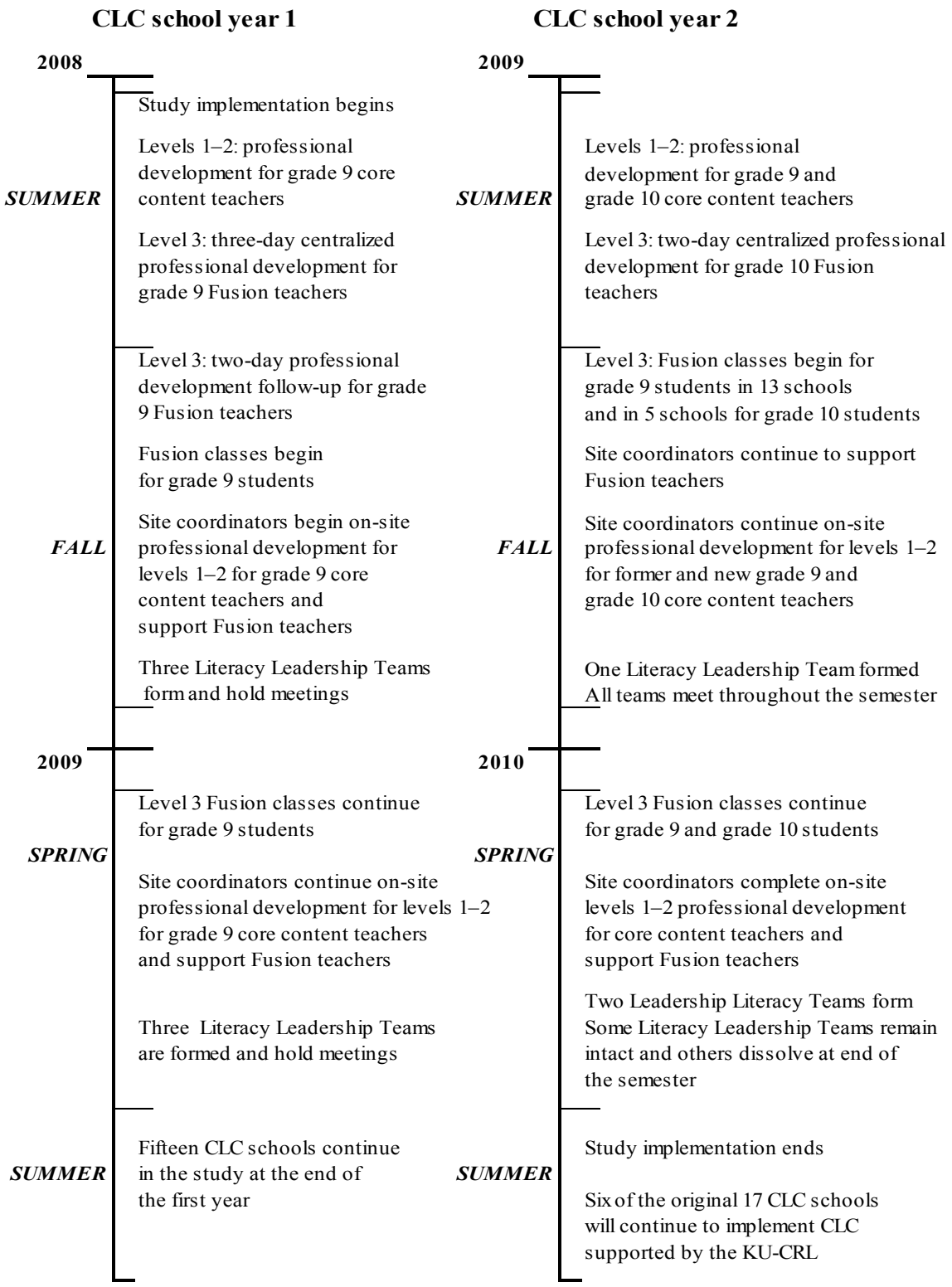
As a literacy-across-the-curriculum intervention, CLC designed its professional development for levels 1 and 2 for teachers from all four core content areas: English language arts, mathematics, science, and social studies. Thus the composition of teachers actually targeted for training would, ideally, reflect the distribution of teachers eligible for training across these core content areas. Schools' master schedules were reviewed to identify teachers of typical grade 9 and grade 10 classes in the four core content areas to determine teachers' eligibility for training in Content Enhanced Routines and Learning Strategies.[101] Self-contained classes serving special education students or English language learners (ELLs) were excluded when defining these core classes.[102] Based on this examination, CLC schools in 2008/09 employed an average of 12.8 grade 9 teachers, and in 2009/10 employed an average of 24.2 grade 9 and grade 10 teachers eligible for professional development in CLC levels 1 and 2 (table 3.2). Eligible teachers were distributed nearly equally across each of the content areas.

[101] These grade 9 core classes and their teachers were identified through consultation with each district and school participating in the CLC study as the basis for conducting classroom observations (see chapter 2).

[102] Action Designs reported that small numbers of grade 9 teachers of self-contained special education or ELL classes received training in CLC levels 1 and 2 within some schools. Training for these teachers was driven by school need, individual teacher interest, and the extent to which special education and ELL teachers worked in inclusion settings with general education core content teachers.

Figure 3.1. CLC implementation timeline

Figure 3.1. CLC implementation timeline



Source: Implementation data provided by Action Designs.

Most of the core content teachers identified as eligible for training in CLC levels 1 and 2 were targeted for training by CLC site coordinators and school leadership. All teachers of grade 9 core content classes in CLC schools were eligible (as well as grade 9 and grade 10 teachers of core content classes in year 2). However, some of these teachers were not targeted for training because the bulk of their teaching load was not grade 9 courses (or grade 9 or grade 10 courses), and a school did not designate them "grade 9 teachers" (or "grade 10 teachers") even if they might have taught one or two sections of grade 9 students. Action Designs could not *require* teacher attendance at professional development trainings, only encourage it. The degree to which teacher attendance was mandatory depended on each local district and/or school administration.

At the beginning of year 1, an average of 11.5 grade 9 core content teachers per school were targeted for training in CLC levels 1 and 2, representing 89.8 percent of the grade 9 core content teachers identified as eligible for training. At the beginning of year 2, an average of 18.3 grade 9 and grade 10 core content teachers per school were targeted for training in CLC levels 1 and 2, representing 75.6 percent of the grade 9 and grade 10 core content teachers identified as being eligible for training.

As with the overall distribution of core content teachers identified as being eligible for training, the teachers actually targeted for training were distributed nearly equally across the content areas. The similarity in distributions between eligible and trained teachers shown in table 3.2 suggests that CLC site coordinators and school leaders attempted to implement CLC across the curriculum in most schools, rather than avoiding content areas where encouraging teachers' uptake of literacy practices might have been more difficult (such as mathematics).

As part of the Action Designs implementation model, site coordinators deliberately focused on a small number of Content Enhancement Routines during the first year in order to encourage teacher buy-in and mastery, as well as to allow site coordinators to gather information about each school's specific instructional needs. Occasionally site coordinators also taught routines or strategies based on teachers' requests for help engaging students with particular content topics. Site coordinators generally planned to train teachers in two to three Content Enhancement Routines during the first year. In addition Action Designs did not expect site coordinators to teach Learning Strategies in all schools during the first year of implementation. Instead site coordinators planned to introduce the Learning Strategies during the second year of implementation after gathering additional information about student needs.

The data provided by CLC site coordinators reveal that they taught an average of 3.1 Content Enhancement Routines and 0.4 Learning Strategies to core content teachers in 2008/09. The number of Content Enhancement Routines taught ranged from 2 to 7 routines, and the number of Learning Strategies taught ranged from zero to 2 strategies. The distribution by routine and strategy and the number of schools that received training in each are provided in table 3.3.[103]

[103] Descriptions of the routines and strategies most frequently introduced at schools can be found in appendix G.

Table 3.2. CLC levels 1 and 2: distribution of core content teachers in year 1 and year 2 receiving professional development across content areas

Content area	Average number of teachers eligible for professional development	Average number of teachers targeted for professional development	Percentage of eligible teachers targeted for professional development
Year 1: grade 9 core content teachers			
English language arts	3.3	2.9	87.9
Mathematics	3.5	3.0	85.7
Science	2.9	2.8	96.6
Social studies	3.1	2.8	90.3
Total	*12.8*	*11.5*	*89.8*
Year 2: grade 9 and grade 10 core content teachers			
English language arts	6.2	4.8	77.4
Mathematics	6.9	4.9	71.0
Science	5.3	4.4	83.0
Social studies	5.8	4.2	72.4
Total	*24.2*	*18.3*	*75.6*

Note: Data shown for all 15 CLC schools.

Source: Implementation data provided for Action Designs used to identify targeted teachers; scheduling and curriculum data provided by CLC schools used to identify eligible teachers.

In the second year site coordinators often revisited level 1 and 2 routines and learning strategies that had been taught to teachers in the first year. The repetition was necessary in order to reinforce the skills with returning staff and also to train new staff. The second year of the study revealed patterns comparable with the first year. The site coordinators taught an average of 2.5 Content Enhancement Routines and 0.6 Learning Strategies to core content teachers. The maximum number of Content Enhancement Routines taught at any school was 5, and in no school were more than 2 Learning Strategies taught. The 2 routines taught most often were the Framing Routine (9 schools) and the Unit Organizer Routine (12 schools). Of the Learning Strategies, the Paragraph Writing Strategy (4 schools) was taught most often.

The Seven-Step Vocabulary Strategy typically taught within Fusion Reading classes (CLC level 3) also was taught to core content teachers. Action Designs reported that select strategies were adopted as level 2 Learning Strategies because of a combination of teacher interest, student need, and site coordinator discretion. Fusion Reading teachers who also taught non-Fusion core classes viewed the Seven-Step Vocabulary Strategy as useful in those classes because of students' struggles with acquiring new vocabulary. Thus CLC site coordinators assisted teachers at fewer

than four schools with embedding the 7-Step Vocabulary Strategy within their core classes during the study.[104]

Support for supplemental reading classes. Prior to the start of each school year, Action Designs worked with CLC schools to assess each school's need for Fusion Reading classes (CLC level 3), which were intended to serve students reading two to five years below grade level.[105] Action Designs then collaborated with schools to identify teachers for Fusion Reading professional development and to schedule sufficient Fusion Reading sections to meet student need. In the first year all CLC schools offered grade 9 Fusion to students. An average of two teachers per school implemented Fusion Reading, and an average of five Fusion sections per school were scheduled for the first year.

Action Designs offered Fusion teachers three days of initial professional development in the first year, beginning in August 2008, focusing on helping teachers establish Fusion classes, introducing Fusion strategies and classroom procedures, and providing teachers with content for the first 10 weeks. In addition Action Designs offered two days of follow-up professional development in October 2008 and the equivalent of at least one day of follow-up training (often informally included within January 2009 site visits). Based on developer data regarding teacher attendance at the August 2008 and October 2008 Fusion trainings, each school's Fusion teachers received an average of 3.8 days of formal professional development by the end of 2008/09 (excluding any professional development received at other times).[106]

In August 2009 professional development training was offered for a new version of Fusion Reading for grade 10 students in 5 CLC schools (Fusion II). These schools committed to providing continued support for previous grade 9 students who took Fusion in the first year. A subset of former Fusion teachers was identified to teach grade 10 Fusion II classes. On average these teachers received 1.7 days of additional formal Fusion professional development. In 2009/10 a total of 5 schools offered Fusion II, and 13 schools continued to offer Fusion to grade 9 students. In the second year an average of two teachers per school implemented Fusion I and/or Fusion II. An average of four Fusion sections and one Fusion II section per school were scheduled.[107]

[104] Source: Action Designs.

[105] Action Designs recommended that CLC-implementing schools use student performance on standardized reading assessments taken between September 2007 and August 2008 in year 1 and September 2008 and August 2009 in year 2 to identify these low-performing students. In the event that such performance data were unavailable, schools were advised to use teacher rankings and/or student self-reports of reading proficiency to identify low-performing students.

[106] Fusion teachers received concentrated initial professional development training because Fusion consists of a highly scripted curriculum. Teachers are supplied with binders to follow that include lesson plans and student exercises. Over the course of the school year, site coordinators continue to coach Fusion teachers as shown in Table 3.6.

[107] In the 2009/10 school year, complete Fusion data provided by Action Designs were available for 13 of the 15 CLC schools based on teacher scheduling assignments.

Table 3.3. CLC levels 1 and 2: Content Enhancement Routines and Learning Strategies in which core content teachers were trained across CLC schools

Name of routine or strategy	Number of schools at which routine or strategy was introduced and supported[a]	
	Year 1	Year 2
Content Enhancement Routines (level 1)		
Framing Routine	14	9
Unit Organizer Routine	14	12
Concept Mastery Routine	7	≤ 4
Course Organizer Routine	6	5
Vocabulary LINCing Routine	5	≤ 4
Question Exploration Routine	≤ 4	≤ 4
Concept Anchoring Routine	≤ 4	≤ 4
Concept Comparison Routine	≤ 4	≤ 4
Clarifying Routine	0	≤ 4
Survey Routine	0	≤ 4
Learning Strategies (level 2)		
7-Step Vocabulary Strategy	≤ 4	≤ 4
Paragraph Writing Strategy	≤ 4	≤ 4
Sentence Writing Strategy	0	≤ 4
Word Mapping Strategy	≤ 4	0
Prediction Strategy	≤ 4	≤ 4
Inference Strategy	0	≤ 4
Bridging Strategy	0	≤ 4
Total number of schools	15	15

Note: Year 1 refers to school year 2008/09, and year 2 refers to school year 2009/10. The sample of 15 schools in both years represents 14 CLC-implementing schools and one no-show school that discontinued CLC implementation. The routines and strategies shown are those implemented in these particular CLC schools but are not inclusive of all routines and strategies in the complete Content Literacy Continuum. The developer's goal is for schools to implement routines and strategies that are most relevant for their staff and students, and not to implement every individual approach that they support.

a. For presentation of findings throughout this report, findings with four or fewer units are masked to protect the confidentiality of those schools, teachers, or students.

Source: Site coordinators' monthly professional development debrief reports provided by Action Designs.

In both versions of the Fusion courses, teachers taught more intensive Learning Strategies geared for struggling readers in comparison with those used in core content classes with students of varying academic abilities. A list of strategies and how many schools received training in each across both grade levels is shown in table 3.4.[108] In the first year the strategies taught most often were the Bridging Strategy (10 schools), the Integration Strategy (7 schools), the Prediction Strategy (5 schools), and Seven-Step Vocabulary Strategy (fewer than 4 schools). In the second year the strategies taught most often were the Bridging Strategy (5 schools) and the Summarization Strategy (4 schools), with the Prediction Strategy and 7-Step Vocabulary Strategy taught next most frequently at fewer than 4 schools each.

Each section of Fusion Reading was designed to serve a maximum of 20 students. Schools were expected to create enough Fusion capacity—that is, to schedule enough class sections of Fusion—to serve all students who were two to five years behind in reading. Based on available developer data, 6 of the 10 CLC schools for which data were available scheduled sufficient Fusion Reading sections to serve their eligible, low-performing grade 9 students (approximately 100 students per school) in year 1.[109] Most of these 6 schools (more than three) actually enrolled and served as many eligible students or more in the scheduled Fusion Reading sections. In year 2 fewer than 4 of the 9 schools for which developer data were available scheduled sufficient Fusion Reading sections to serve their eligible, low-performing grade 9 students (approximately 142 students per school). A subset of these schools actually enrolled and served as many eligible students or more in the scheduled Fusion Reading sections.

Variation in implementation of structural components

The ability of schools to implement the nine structural components[110] listed in this chapter at the developer-defined thresholds provides a gauge for assessing overall implementation fidelity for schools. Schools varied in their success integrating the multifaceted CLC framework into their existing school environment and practices. The distribution of components completed at the expected or adequate level in both years is shown in figure 3.2. In the first year 11 CLC schools implemented five or fewer of the components at an adequate level or better (73 percent). Four CLC schools implemented six or more components at an adequate level or better (27 percent). In the second year the implementation of structural components was less successful. All 15 CLC schools implemented five or fewer of the components at an adequate level or better (100 percent).

[108] Learning Strategies taught in Fusion classes at fewer than four CLC schools include the Paragraph Writing Strategy and the Sentence Writing Strategy. These strategies also are shown in table 3.3 as level 2 Learning Strategies.

[109] In addition to serving students meeting the eligibility criterion of reading two to five years below grade level, some schools served higher performing students within Fusion classes and scheduled sections beyond those required to serve 100 percent of eligible students.

[110] The nine structural components referenced are those shown in table 3.1 for which data were available for all 15 CLC schools. The component describing Fusion Reading sections was omitted from this measure because data were not available for all CLC schools.

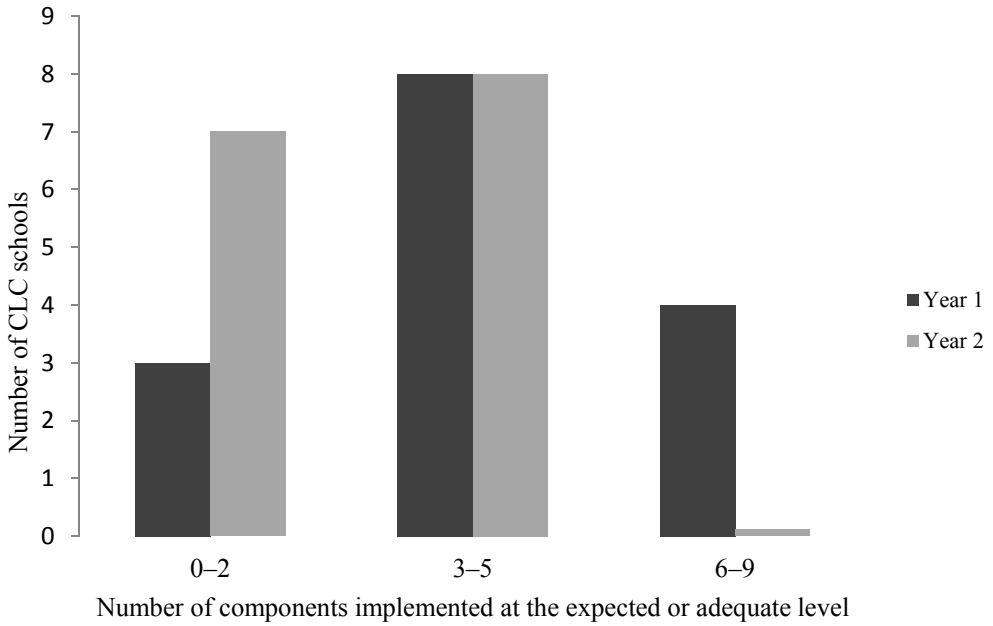
Table 3.4. CLC level 3: embedded learning strategies taught across CLC schools in Fusion classes

Name of strategy	Number of schools at which routine or strategy was introduced and supported	
	Year 1 (Fusion I)	Year 2 (Fusion I and II)
Strategies		
Bridging Strategy	10	5
Integration Strategy	7	≤ 4
Prediction Strategy	5	≤ 4
7-Step Vocabulary Strategy	≤ 4	≤ 4
Summarization Strategy	≤ 4	≤4
Word Mapping Strategy	≤ 4	≤ 4
Inference Strategy	≤ 4	≤ 4
Total number of schools	15	15

Note: Year 1 refers to school year 2008/09, and year 2 refers to school year 2009/10. The sample of 15 schools in both years represents 14 CLC-implementing schools and one no-show school that discontinued CLC implementation. The strategies shown are those implemented in these particular CLC schools but are not inclusive of all strategies in the complete Content Literacy Continuum. The developer's goal is for schools to implement routines and strategies that are most relevant for their staff and students, and not to implement every individual approach that they support.

Source: Site coordinators' monthly professional development debrief reports provided by Action Designs.

Figure 3.2. CLC structural components completed in years 1 and 2



Overall, CLC schools fell short of both the adequate and expected structural thresholds identified by Action Designs in their implementation of the structural components of CLC. The two-year study period represents a relatively short time span for the implementation of a whole-school reform framework such as CLC.[111] In addition circumstances beyond the intervention's design and the support of Action Designs may have limited the development of CLC in some of the schools. Factors that may have affected CLC implementation are discussed in the next section.

Supports and challenges to CLC implementation

Factors internal to a school and factors external to a school at the district level create contexts in which program implementation occurs. Understanding some of these contextual factors can be useful when interpreting program implementation findings. This section of the chapter discusses factors that may have created conditions under which CLC implementation might have been facilitated or better supported, or conditions that may have made implementation more difficult. Variation in school and district contexts may relate to variation in implementation of CLC's structural components across schools. School-level data were gathered from Action Designs site coordinators' monthly site visit reports. Site coordinators independently recorded circumstances observed or reported to them by school staff that either aided or adversely affected the implementation of CLC. District-level data were gathered from interviews conducted with the most knowledgeable district administrator about CLC and literacy reforms. Both site coordinators' monthly records and the administrator interviews were analyzed to create categories of supporting factors and challenges to implementation.

District-level support

As noted previously site coordinators seek assistance with CLC implementation from both district and school leadership because such support can facilitate successful implementation of CLC's structural components. After each of the two years of full implementation, all participating districts were placing some emphasis on literacy reforms in general or on the CLC framework in particular.

After one full year of implementation, district leaders' comments suggest that literacy reform was a goal in the participating districts. Five of the six (83 percent) district administrators interviewed indicated that their districts had at least one major reform effort underway. Of those five, three specifically named literacy as one of, if not their only, major reform effort. Moreover five of the six administrators suggested that literacy or CLC was one of the district's main instructional foci, and all six noted that the district had a formal literacy initiative. As presented in table 3.5, after the second full year of implementation, all seven of the district administrators interviewed indicated that their districts had at least one major reform effort in place, with five of them (71 percent) listing literacy as one of, if not their only, major reform effort. Four of seven (57 percent) administrators named literacy or CLC as a main instructional focus in year 2, and five of seven (71 percent) indicated that the district had a formal literacy initiative in place.

[111] See chapter 1 for a discussion of research on the time it takes to fully implement comprehensive school reform models.

Table 3.5. Academic and literacy support emphases in districts participating in year 2 of the CLC evaluation project

District emphasis	Number of interviewees confirming presence	Number of interviewees asked about emphasis	Percentage
Major reform effort	7	7	100
Literacy as reform	5	7	71
Instructional goals	7	7	100
Literacy/CLC	4	7	57
Formal literacy initiative	5	7	71
Literacy across the content areas	5	6	83
Supplemental reading	4	6	67
Literacy Leadership Team	3	6	50
Literacy-related professional development for teachers	6	7	86

Note: The "confirming presence" column represents the number of administrators who indicated that a particular issue was an emphasis in their district. The "asked about emphasis" column represents the number of administrators who were explicitly asked the relevant question. The percentage represents the number of administrators who discussed a topic divided by the number who were asked about the topic.

Source: District-level interviews conducted for this evaluation of CLC.

When asked about literacy goals for their schools, all district administrators indicated that their districts maintained some literacy-related goals, including improved instruction, higher achievement scores, improved student placement, and successful implementation of literacy programs (for example, CLC and Read 180). In year 1, four administrators (67 percent) stated that literacy skills were supposed to be taught across the curriculum in their districts, and the other two indicated that, within their districts, literacy-across-the-curriculum efforts depended on the school. In year 2, five of six (83 percent) district administrators pointed out that at least some of their high schools should be teaching literacy across the content areas.

District administrator responses varied regarding the establishment of CLC structural components such as school-level literacy teams, literacy-related professional development, and supplemental reading classes. At the end of year 1, all three of the district administrators who were asked explicitly about supplemental reading classes stated that these classes were in place in their high schools. However, only two of the six administrators (33 percent) confirmed the existence of formal literacy teams across schools. Of the six district administrators explicitly asked about supplemental reading classes at the end of year 2, six (67 percent) noted that supplemental reading classes were in place in their high schools. Three of six (50 percent) district administrators confirmed the existence of formal literacy teams across schools. In terms of professional development, all six district administrators indicated that CLC and/or other literacy training were offered across their districts in year 1. Six of the seven district administrators (86 percent) suggested that CLC and/or other literacy training were being offered across their districts in year 2. The presence of such literacy training would be consistent with establishment of CLC structural components.

School-level supports

The factors viewed as supportive overall to CLC implementation that were most commonly reported by site coordinators in their monthly site visit reports are shown in table 3.6.[112] School staff and administrators could display engagement and support for CLC by attending forums for CLC-implementing schools and participating in professional development for teacher leaders. Each year KU-CRL and a partner organization, the Strategic Learning Center, hosted a forum for school leaders across the United States to congregate, share experiences, and learn new approaches to CLC. All CLC schools in the study were invited to participate. In the first year participants from fewer than four schools attended the forum; participants from six schools attended in the second year.

In the second year more schools also sent teachers to the Potential Professional Developers Institute (PPDI) run by KU-CRL. The PPDI trains teacher leaders so that they can facilitate future CLC professional development in their schools. The institute was created as a way to build internal capacity for ongoing local CLC implementation and thus greater sustainability of the framework. Site coordinators also conducted some PPDI training sessions for teacher leaders during their regular on-site professional development visits. Five schools had teachers participating in the PPDI in the second year compared with teachers from less than four schools in the first year. Other schools did not identify teachers interested in attending PPDI trainings.

[112] Although these factors are plausibly relevant to structural implementation, none of them can be conclusively linked to how well or poorly a particular structural component was implemented in CLC schools.

Table 3.6. Factors supportive to CLC implementation in CLC schools

Type of support	Number of schools at which support was reported	
	Year 1	Year 2
Schoolwide supports		
Engagement with other CLC schools[a]	≤ 4	6
Staff participate in Potential Professional Developers Institute[b]	≤ 4	5
Level 1 and 2 supports		
Use of electronic and Web-based resources[c]	8	10
Teachers participate in		
Co-teaching/modeling with site coordinators	5	≤ 4
Coaching with site coordinators	10	7
Classroom observations by site coordinators	12	6
Level 3 supports		
Teachers participate in		
Co-teaching/modeling with site coordinators	≤ 4	≤ 4
Coaching with site coordinators	8	10
Classroom observations by site coordinators	13	12
Total number of schools	15	15

Note: Year 1 refers to school year 2008/09, and year 2 refers to school year 2009/10. The sample of 15 schools in both years represents 14 CLC-implementing schools and one no-show school that discontinued CLC implementation.

a. CLC schools were invited to participate in yearly forums organized by the Strategic Learning Center CLC Institute where practitioners could engage other CLC-implementing districts and schools from across the United States.

b. The Potential Professional Developers Institute offered by the University of Kansas Center for Research on Learning (KU-CRL) trains core content teachers to become CLC professional developers for other staff in their school. The institute helps schools sustain training capacity by developing and mentoring teacher leaders.

c. Use of electronic and Web-based resources include teachers' use of the KU-CRL Teachscape website; GIST graphic-organizer-creator software; and laptop computers and Flipcams, both supplied by Action Designs.

Source: Site coordinators' monthly professional development debrief reports provided by Action Designs.

In professional development sessions for levels 1 and 2, teachers worked both one-on-one and in groups with the site coordinators. Site coordinators reported supporting teachers with co-teaching, coaching, and classroom observations with feedback more frequently in the first year than in the second year. Level 3 professional development sessions had a similar format for how teachers worked with site coordinators. Site coordinators reported providing support to teachers with co-teaching, coaching, and classroom observations similarly in the first year as in the second year.

District-level challenges

The weakening of district stability and support from year to year can present challenges for CLC's integration into schools. Within the two-year study period, six of the nine participating school districts experienced a change in the district administrator who was charged with supporting CLC implementation at the schools. Site coordinators sought to meet with district administrators to provide support and discuss the progress of CLC implementation in participating schools, but the involvement of district administrators in these CLC meetings was not consistent across schools. In the first year site coordinator reports indicate that meetings that included district leadership took place in 7 of the 15 schools. On average these meetings occurred four times during the year. In the second year site coordinator meetings with district leadership were reported for 5 schools. On average these meetings occurred three times during the year.

School-level challenges

The challenges most commonly reported by the site coordinators that may have impeded CLC implementation are shown in table 3.7.[113] The challenges listed here only indicate potential reasons for why CLC was not implemented by schools as intended. Unfortunately, additional data were not collected to provide more insights into schools inability to implement CLC.[114] Action Designs' implementation model is founded on stability of support and cooperation from school leaders and the active engagement of teachers. Seven schools experienced a change in school leaders by the second year.[115] In the first year five of these seven schools implemented five or more structural components (that is, more than half) at an adequate level or better. In the second year four of these seven schools were implementing two or fewer of the key structural components at an adequate level or better. The other schools implemented three to five of these components adequately or better. In addition schools encountered the challenge of maintaining stable staffing. In the second year of implementation, five schools were reconstituted, with all staff being laid off and having to reapply for their jobs alongside additional job candidates. Furthermore, in both years of implementation, a subset of CLC schools had to reduce their teaching force because of financial challenges within the district. During monthly professional development visits, site coordinators also had to be flexible to adjust their planned activities in case staff did not receive substitute coverage to attend. Site coordinators reported that insufficient substitute teacher coverage prevented professional development in seven schools during the first year, and in five schools during the second year. In the first year core content teachers attempting to integrate Content Enhancement Routines and Learning Strategies into their instruction (CLC levels 1 and 2) also reported problems with access to resources for CLC activities, such as the lack of computers, overhead projectors, graphic organizer posters for classrooms, and functioning copy machines to duplicate handouts. By the second year, however, site coordinators did not report those challenges in their monthly site visit reports. Level 3 Fusion teachers faced challenges with Fusion classes starting on time at the beginning of the semester.

[113] Although these factors are plausibly relevant to structural implementation, none of them can be conclusively linked to how well or poorly a particular structural component was implemented in CLC schools.

[114] The data collected do not allow the research team to draw definitive conclusions about the reasons for the lower than intended levels of implementation of structural components of the framework.

[115] School leadership change includes changes in principals, assistant principals, and/or CLC teacher leads.

Students were sometimes not identified for the classes and enrolled into the classes until teaching schedules were finalized (late summer) or in some cases until after the school year had begun.

Table 3.7. Factors challenging CLC implementation in CLC schools

Type of challenge	Number of schools at which challenge was reported	
	Year 1	Year 2
Schoolwide challenges		
Changes in school leadership from year 1 to year 2	na	7
Reconstitution of CLC schools[a]	0	5
Staff reduction in force[b]	≤ 4	≤ 4
Insufficient substitute coverage during professional development[c]	7	5
Level 1 and 2 challenges		
Teacher access to school computers	≤ 4	0
Availability of supplies necessary for CLC activities[d]	6	0
Level 3 challenges		
Fusion classes starting late during semester	5	≤ 4
Under-enrolled Fusion classes	≤ 4	0
Total number of schools	15	15

Note: Year 1 refers to school year 2008/09, and year 2 refers to school year 2009/10. The sample of 15 schools in both years represents the 14 CLC-implementing schools and one no-show school that discontinued CLC implementation. na = not applicable.

a. Reconstitution of CLC schools represents schools where all staff were laid off and given the opportunity to reapply for jobs along with new applicants.

b. Staff reduction in force describes schools where there was a significant shrinkage in school staff.

c. Teachers participating in on-site CLC professional development sessions reported issues with having adequate substitute teacher coverage for their classes.

d. Supplies necessary for CLC activities include classroom overhead projectors, poster-size graphic organizers, and duplication machines to provide copied materials to students.

Source: Site coordinators' monthly professional development debrief reports provided by Action Designs.

In sum, although most schools implemented some of the CLC structural components at adequate levels, the overall implementation of the structural components was incomplete and inconsistent. The structural components, already inconsistently implemented across CLC schools in the first year, were even less evident in the second year. Given that the program model posits that making CLC-related instructional changes is dependent on the establishment of certain preconditions (that is, these structural components), these findings about the low levels of structural fidelity should be considered during the discussions of instructional fidelity and instructional contrast in the next chapter.

Structural contrast between CLC and non-CLC schools

Analytic approach

As discussed in the prior section, the KU-CRL implementation model requires that schools establish several structures to facilitate school-level planning and teacher training. These include establishing the Literacy Leadership Team, setting up Fusion Reading classes and enrolling struggling readers in those classes, setting aside professional development time for the training of teachers in CLC Content Enhancement Routines and Learning Strategies, and training reading teachers in the Fusion curriculum. The degree to which CLC schools were successful in establishing those structures varied.

This section builds on the prior section by investigating whether the implementation of CLC structures resulted in differences from the schools that were not assigned to implement CLC. That is, this section addresses the degree to which some of these structures and schoolwide areas of emphasis in CLC-implementing schools differed from the structures and emphasis areas in non-CLC schools. The information for these comparisons comes from interviews with school-level leaders regarding their schools' approaches to addressing literacy needs of students, the type of professional development that was offered to teachers, schoolwide areas of emphasis, and efforts to align literacy initiatives with other schoolwide initiatives. Interviews included explicit questions regarding the following topics:

- **Literacy across the content areas:** whether the school's literacy programming involves all content area classes
- **Supplemental reading services:** whether the literacy initiatives at the school (if any) also include supplemental support services or classes for struggling readers offered during the school day
- **Continuity of professional development:** whether professional development is literacy focused, and whether professional development is provided during the school year in continuous and focused series (that is, is aligned and provides complementary topics) or addresses multiple topics that are not clearly aligned
- **Sustainability of professional development:** whether professional development activities are used to build school capacity in a self-sustaining manner, thereby increasing internal building and district capacity (that is, provided to train teachers to become trainers themselves for others at the school)
- **Alignment of school improvement efforts:** whether the components of the school improvement plan are complementary (as opposed to operating in isolation or at cross purposes)
- **Working with data to identify instructional priorities:** whether teachers and school administrators use assessment data to identify areas for school-level and classroom-level improvement and to identify students needing supplemental services

If schools that were randomly assigned to implement CLC are implementing the intervention to a high degree of fidelity, then leaders in those schools should respond affirmatively to interview questions regarding the presence of these structures or practices in their schools. Although these

structures and practices are required for CLC implementation, they also may be present among non-CLC schools.[116] Contrast in response frequencies among leaders in CLC schools versus non-CLC schools provides information regarding whether CLC implementation resulted in differences in structures and processes between the two groups of schools.

Findings

The frequencies and percentages associated with school-level interviewees' responses to questions about schoolwide emphasis on literacy and school-level support are presented in table 3.8 (year 1) and table 3.9 (year 2). The right-most column indicates whether the CLC versus non-CLC contrast was statistically significant.[117]

Literacy across the content areas. Several interview questions asked whether schools are addressing literacy by integrating literacy-related activities in all areas of the curriculum. First, school administrators were asked to identify the instructional goals for their staff. Whether or not the administrator mentioned literacy reflects curriculum-wide emphasis on this aspect of learning. Second, interviewees were asked the specific question of whether literacy was taught across content areas. Third, interviewees were asked whether they have a leadership team in place that is focused on the issue of literacy.

Literacy as an instructional goal. When asked about the instructional goals in place within the school during year 1, literacy improvement as a schoolwide goal was mentioned by 55 percent (6 of 11) of interviewees at CLC schools. Forty-six percent (6 of 13) of interviewees mentioned literacy as a goal in non-CLC schools. In year 2, all (15 of 15) interviewees at CLC schools and 83 percent (10 of 12) interviewees at non-CLC schools listed literacy improvement as a schoolwide goal. Differences in the percentages of interviewees mentioning the goal between the CLC and non-CLC interviewees were not statistically significant in either year.

Literacy taught across content areas. For each year, more interviewees from CLC schools than from non-CLC schools mentioned that literacy was taught across content areas. In year 1, the difference between interviewees at CLC schools (11 of 11, or 100 percent) and interviewees at non-CLC schools (8 of 13, or 62 percent) was statistically significant. Year 2 results for the two groups of schools also were statistically significant, as 87 percent (13 of 15) of interviewees at CLC schools indicated that literacy was taught across content areas in their schools compared with less than 33 percent of interviewees at non-CLC schools.

***Presence of a Literacy Leadership Team*.** School administrators were asked whether they had a literacy-focused team made up of faculty from multiple disciplines within the school. Overall 5 of the 25 interviewees in year 1 responded affirmatively to this question. Among interviewees from CLC schools, fewer than 4 of 12 responded affirmatively to the question, and fewer than

[116] The only restrictions placed on schools that agreed to participate in the study but were not assigned to implement CLC is that whatever literacy approach they adopt would not include the combination of supplemental reading classes and formal programming to enhance literacy instruction across the content areas.

[117] Statistical significance is based on chi-square analysis with one degree of freedom (interviewee said *yes* or *no* by type of school [CLC or non-CLC]). Analyses reflect "intent to treat," as one high school assigned to implement CLC was unable to implement the program because of reconstitution during project start-up. Interview data were weighted to account for district size. The *N* for descriptive statistics does not always equal the *N* for the significance testing. For example, for any given question, if only one school is represented in a district with two schools (one CLC, one control), the district is dropped from the statistical test because each assignment is not represented.

4 of the 13 interviewees from non-CLC schools responded affirmatively.[118] Thirteen of 27 interviewees responded affirmatively in year 2. Nine of 15 interviewees from CLC schools and 4 of 12 interviewees from non-CLC schools responded affirmatively. The relationship between response and school type is not statistically significant for either year.[119]

Supplemental reading services. Nearly all of the building administrators were asked if their school had supplemental reading services (for example, separate courses, afterschool support) for students in need of additional reading instruction. For year 1, interviewees in 10 of 12 CLC schools indicated that they had supplemental classes for such students.[120] Fifty percent (6 of 12) of interviewees from non-CLC schools indicated the same. For year 2, 13 of 15 (87 percent) interviewees in most CLC schools indicated that they have supplemental classes for students, and 7 of 12 non-CLC interviewees (58 percent) reported the same. This relationship is not statistically significant for either year.

Continuity of professional development. Interviewees also were asked whether teachers' professional development was cohesive around certain themes. Questions focused on whether professional development was provided as an ongoing series or as single events, whether professional development sessions focused on a single theme, and whether there were professional development sessions that focused on literacy. In year 1, responses from interviewees in CLC schools were not statistically different from responses from interviewees in non-CLC schools. For the question of whether professional development was provided as a series, 50 percent of interviewees from CLC sites responded affirmatively, as did 78 percent of interviewees from non-CLC schools. The percentages of interviewees saying that professional development focused on single themes were less than 25 percent for interviewees from CLC schools and 42 percent for interviewees from non-CLC schools. Finally, 36 percent of administrators from CLC schools stated that teachers took part in literacy-related professional development, compared with 33 percent of administrators from non-CLC schools.

[118] Coding of school administrator interviews demanded specific confirmation of Literacy Leadership Teams in order for a school to be credited for having such a group. Multiple school interviewees noted that leadership teams that were not created specifically for literacy advancement had been assigned greater roles in improving literacy instruction and achievement.

[119] Note the discrepancy between these data from interviews and data supplied by the CLC site coordinators. It might be that the responsibilities of a Literacy Leadership Team (for example, guiding professional development for content area teachers on literacy-related instruction and learning strategies) were subsumed by a broader leadership team within the school. This possibility was not captured by interview questions, which focused on leadership teams devoted to addressing schoolwide literacy only.

[120] Fusion is a supplemental reading component of the CLC treatment. Thus all CLC schools had a supplemental element to their program. It is unclear why any administrators answered negatively to the question.

Table 3.8. Year 1: Aspects of school-level support and emphasis for literacy, per interviews with leaders in participating schools

Contrast aspects	CLC interviewees			Non-CLC interviewees			*p*-value
	Discussed	Asked	Percentage	Discussed	Asked	Percentage	
Literacy across the content areas							
Literacy as an instructional goal	6	11	55	6	13	46	0.652
Literacy taught across content areas	11	11	100	8	13	62	0.017*
Literacy Leadership Team	≤ 4	12	< = 33	≤ 4	13	≤ 30	0.620
Supplemental reading support (i.e., class)	10	12	83	6	12	50	0.052
Literacy across content areas + reading class	9	11	82	≤ 4	11	< 36	0.004*
Continuity of professional development (PD)							
Continuing PD (i.e., a series)	5	10	50	7	9	78	0.235
Singularly focused	≤ 4	12	<33	5	12	42	0.453
Literacy focused	< = 4	11	<36	≤ 4	12	<33	1.000
Sustainability of professional development							
Process of support within school	6	11	55	7	11	64	0.531
District support	9	13	69	8	12	67	0.388
Teacher accountability regarding PD material	6	11	55	7	11	64	0.455
Alignment of school improvement efforts	9	10	90	8	9	89	1.000
Use of data to identify instructional priorities							
Student placement	≤ 4	13	<31	0	11	0	0.041*
School-level strategy development	9	13	69	8	11	73	0.755
Informing instruction	≤ 4	13	<31	≤ 4	11	<36	0.382
Focusing professional development	5	13	38	5	11	45	0.615

**p*-value ≤.05

Note: "Discussed" indicates affirmative response to interview question; "Asked" indicates interviewees who were explicitly asked the question. Chi-squares are based on Yes/No responses among interviewees from CLC and non-CLC schools (1 degree of freedom). These analyses represent "intent to treat." One school classified as a CLC school was unable to implement CLC.

Source: Responses from school administrators to questions during semistructured interviews.

Table 3.9. Year 2: Aspects of school-level support and emphasis for literacy, per interviews with leaders in participating schools

Contrast aspects	CLC interviewees			Non-CLC interviewees			*p*-value
	Discussed	Asked	Percentage	Discussed	Asked	Percentage	
Literacy across the content areas							
Literacy as an instructional goal	15	15	100	10	12	83	0.107
Literacy taught across content areas	13	15	87	≤ 4	12	<33	0.004*
Literacy Leadership Team	9	15	60	≤ 4	12	<33	0.100
Supplemental reading support (i.e., class)	13	15	87	7	12	58	0.125
Literacy across content areas + reading class	12	15	80	≤ 4	12	<33	0.001*
Continuity of professional development (PD)							
Continuing PD (i.e., a series)	6	12	50	≤ 4	8	<50	0.084
Singularly focused	7	15	47	≤ 4	12	<33	0.030*
Literacy-focused	13	15	87	5	12	42	0.029*
Sustainability of professional development							
Process of support within school	11	15	73	6	12	50	0.217
District support	12	15	80	8	12	67	0.550
Teacher accountability regarding PD material	11	15	73	9	12	75	0.843
Alignment of school improvement efforts	8	12	75	≤ 4	6	<67	0.885
Use of data to identify instructional priorities							
Student placement	10	14	71	9	12	75	1.000
School-level strategy development	9	13	69	9	12	75	1.000
Informing instruction	11	15	73	10	12	83	0.543
Focusing professional development	10	14	71	10	12	83	0.632

**p*-value ≤.05

Note: "Discussed" indicates affirmative response to interview question; "Asked" indicates interviewees who were explicitly asked the question. Chi-squares are based on Yes/No responses among interviewees from CLC and non-CLC schools (1 degree of freedom). These analyses represent "intent to treat." One school classified as a CLC school was unable to implement CLC.

Source: Responses from school administrators to questions during semistructured interviews.

In year 2, differences in responses of CLC and non-CLC administrators regarding the singular focus or theme and literacy-focused professional development were statistically different. The percentages of interviewees saying that professional development focused on single themes were 47 percent (7 of 15) for interviewees from CLC schools and less than 33 percent (fewer than 4 of 12) for interviewees from non-CLC schools. Thirteen of fifteen administrators (87 percent) from CLC schools stated that teachers took part in literacy-related professional development, compared with 5 of 12 (42 percent) administrators from non-CLC schools. As in year 1, year 2 results for the provision of professional development as a series were not statistically different, with 50 percent (6 of 12) of CLC interviewees and less than 50 percent (fewer than 4 of 8) of non-CLC interviewees responding affirmatively.

Sustainability of professional development. Responses to school-level interviews showed no statistical differences in year 1 or year 2 between interviewees at CLC schools and non-CLC schools in terms of sustainability of professional development. In the first year, comparable percentages of interviewees at the two types of schools affirmed school-level support for professional development (55 percent and 64 percent for CLC and non-CLC schools, respectively), district-level support for professional development (69 percent and 67 percent for CLC and non-CLC schools, respectively), and teacher accountability for incorporating information from professional development into their teaching practices (55 percent and 64 percent from CLC and non-CLC interviewees, respectively). Similarly, in the second year, comparable percentages of interviewees at the two types of schools affirmed school-level support for professional development (73 percent and 50 percent for CLC and non-CLC schools, respectively), district-level support for professional development (80 percent and 67 percent for CLC and non-CLC schools, respectively), and teacher accountability for incorporating information from professional development into their teaching practices (73 percent and 75 percent from CLC and non-CLC interviewees, respectively).

Alignment of school improvement efforts. Interviewees were asked whether programs, school-wide initiatives, and staff professional development focused on the goals outlined in the schools' annual plan (that is, whether there was alignment in school improvement efforts). In each year, interviewees were asked questions about the extent to which their literacy initiatives aligned with or were complementary to other school goals and initiatives. In year 1, interviewees from 90 percent (9 of 10) of the CLC schools and 89 percent (8 of 9) of non-CLC schools indicated that school improvement efforts were aligned. In year 2, interviewees from 75 percent (8 of 12) of CLC schools and 67 percent (4 of 6) of non-CLC schools indicated that school improvement efforts were aligned. Results were not significantly different in either year.

Use of data to identify instructional priorities. In consideration of the use of data to identify instructional priorities, interviewees from CLC and non-CLC schools expressed no statistically significant differences in either year 1 or year 2 regarding the following: school-level strategy development (69 percent to 73 percent in year 1 and 69 percent to 75 percent in year 2), informing instruction (31 percent to less than 27 percent in year 1 and 73 percent to 83 percent in year 2), or focusing professional development (38 percent to 45 percent in year 1 and 71 percent to 83 percent in year 2). However, in year 1, interviewees from CLC schools were significantly more likely to indicate placing students based on data than were interviewees from non-CLC schools (23 percent to 0 percent). This difference did not persist into year 2, however, as 71 percent of interviewees from CLC schools and 75 percent of interviewees from non-CLC schools indicated that they used data for student placement.

Summary of structural contrast between CLC and non-CLC schools

The one aspect of tables 3.8 and 3.9 that has not been discussed yet is the row labeled "Literacy across content areas + reading class." The recruitment of schools to participate in the study involved finding places where business as usual did not involve both supporting literacy across content areas and providing supplemental reading classes for struggling readers, the combination of which is a key feature of the CLC framework. In other words a clear dichotomy between CLC schools and non-CLC schools should appear when these two fundamental components of CLC are measured together. For an interviewee's response to be given credit (a rating of 1), a positive indication for both the teaching of literacy across content areas and a supplemental reading class was necessary. In year 1, interviewees from 82 percent (9 of 11) of the CLC sites mentioned having both, and less than 36 percent (fewer than 4 of 11) from non-CLC schools mentioned having both. This difference was also apparent in year 2, when 80 percent (12 of 15) of the interviewees from CLC sites indicated having both, and less than 33 percent (fewer than 4 of 12) of the interviewees from non-CLC schools mentioned having both. These differences are statistically significant.

As indicated at the start of this section, the topics investigated about structural contrast through administrator interviews cluster into several categories. In reviewing the topics by cluster, in both years there is little difference between the responses of the CLC and non-CLC administrators around the eight topics at the bottom of table 3.8 and table 3.9: the sustainability of professional development, the alignment of school improvement efforts, and the use of data. For one topic in the first year—the use of data to determine student placement—there is a significant difference, but the numbers of administrators indicating that their schools use data this way are low (fewer than 4 of 13 CLC administrators and none of the 11 non-CLC administrators). For the three topics that relate to the continuity of professional development, no differences exist between CLC and non-CLC administrator responses in year 1. In year 2, however, more CLC administrators than non-CLC administrators reported that professional development activities had a singular focus and were literacy focused ($p<0.05$). While more CLC administrators than non-CLC administrators reported that their schools provided professional development as a related series of sessions, this difference was not statistically significant ($p = 0.08$). For the literacy-related interview topics, there were consistent differences in year 1 and year 2 around literacy being taught across the curriculum and, as mentioned above, the combination of cross-content literacy work with supplemental literacy courses—a key feature of the CLC framework. Although not always statistically significant, more CLC administrators than non-CLC administrators indicated the presence of the literacy-related items in their schools across all items and both years.

Chapter summary

CLC implementation requires the establishment of several structural components within high schools. The focus of this chapter was threefold: to determine whether those structures were put into place, to document factors that may have impeded or supported the creation or maintenance of those structures, and to examine the contrast in the presence of those structures between CLC and non-CLC schools. Data from site coordinators' school visit reports suggest that adequate implementation of the planning phase (for example, the launch call, meetings with district and school leaders, and presentation of the CLC overview) was achieved across implementing schools. Regardless of whether data come from site coordinators' monthly site visit reports or interview data, it is apparent that only a subset of implementing schools established Literacy Leadership Teams. Site coordinators did visit schools and provide professional development on CLC Content Enhancement Routines and Learning Strategies, but across all schools, the time allowed for this professional development was less than the amount considered adequate by KU-CRL.

CLC schools experienced other influences that supported and challenged their ability to implement the CLC framework. Although district officials suggested that literacy reform was a goal, the ability of schools to meet that goal met with obstacles. On average, district officials and school administrators had limited interactions with site coordinators, when such interaction is expected to facilitate implementation. Site coordinators worked consistently with teachers in training and coaching when they had the opportunity to do so. However, not at all teachers could get the necessary dosage of professional development because of changes in staffing, inadequate substitute coverage, and/or lack of supplies for activities.

School administrator interview data indicate that CLC schools differed from non-CLC schools in emphasizing literacy improvement across content areas combined with the provision of a supplemental reading class. However, the contrasts between CLC schools and non-CLC schools for other important structural components (per interviews with school leaders) showed little difference between the two types of schools. The extent to which the topics covered in CLC professional development became integrated within the instruction of core content teachers and reading teachers is the subject of chapter 4.

Chapter 4: Implementing the Instructional Features of the CLC Framework

Whereas chapter 3 focused on whether the schools assigned to implement the Content Literacy Continuum (CLC) framework established the structures deemed necessary for implementation by the developers at the University of Kansas Center for Research on Learning (KU-CRL), this chapter focuses on whether teachers in CLC schools incorporated the topics covered in CLC-related professional development into their instruction. This chapter presents findings on the degree to which classroom observers witnessed the explicit use of specific Content Enhancement Routines and Learning Strategies. Findings also show the degree to which the instruction of teachers within implementing schools reflects the pedagogical practices emphasized in professional development for levels 1–3 of the CLC framework. These pedagogical practices, as mentioned in chapter 1, are sequenced instruction, multiple instructional modalities, and interactive and scaffolded instruction.

This chapter begins with a brief discussion of how classroom observers recorded teachers' use of Content Enhancement Routines and Learning Strategies. The operationalization of the three pedagogical practices emphasized during CLC-related professional development also is described, along with benchmarks for these three practices established by members of the Action Designs implementation team.

Descriptive statistics on the observed presence of Content Enhancement Routines, CLC-specific Learning Strategies, and other learning strategies within the observed *core content classes* in CLC schools and non-CLC schools are provided first. Similar descriptive statistics are provided for *reading* classes in CLC schools as well. The presence of the pedagogical practices within instruction of teachers in CLC-implementing schools then is examined side-by-side with benchmarks established by Action Designs for "adequate" implementation and "exemplary" implementation. All findings are broken out by year of implementation (2008/09 school year and 2009/10 school year). Results are based on the 28 schools that participated in both years of the study; in year 1, classroom observations were conducted in 27 of these schools, and in year 2, observations were conducted in all schools. No statistical comparisons between point estimates for observation outcomes for groups of schools or between observed levels and developer's benchmarks were conducted. Readers are cautioned against drawing strong conclusions from these observational findings.

The key findings presented in this chapter are as follows:

- During year 1, the percentage of observed core content teachers who explicitly used CLC-specific Content Enhancement Routines or Learning Strategies was less than 34 percent. This prevalence is approximately double the percentage of use of routines and strategies witnessed among core content teachers in non-CLC schools. Observed use of CLC-specific routines and strategies among core content teachers in CLC schools was lower in year 2 (11 percent) than in year 1. Nonetheless this prevalence rate in year 2 was double the rate of use among core content teachers in non-CLC schools. These findings are based on descriptive statistics only, no statistical comparisons were made.

- Use of CLC-specific content enhancement routines and strategies during year 1 among reading teachers within CLC schools was observed to be 62 percent. During year 2, this observed use of CLC-specific routines and strategies was about the same (61 percent).
- Observational data from year 1 indicated that instruction among core content teachers in 20 percent of CLC-implementing schools showed adequate fidelity to the pedagogical practices emphasized during CLC professional development. Instruction in none of CLC-implementing schools showed exemplary inclusion of CLC-emphasized pedagogical practices. For year 2, observations of instruction of core content teachers revealed that 100 percent of CLC-implementing schools failed to meet Action Designs' benchmark for adequate fidelity to CLC-emphasized pedagogical practices.

Operationalizing fidelity of instruction

Prior to presenting findings regarding how well teachers incorporate principles from CLC professional development into their instruction, the specific indicators of CLC-based instruction are described. Benchmarks established by KU-CRL for adequate fidelity and exemplary fidelity also are described.

Constructing indicators of CLC-laden instruction

As mentioned in chapter 2, former educators and research staff with experience in observing classroom instruction were trained to use the ACE protocol (*Activating* knowledge, *Constructing* knowledge, and *Extending* knowledge) to record the presence of particular features within teachers' instruction. These features include the following actions and devices emphasized during CLC-related professional development:

- Use of graphic organizers in instruction
- Use of a named strategy or instructional routine, whether or not it is part of CLC
- Mention of the purpose of the instruction for the class period
- Performance of a CLC-aligned literacy activity
- Teacher's checking for understanding
- Blending of oral and written discourse
- Use of interactive discourse between teacher and students
- Teacher's scaffolding of instruction during the class period

These instructional features were combined to create indicators of the three pedagogical practices emphasized across CLC professional development for all three levels. Each of the three indicators was scored 0–4, based on whether the components making up the indicators were present during an instructional period.

Sequenced instruction reflects teachers' use of the "Cue-Do-Review" instructional cycle emphasized in CLC-related professional development. Components include setting a purpose for instruction (Cue phase), leading a CLC-aligned literacy activity (Do phase), and checking for students' understanding (Review phase). A fourth component reflected whether the other three

components were done in order (that is, setting purpose first, followed by CLC-aligned literacy activity, with checking understanding last). Presence of any of these components during a class period received a code of "1"; absence of the components was coded as "0." Scores for the components were summed to create the score for sequenced instruction for a particular class period.

Multiple instructional modalities reflect teachers' use of a combination of verbal, graphical/visual, and written modalities within a single instructional period. Components of the multiple instructional modalities construct include the following: whether a named learning strategy was used (presence during class period = 1, 0 otherwise), whether a graphic organizer was used (presence = 1, 0 otherwise), blending of oral and written discourse (1 if both used, 0 otherwise), and presence of all four modalities (strategy, graphic organizer, oral discourse and written discourse; 1 = all present, 0 otherwise). Scores for the four components were summed for each class period, yielding scores that could range from 0 to 4.

Interactive and scaffolded discourse reflects the degree to which teachers attempt to co-construct knowledge with students through interactive discourse and then gradually release students to acquire and/or apply knowledge independently. Teachers receive scores ranging from 0 to 4 for this pedagogical practice as well. Presence of teacher-student interactive discourse during a class period was coded as 1 (0 otherwise), and occurrence of teacher-student interactive discourse throughout the *entire* class period also was coded as 1 (versus 0 if discourse was never present or present in only a subset of coding segments). Teachers' scaffolding of instruction at any point during the class period was coded 1 (0 otherwise), and continued scaffolding of instruction across all observational segments within a class period was coded as 1 as well (0 otherwise).

Developer's benchmarks for indicators of CLC-laden instructional practices

For each of the three indicators of CLC-emphasized pedagogical practices, the developers of the CLC framework established benchmarks, or cut points, which signify whether teachers in each of their classes are demonstrating less than adequate amount of the practice, an adequate amount of the practice, or an exemplary amount of the practice.[121] Per the developer, teachers who received observation scores of less than 2 for a practice are demonstrating little of the practice and are not facilitating students' understanding of content in the manner recommended in CLC professional development. Teachers scoring between 2 and 3 for an indicator were considered by the developer to have demonstrated an adequate amount of that practice (that is, an amount of the practice that was viewed by the developer as potentially sufficient to influence the ways that students learn content). Teachers who received a score higher than 3 for a practice were classified as "exemplary" for that practice (that is, such teachers were viewed by the developer as demonstrating strong fidelity to that CLC-emphasized pedagogical practice, and according to Action Designs, were very likely to be influencing students' learning of content).

By averaging the indicators across class periods within a school, it is possible to create a schoolwide indicator of whether CLC-emphasized pedagogical practices are being integrated within instruction. The same cut points for adequate and exemplary can be used to assess the degree to which CLC-implementing schools are using the CLC-emphasized pedagogical practices.

[121] Project partners at Action Designs established these cut points by reviewing the components of each indicator and the possible range for each indicator. They were unaware of schools' actual scores or the range of actual scores.

Findings

The following sections summarize what observers saw in the classrooms. First, the degree to which teachers used Content Enhancement Routines and Learning Strategies within their instruction is presented. The second set of analyses focuses on the degree to which CLC-emphasized pedagogical practices were present in classrooms within CLC-implementing schools. These latter findings examine the presence of these practices against the previously mentioned benchmarks established by KU-CRL. Each summary of findings separately examines the presence of pedagogical practices in core content classes and in reading classes for both year 1 observations and year 2 observations.

Use of CLC's Content Enhancement Routines and Learning Strategies in core content classrooms

One way to examine the degree to which core content teachers incorporated CLC-related Content Enhancement Routines and Learning Strategies within their instruction is to record the degree to which teachers explicitly name and use the routines and strategies during the observed class periods. This information provides an indication of the *saturation* of these routines and strategies within the classrooms of CLC schools. The amount of *contrast* also can be examined by viewing the percentage of classrooms in CLC schools in which these routines and strategies are used alongside the percentage of classrooms using the routines and strategies in non-CLC schools. This information for core content classrooms is presented in table 4.1 (information from year 1 observations) and table 4.2 (information from year 2 observations). Statistical comparisons on observational outcomes between the two groups of schools or between observations made in year 1 and year 2 were not conducted.

Year 1. During the year 1 observations, teachers in 34 percent of the observed core content classes incorporated a graphic organizer into their instruction. Although this percentage falls short of the level that would show maximum saturation (that is, 100 percent), it still is double the percentage of use of graphic organizers found in non-CLC schools (17 percent). Observations also indicated the presence of CLC Content Enhancement Routines and Learning Strategies in 22 percent of the classes observed, which was double the percentage of classrooms in non-CLC schools in which CLC routines and strategies were evident (11 percent).[122] Much of the contrast reflects teachers' use of the FRAME routine within CLC schools (21 percent) compared with non-CLC schools (9 percent). Whereas CLC-related routines and strategies were more prevalent in classrooms within CLC schools, more teachers in non-CLC schools incorporated other types of learning strategies into their instruction than did teachers in CLC schools (33 percent for non-CLC versus 21 percent for CLC).

[122] Only speculation is possible about why observers may have witnessed CLC-specific content enhancement routines and learning strategies within non-CLC schools. Possible reasons include (1) district-level staff attending Action Designs's professional development may have shared routines and strategies across both types of schools (that is, contamination), (2) teachers who were transferred from CLC schools to non-CLC schools during planning or full implementation phases may have used CLC routines or strategies within their classes, (3) teachers within non-CLC schools may have inadvertently used names of CLC routines and strategies for the routines and strategies that they independently used within their courses, and (4) observers may have misidentified instructional techniques as CLC routines or strategies.

Table 4.1. Percentages of year 1 grade 9 core content CLC and non-CLC classrooms using Content Enhancement Routines and Embedded Learning Strategies

Devices, routines, and strategies	CLC classrooms (*N* = 131)	Non-CLC classrooms (*N* = 127)
Any graphic organizer	34	17
Presence of CLC routines or strategies	22	11
Concept Routine	6	8
FRAME/Framing Routine	21	9
Planning/Organizing Routine	6	10
Question Exploration Routine	6	9
Vocabulary LINCing Routine	<2	<2
Bridging Strategy	2	0
Paragraph Writing Strategy	0	<2
Prediction Strategy	2	8
Word Maps—morphology/decoding	<2	0
Seven-Step Vocabulary Process	0	0
Sentence Writing Strategy	0	0
Summarizing/Summarization Strategy	0	0
Presence of other learning strategies	21	33
Brainstorming	0	4
KWL	0	3
Question-Answer Relationship (QAR)	5	3
Quick Writes	3	6
Reciprocal Teaching	<2	2
Response Journals	3	6
Structured Overview	7	11
Think Aloud	3	8
Webbing	0	<2
Word Maps—concepts	2	3
Other	16	19

Note: This table is based on the 27 schools where classroom observations could be conducted in year 1 (of the 28 that participated in both years of the study). Observations are weighted to account for different sampling probabilities across subject areas.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

Table 4.2. Percentages of year 2 grade 9 and grade 10 core content CLC and non-CLC classrooms using Content Enhancement Routines and embedded Learning Strategies

Devices, routines, and strategies	CLC classrooms (*N* = 131)	Non-CLC classrooms (*N* = 127)
Any graphic organizer	32	37
Presence of CLC routines or strategies	11	6
Concept Routine	<2	<2
FRAME/Framing Routine	7	0
Planning/Organizing Routine	0	0
Question Exploration Routine	<2	2
Vocabulary LINCing Routine	<	3
Bridging Strategy	0	0
Paragraph Writing Strategy	0	0
Prediction Strategy	0	0
Word Maps—morphology/decoding	0	0
Seven-Step Vocabulary Process	0	0
Sentence Writing Strategy	0	0
Summarizing/Summarization Strategy	0	0
Presence of other learning strategies	20	23
Brainstorming	0	<2
KWL	0	0
Question-Answer Relationship (QAR)	6	13
Quick Writes	2	4
Reciprocal Teaching	0	<2
Response Journals	2	<2
Structured Overview	0	0
Think Aloud	10	2
Webbing	<2	0
Word Maps—concepts	0	0
Other	5	6

Note: This table is based on the 28 schools that participated in both years of the study. Observations are weighted to account for different sampling probabilities across subject areas.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

Year 2. Observations of core content classrooms during year 2 showed that graphic organizers were used by 32 percent of teachers in CLC schools and by 37 percent of core content teachers in CLC schools. Observational data also indicate that the prevalence of the use of CLC Content Enhancement Routines and Learning Strategies within CLC schools were still nearly double that observed in classrooms in non-CLC schools (11 percent versus 6 percent). Again the most frequently observed CLC routine or learning strategy was the FRAME routine, observed in 7 percent of classrooms. Observation of instruction in CLC schools revealed use of other types of learning strategies in 20 percent of classrooms, compared with 23 percent in non-CLC schools).

Core content teachers' use of routines and strategies over time. Observation data on prevalence rates of Content Enhancement Routines and Learning Strategies in core content classrooms in CLC schools from year 1 to year 2 indicate *less* inclusion of these routines and strategies in year 2 than in year 1. In year 1, these routines and strategies were apparent in 22 percent of observed core content classrooms. The inclusion of CLC routines and strategies in year 2 was about half as prevalent (11 percent of core content classrooms included these routines and strategies).

Use CLC's Content Enhancement Routines and Learning Strategies in reading classes

Observations of reading classes during year 1 and year 2 provide information of saturation of CLC devices, instructional routines, and learning strategies by reading teachers trained in the Fusion curriculum. This information is provided in tables 4.3 and 4.4. These findings are summarized for each year.

Year 1. Observations of reading classes in year 1 revealed that instruction in 31 percent of the classes within CLC schools (that is, Fusion Reading classes) included use of a graphic organizer. Sixty-two percent of the observed reading classes in CLC schools included use of CLC's Content Enhancement Routines and Learning Strategies. This use of CLC-prescribed routines and strategies among teachers of Fusion Reading was more prevalent than among the core content teachers within the same schools, which was 22 percent usage. Roughly one in four of the Fusion Reading teachers who were observed included learning strategies within their instruction that were not part of the CLC Learning Strategy Curriculum.

Year 2. Observations of instruction in reading classes in CLC schools during year 2 indicated slightly greater prevalence of graphic organizer use than during year 1 (31 percent in year 1, 42 percent in year 2). Use of CLC Content Enhancement Routines and Learning Strategies remained roughly the same during year 2 (62 percent) as in year 1 (62 percent). Observations of reading classes in CLC schools showed less use of non-CLC learning strategies in year 2 than in year 1 (15 percent versus 28 percent).

Table 4.3. Percentages of year 1 grade 9 Fusion Reading classrooms where graphic organizers, Content Enhancement Routines, and embedded Learning Strategies were observed[123]

Devices, routines, and strategies	Reading classrooms in CLC schools (*N* = 29)
Any graphic organizer	31
Presence of CLC routines or strategies	62
Concept Routine	<10
FRAME/Framing Routine	<10
Planning/Organizing Routine	<10
Question Exploration Routine	10
Vocabulary LINCing Routine	0
Bridging Strategy	34
Paragraph Writing Strategy	<10
Prediction Strategy	29
Word Maps—morphology/decoding	0
Seven-Step Vocabulary Process	10
Sentence Writing Strategy	0
Summarizing/Summarization Strategy	0
Presence of other learning strategies	28
Brainstorming	0
KWL	0
Question-Answer Relationship (QAR)	0
Quick Writes	<10
Reciprocal Teaching	<10
Response Journals	<10
Structured Overview	<10
Think Aloud	10
Webbing	0
Word Maps—concepts	0
Other	21

Note: This table is based on the 15 CLC schools that participated in both years of the study. Observations are weighted to account for different sampling probabilities across subject areas.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

[123] Data on use of graphic organizers, instructional routines, and learning strategies in reading classes in non-CLC schools are not provided due to lack of reading classes in many of the non-CLC schools.

Table 4.4. Percentages of year 2 grade 9 and grade 10 Fusion Reading classrooms where graphic organizers, Content Enhancement Routines, and embedded Learning Strategies were observed

Devices, routines, and strategies	Reading classrooms in CLC schools (*N* = 26)
Any graphic organizer	42
Presence of CLC Routines or Strategies	62
Concept Routine	0
FRAME/Framing Routine	<11
Planning/Organizing Routine	0
Question Exploration Routine	0
Vocabulary LINCing Routine	<11
Bridging Strategy	35
Paragraph Writing Strategy	0
Prediction Strategy	12
Word Maps—morphology/decoding	<11
Seven-Step Vocabulary Process	23
Sentence Writing Strategy	0
Summarizing/Summarization Strategy	0
Presence of other learning strategies	15
Brainstorming	0
KWL	0
Question-Answer Relationship (QAR)	0
Quick Writes	0
Reciprocal Teaching	0
Response Journals	0
Structured Overview	0
Think Aloud	<11
Webbing	0
Word Maps—concepts	0
Other	12

Note: This table is based on the 15 CLC schools that participated in both years of the study. Observations are weighted to account for different sampling probabilities across subject areas.

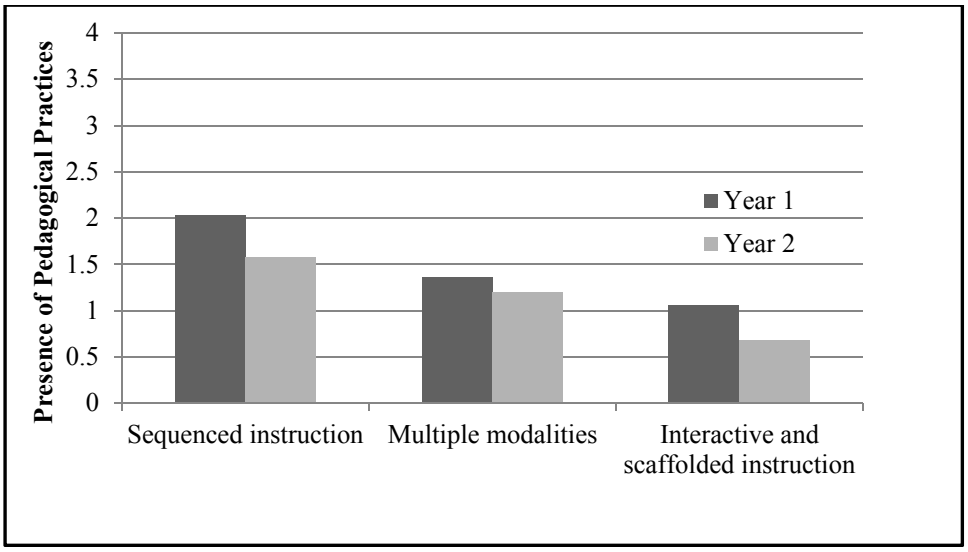
Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

Use of CLC-emphasized pedagogical practices among core content teachers

Observational data on the three pedagogical practices that are emphasized during CLC professional development (that is, sequenced instruction, multiple instructional modalities, and interactive and scaffolded instruction) were analyzed. Caution is warranted in placing too much confidence in these aggregated findings, since interrater reliability estimates calculated on pilot observations and observations of film clips were low (see appendix I for more information).[124] The average levels of use of these practices by core content teachers in CLC schools during year 1 and year 2 are presented in figure 4.1. Statistical tests of differences between groups of schools, between observed scores and developers' cut points, or between years were not conducted.

Use of CLC-emphasized pedagogical practices compared with KU-CRL cut points. Across all 15 CLC-implementing schools, only the practice of "sequenced instruction" (that is, "cue-do-review") for year 1 was apparent at mean levels deemed adequate per the cut point established by developers at KU-CRL (M = 2.03, greater than cut point of 2). The mean level for sequenced instruction across the CLC schools during year 2 was M = 1.58. For "multiple instructional modalities," the mean levels in CLC schools were 1.36 for year 1 and 1.06 for year 2. Observation scores for interactive and scaffolded instruction across all CLC schools was M = 1.06 for year 1 and M = 0.68 during year 2.

Figure 4.1. Use of CLC-emphasized pedagogical practices among core content teachers in CLC schools during year 1 and year 2



Note: This figure is based on the 15 CLC schools that participated in both years of the study. Observations are weighted to account for different sampling probabilities across subject areas. Rounding may cause slight discrepancies in calculating sums and differences.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

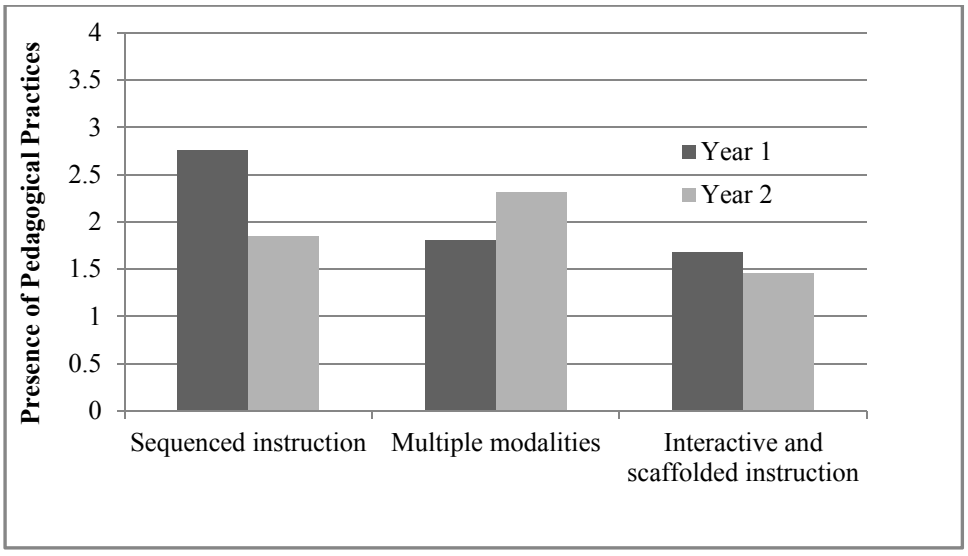
[124] Percent agreements for specific pedagogical features during year 1 ranged from 57 percent to 96 percent. For the aggregated pedagogical constructs, however, percent agreements average 42 percent. During year 2, percent agreements for specific pedagogical features ranged from 47 percent to 100 percent. Average percent agreement for aggregated pedagogical constructs was 40 percent.

Use of CLC-prescribed pedagogical practices among reading teachers

Use of CLC-emphasized pedagogical practices among reading teachers in CLC schools also was examined. The mean levels for the three practices are provided in figure 4.2. Statistical comparisons between groups of schools, between implementation years, or between observed scores and developers' cut points were not conducted. Across all reading classes observed in CLC schools during year 1, only the mean level for the practice of sequenced instruction was higher than the cut point specified by KU-CRL for adequate implementation (mean of 2.76, greater than the cut point of 2). Scores from observations of reading classes in CLC schools showed mean levels on the other two CLC-emphasized pedagogical practices as below the cut point for adequate implementation (means of 1.80 and 1.68 for multiple instructional modalities and interactive instruction, respectively).

Observations of reading classes in CLC schools during year 2 revealed that the only CLC-emphasized practice apparent across schools at levels considered adequate by the program developer was multiple instructional modalities (M = 2.31, above the cut point of 2 set as signifying adequate implementation). The amounts of sequencing of instruction and interactive and scaffolded instruction were less than adequate, based on the cut point of 2 established by KU-CRL (mean levels of 1.85 and 1.46 for sequenced instruction and interactive and scaffolded instruction, respectively).

Figure 4.2. Use of CLC-emphasized pedagogical practices among reading teachers in CLC schools during year 1 and year 2



Note: This figure is based on the 15 CLC schools that participated in both years of the study. Observations are weighted to account for different sampling probabilities across subject areas. Rounding may cause slight discrepancies in calculating sums and differences.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

Classification of instructional fidelity in CLC schools

Up to this point the analyses presented in this chapter have examined prevalence of use of CLC Content Enhancement Routines and Learning Strategies within observed core content and reading classes within CLC schools. Findings on the degree to which observed teachers are including the pedagogical practices emphasized in CLC professional development also have been summarized. Those previously presented findings used aggregated scores across all schools. This section examines whether observational data show adequate amounts of CLC-emphasized pedagogical practices school-by-school. That is, this section examines whether observed instruction in any schools exceeded the developer-established cut point for adequate implementation. This examination is done by averaging across scores for sequenced instruction, multiple instructional modalities, and interactive and scaffolded instruction (each uses a 0–4 scale) within schools and examining those averages against the program developers' cut point for adequate implementation. The findings are presented in table 4.5. Tests of statistical differences between years or between schools' average scores and cut points were not conducted.

The analysis for year 1 observational data suggests that instruction within none of the CLC schools would be classified as exemplary in terms of incorporation of CLC-emphasized pedagogical practices. Instruction in fewer than 4 of the 15 implementing schools would be considered adequate. Observed instruction in the remaining 12 schools included less-than-adequate amounts of the CLC-emphasized pedagogical practices.

For year 2, observations of core content classes suggest that instruction within CLC schools did not include adequate levels of CLC-emphasized pedagogical practices.

Table 4.5. Categorization of CLC schools based on degree to which CLC-emphasized pedagogical practices were apparent among core content teachers in the school

	Year	
Fidelity category	**Year 1**	**Year 2**
Exemplary	0	0
Adequate	≤ 4	0
Not adequate	≥ 11	15

Note: This table is based on the 28 schools that participated in both years of the study. In year 1, classroom observations were conducted in 27 of these schools; in year 2, observations were conducted in all schools. Fidelity categorizations are based on the mean of the scores for three CLC-emphasized pedagogical practices. Schools with means across the three practices that are less than 2 are classified as not adequate, schools with mean scores between 2 and 3 are classified as adequate, and schools with means that are greater than 3 are classified as exemplary.

Source: Classroom observations conducted for the CLC study in the spring of each study year.

Chapter summary

Observations of classroom instruction conducted during the spring of the 2008/09 school year (first full year of implementation) and spring of the 2009/10 school year suggest that, at best, 34 percent of core content teachers within CLC schools used graphic organizers, and 22 percent explicitly used CLC Content Enhancement Routines or Learning Strategies. Observed use of these graphic organizers, routines, and embedded Learning Strategies was even less prevalent

during observations conducted in year 2 of the project. This observed prevalence in CLC schools was about twice that observed in non-CLC schools.

For reading classes within CLC schools, observers saw an increase in the use of graphic organizers from year 1 to year 2 (31 percent and 42 percent for years 1 and 2, compared with 34 percent and 32 percent use among core content teachers). During both years, observers witnessed CLC routines and strategies being used by 62 percent of reading teachers. The difference in use of CLC Content Enhancement Routines and Learning Strategies among reading teachers compared with core content teachers is consistent with the prescribed nature of the Fusion Reading curriculum. Moreover the training that reading teachers received was more focused (that is, three days during the summer plus two additional days during the fall in off-site training sessions), compared with the monthly one-hour professional development sessions offered to core content teachers.

It is possible that core content teachers and reading teachers in CLC schools were adhering to the spirit of CLC routines and strategies without explicitly mentioning them by name. Teacher-specific scores for the CLC-emphasized pedagogical practices (that is, sequenced instruction, multiple instructional modalities, and interactive and scaffolded instruction) were created to gauge teachers' adherence to concepts underlying routines and strategies. Examination of aggregated scores across teachers in CLC schools for these three pedagogical practices showed that the mean levels of sequenced instruction during year 1 were at the level considered adequate by the program developer. All other year 1 aggregated scores and scores for year 2 were below adequate. Regarding the apparent decreases in instructional fidelity between year 1 and year 2, additional follow-up observations of teachers' instruction would be needed to determine whether these decreases represent an "implementation dip"' (Fullan, 2001) or whether they reflect decisions among teachers not to include the CLC-emphasized pedagogical practices within their instruction.

For reading teachers in CLC schools, only aggregated scores for sequenced instruction during year 1 and aggregated scores for multiple instructional modalities during year 2 were at levels considered adequate by the program developer. All other scores for the CLC-emphasized pedagogical practices were below the cut point for adequate.

Chapter 5: Impacts on Reading Comprehension and Credit Accumulation in the Second Year of the Study

The Content Literacy Continuum (CLC) is a literacy-across-the-curriculum intervention that aims to improve adolescent students' reading skills as well as their academic performance in high school. Accordingly this chapter examines the following primary impact questions:[125]

- What are the impacts of the CLC framework on grade 9 and grade 10 students' reading comprehension in the second year of the study?
- What are the impacts of the CLC framework on grade 9 and grade 10 students' accumulation of credits in core content areas (English language arts, social studies, science, and math) in the second year of the study?

The findings in this chapter show that, overall, implementing the CLC framework for two school years did not have an impact on grade 9 and grade 10 students' reading comprehension or accumulation of credits in core subject areas.[126] The following key findings are discussed:

- It cannot be concluded that the CLC framework improved students' reading comprehension scores in the second year of the study, in either grade level. Estimated effect sizes are 0.06 for grade 9 students and 0.10 for grade 10 students, but these differences are not statistically significant.
- Nor did CLC have an impact on students' accumulation of core credits in the second year, in either grade level. Estimated effect sizes are −0.17 for grade 9 students and 0.02 for grade 10 students, but these differences are not statistically significant.

Overall the findings presented in this chapter—as well as chapters 3 and 4—contribute to a growing body of research showing that effective comprehensive school reform is challenging, especially in the initial years of implementation.[127]

The remainder of this chapter discusses the key CLC impact findings in greater detail.[128] The first section begins by discussing CLC's impact on students' reading comprehension. This is followed by a presentation of impacts on students' credit accumulation.[129]

---

[125] See chapter 2 for a discussion of the rationale for choosing these two primary indicators.

[126] These conclusions are robust to alternate methods of weighting the results across random assignment blocks. (See appendix J for the results of these sensitivity analyses.)

[127] As found by Borman et al. (2003), for example, the effect of comprehensive school reforms is most evident after five years of implementation: effect sizes are 0.13 to 0.17 in the first four years of implementation, and 0.23 to 0.50 after five years.

[128] See box 2.1 in chapter 2 for information about how to read and interpret the columns in the impact tables in this report. As explained in that box, the mean outcomes presented in the impact tables are for the average CLC *school* in the study. See appendix L for the mean outcome levels for the average *student* in the CLC study, as well as standard deviations by program group.

[129] See appendix J for the standard error and confidence interval for the primary impact findings presented in this chapter, and appendix K for model fit information (R^2 and the intraclass correlation). Appendix J also presents impact estimates that are *not* adjusted for students' background characteristics and prior achievement; the main conclusions discussed in this chapter are not sensitive to these adjustments.

Impacts on reading comprehension

The CLC framework's estimated impact on reading comprehension scores in the second year of the study, by grade level, is presented in table 5.1. As discussed in chapter 2, reading comprehension in this study was measured using the reading comprehension subset of the GRADE assessment, which was administered to students in the spring of each study year. To facilitate the understanding of students' average reading comprehension level, the national grade equivalent and national percentile that correspond most closely to the average score for students in CLC and non-CLC schools are shown in table 5.1. A grade equivalent is the grade at which the mean score represents the median for the test's norming population. For example, a grade equivalent score of 9.0 refers to a median performance at the beginning of grade 9, and a 9.9 grade equivalent indicates a median performance at the end of grade 9.[130] The standard deviations used to calculate the effect sizes in the table are presented in appendix L.

As shown in table 5.1, it cannot be concluded that the CLC program improved students' reading comprehension scores in the second year of the study, in either grade level. Although reading comprehension scores were higher in CLC schools than in non-CLC schools (effect size = 0.06), this difference is not statistically significant (p-value = 0.262). Therefore it cannot be concluded that the program had an impact on grade 9 students' reading comprehension skills. This pattern of findings—that is, estimated impacts on reading comprehension that are positive but not statistically significant—is also observed for grade 10 students. The magnitude of the impact on the reading comprehension test scores of grade 10 students represents an effect size of 0.10 (p-value = 0.203).

Impacts on credit accumulation

CLC's impact on students' accumulation of credits in core subject areas (English language arts, social studies, science, and mathematics) in the second year of the study are presented in table 5.2. As discussed in chapter 2, in order to standardize the meaning of credit accumulation across school districts, the number of core credits earned was divided by the number of core credits required for graduation in a student's district. In this way the outcome measure is an indicator not only of students' content learning but also the extent to which students have progressed toward meeting the graduation requirements in their district. Impacts for each core subject area are looked at *separately* in table 5.2; however, these subject-specific results are secondary outcomes and are presented for descriptive purposes only.

[130] Differences in grade equivalents and percentiles between CLC and non-CLC schools are not shown because these measures are not equal-interval scales. Grade equivalents indicate a student's place along a growth continuum, which may not increase at regular intervals. For example, the difference between a vocabulary grade equivalent of 1.0 and 2.0 represents a greater difference in vocabulary knowledge than the difference between a grade equivalent of 8.0 and 9.0. Percentiles indicate the percentage of students in the test's norming group who performed at or below a given student's score. As such, percentiles provide information only about the rank order of students' scores; they do not provide any information about students' actual performance. Because they do not reflect equal intervals between units of measure, neither grade equivalents nor percentiles can be manipulated arithmetically (see American Guidance Service, 2001a, pp. 55–60). Thus, readers should exercise caution when interpreting differences in grade equivalents or percentiles between the CLC group and the non-CLC group.

Table 5.1. Impacts on reading comprehension, GRADE respondent sample, year 2

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
Average standard score	91.6	90.7	0.9	0.06	0.262
Corresponding grade equivalent	*6.5*	*6.2*			
Corresponding percentile	*28*	*25*			
Sample size					
Students (total = 5,011)	2,975	2,036			
Schools (total = 28)	15	13			
Grade 10 sample					
Average standard score[a]	96.7	95.1	1.6	0.10	0.203
Corresponding grade equivalent	*7.7*	*7.2*			
Corresponding percentile	*40*	*35*			
Sample size					
Students (total = 4,546)	2,908	1,638			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table 5.2. Impacts on credit accumulation, school records sample, year 2

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
Credits earned in core subject areas (%)	22.2	23.8	−1.6	−0.17	0.058
English language arts	18.9	20.9	−2.0	−0.22	0.015*
History	25.0	26.9	−1.9	−0.16	0.088
Science	25.9	26.9	−0.9	−0.06	0.505
Math	21.0	22.4	−1.5	−0.13	0.113
Sample size					
Students (total = 7,951)	4,467	3,484			
Schools (total = 28)	15	13			
Grade 10 sample					
Credits earned in core subject areas (%)[a]	41.3	40.9	0.4	0.02	0.726
English language arts	36.3	35.5	0.7	0.04	0.333
History	45.1	45.6	-0.5	-0.02	0.752
Science	47.5	46.9	0.6	0.02	0.735
Math	39.6	39.1	0.5	0.02	0.675
Sample size					
Students (total = 8,514)	4,888	3,626			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

As seen in table 5.2, it cannot be concluded that the CLC framework affected accumulation of core credits in the second year of the study for grade 9 students (effect size = −0.17; *p*-value = 0.058). Nor did the CLC framework affect grade 10 students' credit earning by a statistically significant amount (effect size = 0.02; *p*-value = 0.726).

In terms of the subject-specific findings (English language arts, social studies, science, and mathematics), the magnitude of the impact estimates is similar to the overall finding, which suggests that no one subject area drives or deviates from the overall finding. Although the estimated impact of CLC on grade 9 students' English language arts credits is negative and statistically significant (*p*-value = 0.015), this finding should be interpreted with caution because this is a secondary outcome, and statistical significance may be due to a type I error.

The following chapter presents findings from several supplemental analyses whose purpose is to better understand why CLC did not have a statistically significant impact on reading comprehension and credit accumulation in the second year of the study.

Chapter 6: Exploring Hypotheses about the Pattern of Effects on Student Outcomes

As discussed in the previous chapter, the Content Literacy Continuum (CLC) framework did not improve students' reading comprehension and credit earning in the second year of the study. This chapter includes several exploratory analyses that examine the conditions under which the CLC framework might have had greater success at improving student outcomes. The purpose of these analyses is to better understand and contextualize the primary impact findings, with the goal of guiding future research on adolescent literacy and comprehensive school reform.

Four hypotheses about the primary impact findings are examined and tested. These hypotheses are informed by CLC's design and the implementation findings presented in chapters 3 and 4 of this report. The first hypothesis is that impacts were larger in the first year of the study, when implementation of the CLC framework was relatively stronger. The second hypothesis is that although the CLC framework did not affect the primary outcomes—reading comprehension and credit earning—it may have had an impact on other related outcomes such as students' grade point average (GPA) and their reading vocabulary. The third hypothesis is that the CLC intervention was more effective for struggling readers than for more proficient readers because the former group is more likely to receive intensive CLC services. The fourth hypothesis is that the CLC framework was more effective in some districts than others because of variable implementation fidelity and service contrast across the study sites. These four hypotheses map into the following exploratory impact questions:

- What are the impacts of the CLC intervention on grade 9 students' reading comprehension scores and credit accumulation in the first year of the study? Do estimated impacts in the first year differ from estimated impacts in the second year?
- What are the impacts of CLC on students' reading vocabulary and their GPA in core content areas?
- Are the CLC framework's impacts on reading and course performance outcomes greater for some subgroups of students than for others?
- Was the impact of CLC greater in some school districts than in others?

Overall the findings in this chapter provide no conclusive support for the hypotheses that CLC was more successful at improving student outcomes under various specified conditions. The following key findings are discussed:

- **Impacts in the first year of the study.** Although implementation was relatively stronger in the first year, the CLC program did not improve students' reading comprehension or their credit accumulation in the first year of the study, and estimated impacts on the primary outcomes do not differ statistically across the two study years.
- **Impacts on related student outcomes.** The CLC program improved grade 9 students' reading vocabulary in the first year of the study. However, it did not improve their GPA by a statistically significant amount, in either grade level or in either study year.

- **Impacts by student subgroup.** There is no conclusive evidence that the programs were more effective for one subgroup of students than for another.
- **Impacts by school district.** It does not appear that CLC was more effective in some districts than in others. The estimated impacts of CLC on students' reading comprehension and credit accumulation do not vary by a statistically significant amount across school districts, in either grade level.

Taken together, these findings resonate with the conclusions drawn by Borman, Hewes, and Overman (2003) from their meta-analysis of research on comprehensive school reform efforts. Their findings suggest that impacts are difficult to achieve in the early years of a comprehensive reform effort, even for the most struggling students and the most motivated districts.[131] The remainder of this chapter discusses the findings in greater detail. It is important to remember that all results in this chapter are exploratory and should not be used to make conclusive inferences about CLC's effectiveness.

Impacts on reading comprehension and credit accumulation in the first year of implementation

Student outcomes in the second year of the study are the primary focus of the impact evaluation. This decision was made because it was expected that CLC implementation would be stronger after two years of implementation than after one year. Implementation data collected for the study, however, actually show the reverse: although still inadequate, implementation of the CLC program elements was relatively stronger in the first year of the study than in the second.

Thus the first question examined in this chapter is whether the CLC framework improved student outcomes in the first year of the study, and whether its impact in the first year was greater than in the second. Accordingly the estimated impact of CLC on students' GRADE reading comprehension scores in the first year of the study is presented in table 6.1; first-year impact findings for credit accumulation are presented in table 6.2. Results are shown for grade 9 students only because in the first year, the CLC framework was implemented in this grade level exclusively. Also note that these first-year findings are based on the 28 schools that participated in both years of the study, which makes it possible to draw inferences about the pattern of impacts across study years.[132]

Based on these findings, it does not appear that the CLC framework improved student outcomes in the first year of the study. The estimated effect size on grade 9 reading comprehension scores in year 1 was 0.13, but this estimate is not statistically significant (p-value = 0.135). The estimated effect size on credit accumulation in year 1 -0.04 but again this first-year impact is not statistically significant (p-value = 0.661).

[131] See Borman et al. (2003).

[132] Appendix M presents primary impact findings in year 1 based on all 33 schools that participated in the first year of the study. Conclusions about the CLC intervention's impacts in year 1 of the study are the same for this larger group of schools.

Table 6.1. Impacts on reading comprehension, GRADE respondent sample (grade 9), year 1

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
Average standard score[a]	92.5	90.4	2.1	0.13	0.135
Corresponding grade equivalent	*6.8*	*6.1*			
Corresponding percentile	*31*	*24*			
Sample size					
Students (total = 4,786)	2,869	1,917			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, English-as-a-second-language (ESL) status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table 6.2. Impacts on credit accumulation, school records sample (grade 9), year 1

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Credits earned in core subject areas (%)[a]	22.9	23.3	–0.4	–0.04	0.661
English language arts	19.6	20.0	–0.4	–0.04	0.632
History	25.7	26.9	–1.2	–0.10	0.318
Science	26.9	26.2	0.7	0.05	0.605
Math	21.4	22.0	–0.6	–0.05	0.529
Sample size					
Students (total = 7,365)	4,254	3,111			
Schools (total = 28)	15	13			

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Impacts on secondary outcomes

In order to add more depth to the primary impact findings, data on several secondary outcomes targeted by the CLC framework were collected. These secondary outcomes include GRADE reading vocabulary subtest scores and students' GPA in their core courses (English language arts, science, social studies, and math). This section examines whether CLC had a positive impact on these secondary measures.

Impacts on reading vocabulary

As noted previously in this report, the CLC framework aims to improve reading achievement, which includes both reading comprehension and reading vocabulary. Of these two outcomes, reading comprehension is the primary focus of the CLC, but the intervention also includes routines and strategies aimed at vocabulary. Accordingly the CLC framework's estimated impact on students' GRADE reading vocabulary scores in the second year of the study is presented in table 6.3; results for the first year of the study appear in table 6.4.

Table 6.3. Impacts on vocabulary, GRADE respondent sample, year 2

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
Reading vocabulary					
Average standard score	94.3	92.8	1.5	0.09	0.089
Corresponding grade equivalent	*7.9*	*7.7*			
Corresponding percentile	*35*	*31*			
Sample size					
Students (total = 5,011)	2,975	2,036			
Schools (total = 28)	15	13			
Grade 10 sample					
Average standard score[a]	100.5	98.9	1.6	0.10	0.142
Corresponding grade equivalent	*9.6*	*8.9*			
Corresponding percentile	*50*	*45*			
Sample size					
Students (total = 4,546)	2,908	1,638			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalents and percentiles are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

Source: Calculations from GRADE assessment administered as part of the CLC study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students; Form B to grade 10 students).

As seen in table 6.3, CLC's impacts on reading vocabulary scores in the second year of the study follow a similar pattern to its impacts on reading comprehension—that is, estimated impacts are positive in magnitude but not statistically significant. Effect sizes are 0.09 and 0.10 for grade 9 and grade 10 students, respectively, but these estimates cannot be reliably distinguished from zero. Thus it cannot be concluded that CLC improved students' reading vocabulary in the second year of the study, in either grade level.

Table 6.4. Impacts on vocabulary, GRADE respondent sample (grade 9), year 1

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Reading vocabulary					
Average standard score[a]	95.1	93.1	2.0	0.13	0.011*
Corresponding grade equivalent	*8.0*	*7.7*			
Corresponding percentile	*36*	*32*			
Sample size					
Students (total = 4,786)	2,869	1,917			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

There is some indication that CLC may have improved grade 9 students' reading vocabulary scores in the first year of the study. As shown in table 6.4, the CLC framework had a positive and statistically significant impact on grade 9 students' vocabulary scores in year 1—improving their scores by 2.0 standard score points (*p*-value = 0.011). This impact corresponds to an effect size of 0.13 and represents an improvement from the 32nd to the 36th percentile nationally. However, caution should be exercised when interpreting this finding for several reasons: first, students' vocabulary scores are a secondary outcome in this study; and second, the positive impact on vocabulary scores does not persist in the second year of the study.

Impacts on grade point average

As reported in the previous chapter, the CLC framework did not help students earn more credits in core subject areas. Nonetheless the intervention may have improved students' understanding of course content, as measured by their GPA and their course marks. Accordingly the estimated impact of CLC on students' GPA in core subject areas in the second year of the study is presented in table 6.5; results for the first year of the study appear in table 6.6. These findings are

based on the subset of students in the school records sample for whom course marks are available.[133]

Table 6.5. Impacts on grade point average, school records sample, year 2

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
GPA in core subject areas	1.81	1.93	−0.13	−0.11	0.121
English language arts	1.92	2.06	−0.14	−0.12	0.170
History	1.86	2.08	−0.22	−0.17	0.047*
Science	1.76	1.81	−0.05	−0.04	0.611
Math	1.68	1.83	−0.15	−0.12	0.274
Sample size					
Students (total = 7,917)	4,453	3,464			
Schools (total = 28)	15	13			
Grade 10 sample					
GPA in core subject areas[a]	1.87	1.85	0.02	0.02	0.820
English language arts	2.00	1.97	0.03	0.02	0.728
History	1.91	1.95	−0.04	−0.03	0.784
Science	1.80	1.83	−0.04	−0.03	0.804
Math	1.74	1.69	0.05	0.04	0.555
Sample size					
Students (total = 8,209)	4,718	3,491			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences. The analysis is based on students in the school records sample for whom course marks are available; therefore the sample size differs across subject areas. The sample sizes reported in this table are for students who have a course mark in at least one of the four core content areas.

a. GPA in core subject areas is based on a four-point scale: A+/A/A− = 4.0, B+/B/B− = 3.0, C+/C/C− = 2.0, D+/D/D− = 1.0, F = 0.0.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

[133] For some students, information is available on whether or not they earned a credit, but their mark in that course is either not provided or coded on the basis of an atypical scheme (not using letter grades A–F or numerical marks). In both cases these students' marks are treated as missing.

Table 6.6. Impacts on grade point average, school records sample (grade 9), year 1

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
GPA in core subject areas[a]	1.82	1.89	–0.07	–0.06	0.414
English language arts	1.86	1.95	–0.10	–0.08	0.555
History	1.87	2.08	–0.21	–0.17	0.080
Science	1.85	1.74	0.11	0.09	0.200
Math	1.75	1.80	–0.05	–0.04	0.639
Sample size					
Students (total = 7,315)	4,220	3,095			
Schools (total = 28)	15	13			

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences. The analysis is based on students in the school records sample for whom course marks are available; therefore the sample size differs across subject areas. The sample sizes reported in this table are for students who have a course mark in at least one of the four core content areas.

a. GPA in core subject areas is based on a four-point scale: A+/A/A− = 4.0, B+/B/B− = 3.0, C+/C/C− = 2.0, D+/D/D− = 1.0, F = 0.0.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

As shown in these tables, CLC's impact on students' GPA is similar to its impact on credit accumulation. In the second year of the study, the estimated effect of CLC on grade 9 students' GPA is negative but not statistically significant (effect size = −0.11), and for grade 10 students the estimated impact is positive but not statistically significant (effect size = 0.02). The direction and magnitude of these estimated effects closely mirror the estimated effect of CLC on credit accumulation.[134] This result makes intuitive sense because GPA is a key determinant of credit earning.[135]

[134] Impacts on GPA do not differ by a statistically significant amount across study years (for grade 9 students) or across grade levels (in year 2).

[135] Whether or not a student ultimately earns a course credit also depends on two other factors: (1) whether or not the student enrolls in the core courses needed for graduation (that is, credits attempted) and (2) the student's attendance in the course. Additional analyses, however, indicate that the CLC program does not have an impact on these two factors, and that estimated impacts on credit accumulation are most closely aligned with the pattern of impact findings for GPA. Appendix N presents the estimated impact of CLC on credits attempted and attendance.

Impacts for student subgroups

This section examines whether the CLC intervention's impact on reading comprehension and credit accumulation is greater for some subgroups of students than for others. In particular, it was hypothesized that the CLC intervention may be more beneficial for students with low levels of reading achievement. As part of the CLC framework, students whose reading skills are below grade level are more likely to receive all three levels of the CLC framework—that is, the supplemental reading class targeted at struggling readers (level 3 Fusion Reading) as well as the supports provided by content area teachers trained in CLC strategies (levels 1 and 2 of the framework). In contrast, students whose reading skills are not below grade level are not eligible for the level 3 reading class; therefore they receive a less intensive "dose" of the CLC intervention.

Based on this hypothesis, the impact of CLC on the primary outcomes was examined for student subgroups defined by several predictors of reading achievement, as follows:

- **Reading proficiency:** whether or not a student had scored below the proficiency cutoff on the grade 8 reading/English language arts assessment in his or her state
- **Overage for grade:** whether or not a student was overage for grade when he or she entered grade 9
- **Special education:** whether or not a student was classified as being eligible for special education services at the start of grade 9

It was hypothesized that impacts would be greater for the subgroups most likely to be struggling readers—that is, students who were not proficient on the grade 8 reading assessment, students who were overage for grade at the start of grade 9, and special education students.[136] Impacts on credit accumulation were examined for all student subgroups, and impacts on reading comprehension were examined for the first two subgroups only because few special education students took the GRADE assessment.[137]

Overall, these subgroup analyses do not offer conclusive evidence that the CLC intervention was more effective in improving the outcomes of any of these subgroups of students over others. In no case does the estimated impact for one subgroup differ statistically from the estimated impact for the corresponding subgroup (for example, the estimated impact of CLC on reading comprehension does not reliably differ between overage students and students who are not overage). Subgroup impacts for the primary outcomes are reported in appendix O.

[136] Students whose second language is English (English language learners or ELLs) also are more likely to be struggling readers and to receive intensive CLC supports. However, it was not possible to look at impacts for ELL students because not all schools in the study have ELLs.

[137] As explained in chapter 2, students requiring special testing accommodations did not take the GRADE. Thus, in some districts, the GRADE respondent sample does not include any students classified as ELL or special education.

Variation in impacts across school districts

This section examines whether the overall impact findings presented in chapter 5 mask differences in the impact of CLC across school districts. Such variation could occur for two reasons. First, the quality of CLC implementation could vary in the study schools. As documented earlier in this report, some schools implemented the structural and instructional components of the program with relatively greater fidelity, so one might expect impacts to be larger in these sites, all else equal. Second, some school districts may have had fewer literacy supports in place as part of "business as usual"; the effect of CLC could be larger in these districts because the service contrast between CLC and business as usual in the non-CLC schools is greater.

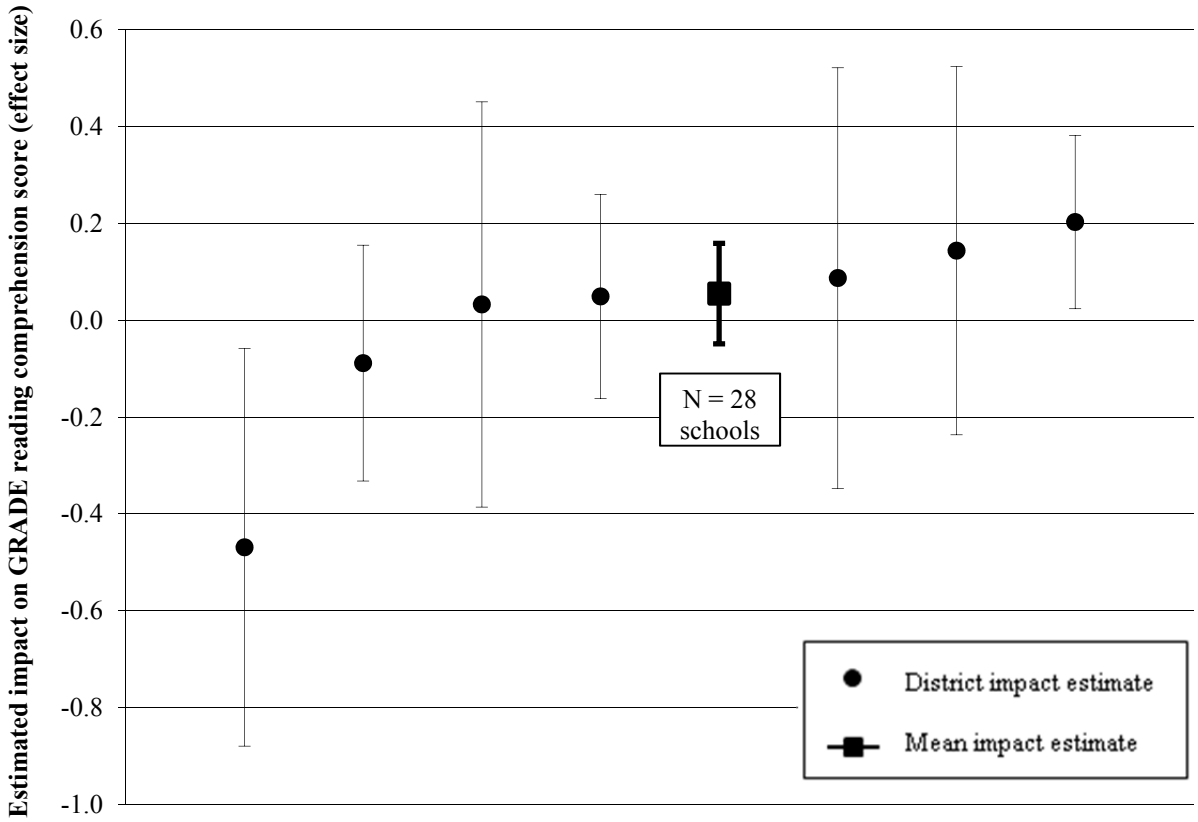
Thus the following are examined: (1) whether the intervention had a positive impact on reading comprehension and credit accumulation in some districts, even though on average the estimated impact of CLC was not statistically significant and (2) whether the CLC framework had a larger impact on the primary outcomes in some districts than in others. The findings in this section focus on the CLC framework's impact on the primary outcomes: reading comprehension and credit accumulation in the second year of implementation.

The estimated impact of CLC on GRADE reading comprehension scores for each of the seven school districts, for grade 9 and grade 10 students, respectively is presented in figures 6.1 and 6.2.[138] Similar but district-specific impact findings on credit accumulation appear in figures 6.3 and 6.4. The figures also show the 95 percent confidence interval around each impact estimate; the wider the confidence interval, the broader the margin of error and the greater the uncertainty about the impact estimate. Confidence intervals that do not include zero are statistically significant (*p*-value is less than or equal to 5 percent). To facilitate interpretation, impacts on GRADE scores are shown in effect sizes, and impacts on credit accumulation are shown in their original metric (that is, core credits earned as a percentage of core credits required for graduation).

It is important to issue two cautions about the district-specific estimates in these figures. The first is that some of these estimates are based on as little as two schools. The benefit of random assignment is that it yields two groups that, on average, are similar to each other at baseline. With very few schools, however, it becomes much less likely that the treatment and control group are similar on average. At the extreme, with only two schools, the internal validity of the results is highly questionable because the average outcome for each group is based on one school only. Thus estimated impacts based on only two schools should not be interpreted as a stand-alone finding; they are presented solely for the purpose of depicting variation in impacts across sites.

[138] As noted in chapter 2, one of the random assignment blocks is a consortium of two rural school districts. These two school districts are considered to be part of the same "district" for the purposes of this analysis.

Figure 6.1. Estimated impact on reading comprehension in year 2 of the study, by school district, GRADE respondents sample (grade 9)

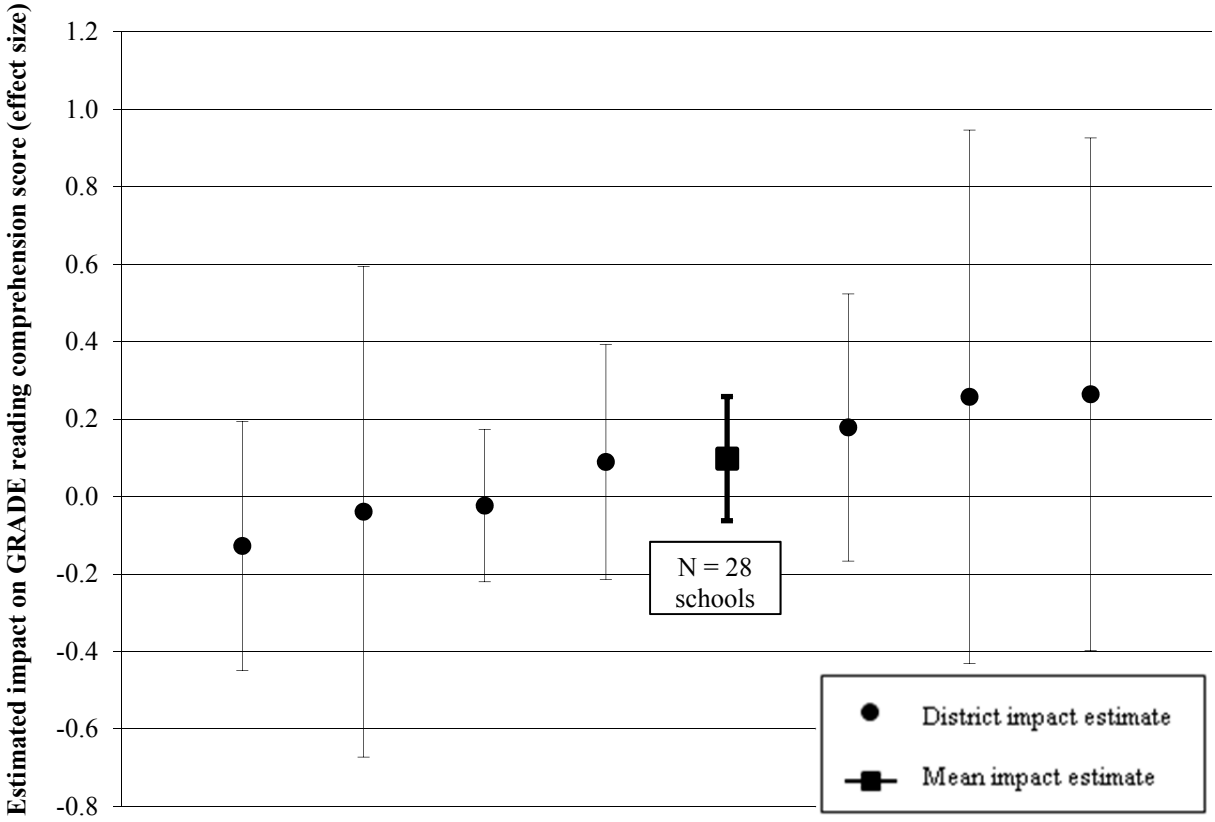


Note: This figure is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. The impact estimates in this figure are the coefficients on the interaction between districts and the treatment indicator. These estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender.

Rounding may cause slight discrepancies in calculating sums and differences. Sample size: 5,011 students.

Source: Calculations from the school records provided by participating school districts as part of the Content Literacy Continuum study.

Figure 6.2. Estimated impact on reading comprehension in year 2 of the study, by school district, GRADE respondents sample (grade 10)

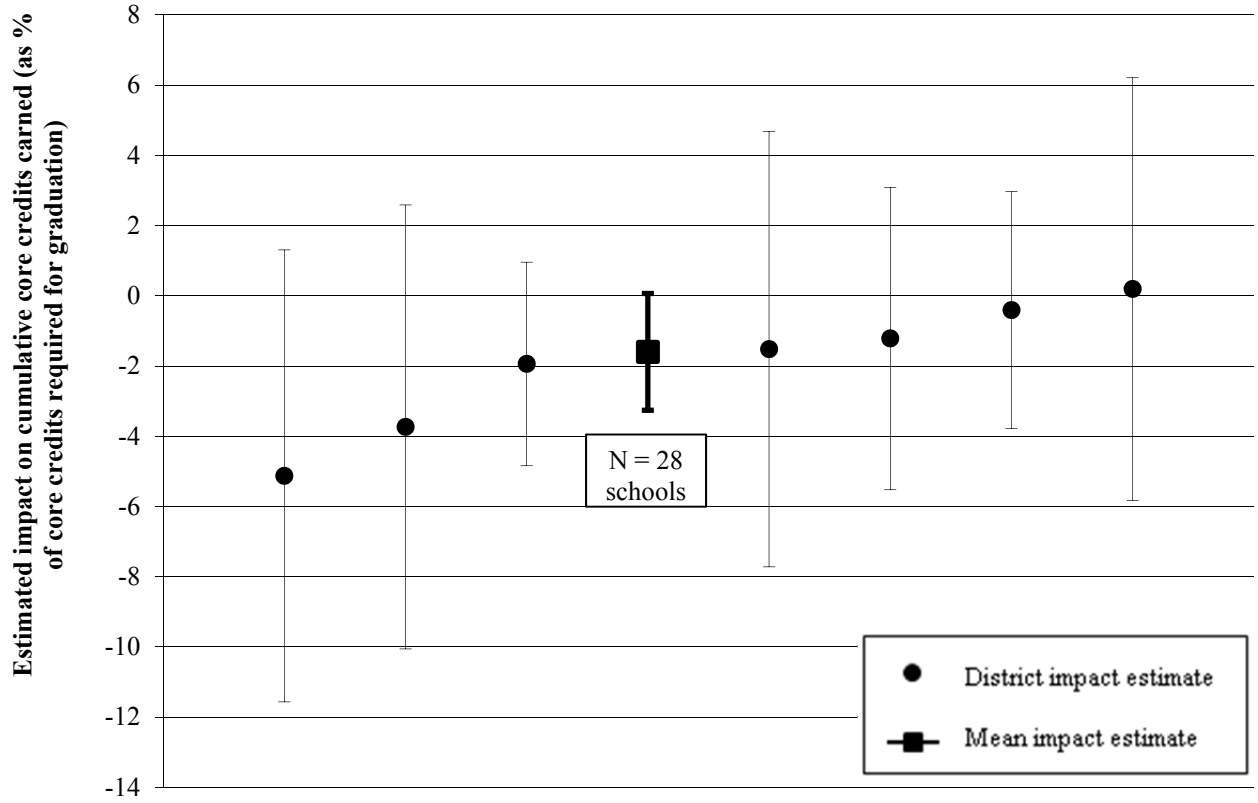


Note: This figure is based on the 28 schools that participated in both years of the study. Findings for grade 10 are based on grade 10 students and retained grade 9 students. The impact estimates in this figure are the coefficients on the interaction between districts and the treatment indicator. These estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender.

Rounding may cause slight discrepancies in calculating sums and differences. Sample size: 4,546 students.

Source: Calculations from the school records provided by participating school districts as part of the Content Literacy Continuum study.

Figure 6.3. Estimated impact on credit accumulation in year 2 of the study, by school district, school records sample (grade 9)

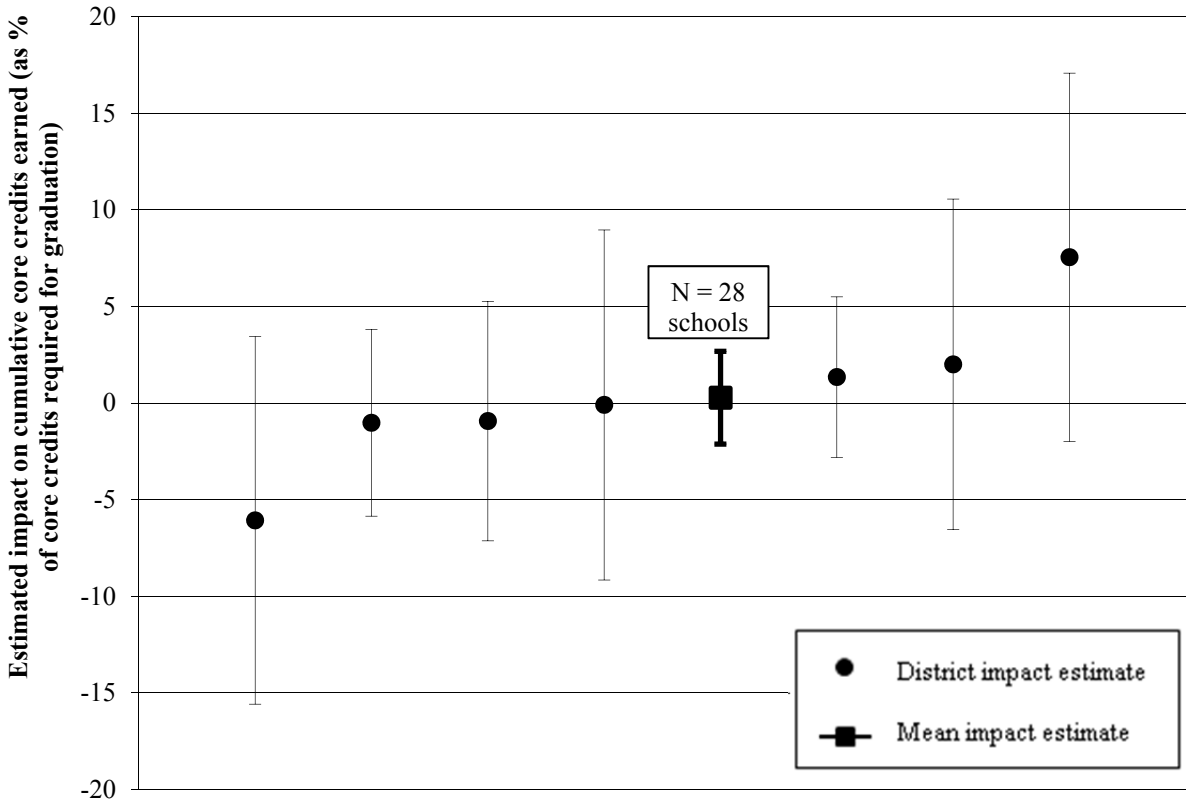


Note: This figure is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. The impact estimates in this figure are the coefficients on the interaction between districts and the treatment indicator. These estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender.

Rounding may cause slight discrepancies in calculating sums and differences. Sample size: 7,951 students.

Source: Calculations from the school records provided by participating school districts as part of the Content Literacy Continuum study.

Figure 6.4. Estimated impact on credit accumulation in year 2 of the study, by school district, school records sample (grade 10)



Note: This figure is based on the 28 schools that participated in both years of the study. Findings for grade 10 are based on grade 10 students and retained grade 9 students. The impact estimates in this figure are the coefficients on the interaction between districts and the treatment indicator. These estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender.

Rounding may cause slight discrepancies in calculating sums and differences. Sample size: 8,514 students.

Source: Calculations from the school records provided by participating school districts as part of the Content Literacy Continuum study.

The second caution is that the impact variation displayed in these figures overstates the *true* variation in impacts. The district-specific impact estimates are measured with error; this "noise" inflates the amount of variation in the observed impacts and therefore overstates the amount of true variation in impacts across districts. To examine variability in impacts across schools more systematically, an omnibus test was used to assess whether the district-level impacts on reading comprehension test scores are statistically equivalent. This test accounts for estimation error in district-level impacts and provides an indication of the confidence one might have that there is variation in true impacts across school districts.[139]

The key findings from the analysis of site variation are as follows:

- The estimated impact of CLC on reading comprehension does not vary by a statistically significant amount across school districts, in either grade level.[140] In one school district the estimated impact of CLC on grade 9 students' reading comprehension scores is positive and statistically significant (effect size = 0.20, *p*-value = 0.03). In another school district, there is a statistically significant negative impact (effect size = -.47, *p*-value = 0.03). However, the statistical significance of both of these result should be interpreted with caution: overall variation in impacts is not statistically significant, and so this one statistically significant impact estimate could be due to a type I error.
- The estimated impact of CLC on credit accumulation does not reliably differ across school districts, in either grade level.[141] None of the individual impact estimates is statistically significant.

Overall these findings suggest that CLC's impacts were similar across school districts—the intervention was not more successful in some districts than in others.

[139] This test is conducted by (1) estimating a model that includes random assignment blocks, student-level covariates, a treatment indicator, and a set of interactions between the treatment indicator and districts and (2) testing whether this latter set of interactions is jointly statistically significant.

[140] The *p*-values are 0.17 for grade 9 impacts and 0.92 for grade 10 impacts.

[141] The *p*-values are 0.70 for grade 9 impacts and 0.33 for grade 10 impacts.

Chapter 7: Conclusions and Study Limitations

As stated at the outset of this report, there is a need to develop and test interventions focused on improving adolescent literacy within secondary schools. The developers of the Content Literacy Continuum (CLC) at the University of Kansas Center for Research on Learning (KU-CRL) designed their schoolwide instructional framework in response to that need. Their intervention aims to improve literacy skills and academic achievement of adolescent students by having core content teachers enact intervention-specific instructional routines, by having core content teachers embed content-specific learning strategies within their instruction, and by providing students who are two to five years below grade level in reading with supplemental reading instruction.

The high schools with low-income and low-achieving students assigned to implement the CLC framework attempted to do so over two full years. Random assignment of half of the eligible schools to implement CLC and the other half to serve as comparison schools (that is, not implementing CLC or similar programs that include both content area and supplemental literacy support) enabled evaluators at REL Midwest and MDRC to examine differences in performance among students in the two types of schools and to attempt to attribute those performance differences to CLC.

This chapter provides concluding remarks about the fidelity of implementation of the intervention in CLC schools and whether this study can state definitively that CLC does or does not lead to improvements in students' academic achievement. The chapter also makes explicit some of the methodological limitations of the study.

Conclusions

This section highlights some of the conclusions that can be drawn regarding the CLC intervention, the fidelity of implementation of CLC in schools for this study, and impacts of the intervention on students' academic achievement within the study schools.

The intervention

CLC is a tiered, schoolwide intervention that aims to support core content teachers within secondary schools (that is, teachers in the content areas of English language arts, mathematics, science, and social studies) in the use of developer-specified instructional routines and to share with students content-specific learning strategies (that is, strategies that allow them to access, understand, and retain course content). The intervention also seeks to establish supplemental reading classes for students who are two to five years below grade level, in which enrolled students are provided more intense exposure to learning strategies.

The theory of action underlying the intervention (see chapter 1) posits that representatives of the program developer's implementation team (that is, Action Designs for this project) initially engage with school and district leaders who provide support and direction for implementation. According to the theory, site coordinators need to engage with school leaders in order to accomplish the following:

- Understand better the strengths and weaknesses among teachers and students
- Co-construct the order in which instructional routines and learning strategies will be presented to core content teachers in professional development
- Establish dates during which the professional development for core content teachers will take place
- Set up and staff the supplemental reading classes (Fusion Reading)
- Establish a mechanism for enrolling students into sections of Fusion Reading

Implementation of CLC involves putting in place several structural components (see chapter 3). According to the theory of action, these structural components—including the provision of CLC professional development to core content teachers on Content Enhancement Routines and Learning Strategies and to Fusion teachers around more intensive learning strategy instruction—*should* lead teachers to enact those routines and share the learning strategies with students. The theory specifies that the culmination of these schoolwide activities should produce better instruction, additional support for struggling readers, and thus improved reading achievement and general academic achievement for students.

Fidelity of structural implementation

The following sections summarize findings on fidelity of implementation according to the sequence specified in the theory of action. The number of meetings between site coordinators and school leaders is presented first, followed by the amount of time set aside for professional development, teacher participation in professional development, setting up supplemental reading classes, and enrolling students into those classes.

Leadership support. As specified in KU-CRL's description of CLC, "A hallmark of the entire adoption process is that it [CLC] is co-constructed with school leaders" (Lenz et al., 2005, p. 4). Thus the program developers intend for school leaders and the Literacy Leadership Team to play an important role in coordinating the implementation of CLC. Several indicators from site coordinators' implementation logs reflect the amount of support obtained from school leaders during this project, as follows:

- On average, school leaders had 5.3 meetings with site coordinators during year 1 and 3.3 meetings during year 2. The program developer considers 8 meetings per year to be necessary for successful implementation.
- Across the 15 high schools in the CLC condition, 6 schools had established a Literacy Leadership Team during year 1 (or had a schoolwide leadership team perform the functions of this group), and 9 had established a Literacy Leadership Team during year 2. The program developer considers it essential that *all* schools implementing CLC establish a Literacy Leadership Team (or relegate the functions of this team to the school's general leadership team).

In addition, at 7 of the 15 schools, leaders who originally agreed to support involvement in the study and support implementation of the CLC framework had been replaced by the end of year 2 of implementation. The subsequent leaders inherited a partially implemented program that they may or may not have bought into, despite additional work by the site coordinators and the study

team to introduce the intervention and the study and to secure their support for continuation of the project.

Professional development. School leaders' accommodation of time for site coordinators to provide CLC professional development and teachers' participation in that professional development are necessary components of CLC implementation. According to site coordinators' monthly reports, on average, many of the schools did not meet the implementation thresholds for professional development defined by the program developer. The reports indicate the following:

- Site coordinators were able to schedule 1.9 days per month with CLC schools during year 1 and 1.6 days per month with CLC schools during year 2 (KU-CRL considers 2 days per month to be necessary for adequate implementation).
- Site coordinators were on-site performing professional development for an average of 18.7 days during year 1 and 15.7 days during year 2 (the program developer considers18 days as "adequate").
- During year 1, site coordinators conducted professional development during 7.4 months (8 months is considered adequate), with professional development visits conducted in 6.7 months during year 2.
- For year 1, 89.8 percent of core content teachers participated in CLC professional development (with 80 percent considered adequate for implementation). A smaller percentage of core content teachers (75.6 percent) participated in professional development during year 2.

Setting up supplemental reading classes. During the recruitment process, districts and schools assigned to the CLC condition agreed to schedule a sufficient number of sections of Fusion Reading to accommodate all students within the school who were two to five years below grade level in reading, according to grade 8 reading tests or other assessment data. They also agreed to staff these sections with teachers who would participate in professional development specific to Fusion Reading curriculum and instruction. Site coordinators' monthly reports indicated the following:

- Six of 15 schools during year 1 and fewer than 4 of 15 schools in year 2 set up a sufficient number of sections of Fusion Reading to accommodate the students needing such a class.

On average, Fusion Reading teachers participated in 3.8 days of training on Fusion curriculum during year 1 (5 days considered adequate by KU-CRL) and 1.7 days of training during year 2 (3 days considered adequate by KU-CRL during year 2).

Summary of structural implementation. In sum, during year 1 of implementation, the implementation thresholds for 4 of 10 of the indicators of structural fidelity (from table 3.1) were met. One of 9 of the structural components was adequately established during year 2 of implementation. These data suggest that collectively schools assigned to implement CLC failed to establish the structures at a level considered "adequate" by the program developers as a platform for enacting CLC instructional practices in classrooms.

Fidelity of instructional implementation

The CLC theory of action suggests that successful implementation of the structural components *should* lead to CLC-aligned instruction among core content teachers. Such instruction should include the presence of more graphic organizers and more CLC instructional routines and learning strategies. High fidelity instruction also should include the pedagogical practices emphasized by site coordinators during professional development for levels 1, 2, and 3 of CLC (that is, sequenced instruction, multiple instructional modalities, and interactive and scaffolded instruction).

CLC implementation did create a contrast between CLC and non-CLC schools regarding the use of CLC instructional routines and strategies. These routines and strategies were observed in use twice as often in classrooms in CLC schools as in classrooms in non-CLC schools in both years of the study. Despite this contrast, CLC routines and strategies were not observed to be widely used within CLC schools (22 percent in year 1 and 11 percent of classrooms during year 2). Observers witnessed teachers sharing *other* learning strategies with students in classrooms in non-CLC schools as well.

Observation data suggest that across all 15 CLC schools during both years of implementation, only the average level of sequenced instruction for year 1 reached a level considered "adequate" by the program developer. These findings suggest that core content teachers did not incorporate content enhancement routines or learning strategies within their instruction.

Moreover, comparison between CLC and non-CLC schools on the presence of CLC-emphasized pedagogical practices revealed no statistically significant differences between the two types of schools during year 1 and one statistically significant difference (interactive instruction was *greater* in comparison schools than in CLC schools) during year 2.

Impacts

The study described in this report was designed to provide a rigorous test of whether CLC produces improved reading comprehension and greater accumulation of course credits among students. The cluster-randomized research design allows evaluators to attribute observed differences in outcomes to the intervention, provided the intervention was implemented with fidelity.

After two years of implementation of levels 1–3 of CLC, findings indicate no statistically significant differences in reading comprehension or accumulation of course credits among grade 9 or grade 10 students between CLC schools and non-CLC schools. As noted in the previous section, these findings may not reflect the lack of effectiveness of CLC as designed, but rather the lack of fidelity to that design in implementation of CLC by the schools in this study.

Study limitations

This study was designed to determine whether CLC produced impacts on students' reading achievement and general academic achievement after two years of implementation. Several study-related issues must be kept in mind when examining the findings contained within this report.

Duration of implementation

Although the developers of CLC have not published clear guidelines on an amount of time needed for the intervention to produce impacts, they do state that "initial adoption takes place over a three- to five-year period as school staff work through activities associated with the phases of planning, implementing, and sustaining a literacy improvement initiative" (Lenz et al., 2005, p. 4). This statement is consistent with previously mentioned research that suggests that comprehensive school reform models (that is, interventions such as CLC that emphasize coordination and change of multiple school processes) typically are unable to attain significant impacts until at least three or more years of high-fidelity implementation (Aladjem et al., 2006; Borman et al., 2003; Desimone, 2002).

The developers of CLC have published statements indicating that implementation of the full CLC framework requires at least three years. However, it was expected that impacts would emerge by the second year of implementation, provided that the following conditions were met: (1) comparison schools refrained from implementing a schoolwide, tiered literacy approach; (2) CLC schools implemented the structural components with fidelity; and (3) core content teachers in CLC schools included Content Enhancement Routines and Learning Strategies within their instruction. Overall, less than adequate implementation of the structural components of the intervention in CLC schools and of the CLC-emphasized pedagogical practices in these schools' classrooms were observed. Since CLC was not fully implemented as suggested by the developers, the results of this study cannot or should not be interpreted to suggest that CLC could not be effective if fully implemented as the developers intended.

Statistical power

As discussed in chapter 2, the short time horizon of this evaluation and the fact that five schools left the study after the first year[142] created challenges to the ability of this evaluation to detect impacts on the primary student outcomes of magnitudes that might be anticipated based on previous research on comprehensive school reform. Borman et al. (2003) find in their meta-analysis of comprehensive school reform programs that these programs improve student outcomes by an effect size of 0.23 to 0.50 after five years of implementation. However, because of the complexity of whole-school reforms, impacts in the first few years of implementation are smaller in magnitude—ranging from 0.13 to 0.17 in effect size.[143] Three of the Striving Readers studies presented in table 1.2 are evaluations of related literacy interventions from the same developer, KU-CRL. These evaluations found two-year impacts on reading assessment outcomes ranging from 0.10 to 0.29 standard deviation, and two impacts, both for middle school students,

[142] As noted in chapter 2, the original sample of schools was 37, almost at the total of 40 schools that was the high end of the site recruitment target set at the start of this study.

[143] See Borman et al. (2003).

were statistically significant (effect sizes = 0.13 and 0.29 standard deviation). As reported in chapters 5 and 6 of this report, impacts on reading comprehension ranged from 0.06 to 0.10 standard deviation, magnitudes that overlap with the range represented by the Striving Readers studies. However, these impact estimates were not statistically significant. Greater statistical power resulting from a larger sample size could have allowed for a stronger test of the effectiveness of CLC under the conditions in which it was implemented for the study.

Contrast between CLC schools and non-CLC schools

Readers are reminded that the counterfactual condition for this evaluation is not the absence of literacy programming within control schools but rather the absence of a tiered, schoolwide approach to literacy (that is, approaches similar to CLC) within those schools. Interview data suggest that most non-CLC schools upheld the primary provision specified for control schools: to refrain from implementing the tiered, schoolwide approach to adolescent literacy (that is, literacy-across-the-curriculum combined with supplemental reading classes). However, most control schools (6 of 13 schools during year 1 and 10 of 12 schools in year 2) were emphasizing literacy as an instructional goal for teachers within their buildings (per interview data, see tables 3.8 and 3.9). Thus the findings should not be interpreted as representative of the impact of CLC compared with no literacy programming, but rather as compared other forms of literacy programming.

Generalizability of GRADE findings

The GRADE, the assessment used to measure students' reading comprehension and vocabulary, was administered in participating schools between March and May of each year of implementation. However, not all grade 9 students within the study schools took the GRADE assessment during year 1, nor did all the grade 9 and grade 10 students take the GRADE assessment during year 2. Rather, some schools were allowed the option of administering the GRADE to students within a random sample of classrooms, and the assessment was administered to just those students who were in attendance on the testing days.

As mentioned in chapter 2, an analysis was performed to compare students who completed the GRADE versus students who did not complete on academic performance measures obtained through school records. Findings from this analysis suggest that those who completed the GRADE were higher performing on average than those students who did not complete it. The results of this analysis suggest that GRADE-related findings may not be generalizable to all students within the types of schools comprising the study sample.

Validity of observation-based findings during year 2

Comparison of classrooms randomly chosen for observation and classrooms actually observed indicated that during year 2, more of the classrooms that were chosen randomly for observation went unobserved in the non-CLC schools than in the CLC schools. This finding suggests that the statistical contrasts of instruction between teachers in CLC schools and non-CLC schools may not be reflecting true differences in instruction between the two types of schools. These findings are best interpreted with caution.

Reliance of site coordinators' monthly reports for implementation data

The examination of fidelity of implementation of structural components of CLC (components listed in table 3.1) was based primarily on data supplied by site coordinators who were employed by the program developer. These data are in the form of components of a uniform system for monitoring implementation created by the program developer. This system included components such as site coordinators' assessments of teachers' implementation of Content Enhancement Routines and Learning Strategies, logs of professional development activities (coaching, training on Content Enhancement Routines and Learning Strategies, meetings with Fusion Reading teachers), teacher attendance at professional development activities, logs of contacts with school and district administrators, and comments regarding Literacy Leadership Team activities.

These data were the sole source relied on for many of the structural implementation measures, and they usually could not be validated based on other data collected by the study team. However, the data in site coordinators' reports are consistent with challenges expressed by site coordinators and program developers to the evaluation team throughout the course of the implementation. That said, the implementation story in actuality may be more nuanced or might vary in the eyes of school and district staff or strictly independent observers.

Lack of reliability of observational measures

Observations were conducted by former educators, former education researchers, or full-time evaluation staff following a12-hour training on the project-developed observation protocol. This observational protocol required observers to record whether CLC-emphasized instructional practices were being used and whether other instructional practices that are cited as influencing content literacy were being used (see appendix B for more information).

Interrater reliability estimates were calculated on the basis of (1) co-observation of individual teachers by pairs of observers during the spring 2009 pilot observations (see table I.1) and (2) observation ratings following viewing of a film clip by all observers during fall 2009 (see table I.2). Interrater reliability estimates for observation data are provided in appendix I. Estimates for individual behaviors (for example, use of "named" strategies) for year 1 ranged from 0.57 to 0.96. Estimates of reliability for individual behaviors in year 2 ranged from 0.47 to 1.00, with 9 of 12 behaviors in year 2 greater than 0.50. When observers' data on these individual behaviors were summed to form composite indicators of CLC-emphasized pedagogical practice, the interrater reliabilities ranged from .36 to .47 across the two years. Thus readers should consider observations of individual behaviors as more trustworthy than the aggregates.

It should be noted that following the pilot observations during spring 2009 and the joint ratings of video clips of classroom instruction in fall 2009, observers were provided with additional clarification on how to make ratings of various actions. However, interrater reliability information was not gathered again prior to actual observations in spring 2009 (following the pilot observations) or before the observations performed in spring 2010 (after the joint ratings of video clips were taken).

Because of the lack of reliability information on actual observations, combined with the limited number of "snapshots" of instruction that these observations represent, readers should interpret with caution the findings generated from observations (for example, teachers' use of the CLC pedagogical practices).

Final comments

An alternative explanation for the lack of fidelity of implementation of structural components is that the CLC intervention may be too complex to implement in only two years. The comprehensive instructional framework that CLC represents has many components that require hard work on the part of both school staff and the program developer's implementation support staff, and a two-year timeline may not be adequate to complete the necessary implementation tasks. The program developer can cite examples of successful implementation of the intervention in secondary schools. However, implementation in those sites took place over the course of six years (Deshler, 2010).

In addition it is possible that the potential for CLC to have impacts on student literacy and course outcomes may be highest for the level 3 students. Unlike the students who do not receive level 3 support, these students have an opportunity to receive CLC instruction through Fusion Reading (for which the curriculum and instruction are more closely controlled by the program developer) and through their content area classes. This report does not examine the outcomes of this important subgroup of students explicitly, but it would be valuable for future studies to try to understand better whether there is an association between impacts and variation in the degree of CLC support students receive.[144]

More generally, more analysis should be done that tries to understand better whether there are contexts or conditions in which CLC is more or less successful in achieving its goals of improving student outcomes. The findings from such analyses might benefit the program developer in terms of understanding whether and how to continue to refine CLC. The findings also might help schools and districts understand whether their particular contexts might make for a better or worse match for CLC.

[144] Such an analysis would require "matching" students in level 3 reading classes in CLC schools with students with similar reading performance at baseline from non-CLC schools. The project team was unable to perform that supplemental analysis within the time frame established for this project.

Appendix A. Sampling of Classrooms for GRADE Testing and Classroom Observations

This appendix supplements chapter 2 by providing a more detailed description of how classrooms were sampled for GRADE testing and observations of classroom instruction.

Classroom sampling for GRADE testing

As explained in chapter 2, the GRADE assessment was administered to grade 9 and grade 10 students enrolled in the study schools in the spring of each study year. To ease the testing burden, schools were given the option of either testing all students or a random sample of their students. The random sample option was chosen by one school district (two random assignment blocks, four schools) in both year 1 and year 2 and by another school district (one random assignment block, two schools) in year 2.

In these schools, classrooms of grade 9 and grade 10 students were randomly sampled from a master schedule of the classrooms listed during the class period in which testing was scheduled. The number of sampled classrooms was chosen based on the goal of selecting 125–150 students for testing, to ensure that even with absences, at least 100 students per grade would write the exam on the day of testing.

Classroom (cluster) sampling was used in both year 1 and year 2. In year 2, classroom sampling was stratified by both grade level (grades 9 and 10) and class size (large and small classrooms).[145] Sampling was stratified by grade to ensure that the target sample size of 100 tested students would be achieved in each grade.[146] Stratification by class size was further imposed because there were more small classes in grade 10.[147] Stratifying by class size made it possible to oversample large classrooms, thereby ensuring that the target sample size would be met. In addition within the strata of small classrooms, classes were selected with a probability proportional to size to increase the probability of selecting the small classrooms with the greatest enrollment.[148] Sampling weights are used in the analyses to account for this stratified sampling scheme.

[145] In year 1, only grade 9 students were tested. Therefore it was not necessary to stratify; grade 9 classrooms were selected using simple random sampling.

[146] "Mixed" classes with students in both grades were categorized as either grade 9 or grade 10 based on the official grade level of the majority of students in the class. Schools provided enrollment data for each class by grade, which made it possible to categorize classes as either grade 9 or grade 10.

[147] This is primarily due to the fact that more electives are offered in grade 10.

[148] A class was categorized as small if it had fewer than 10 students in the grade of the class (that is, fewer than 10 grade 9 students in grade 9 classes and fewer than 10 grade 10 students in grade 10 classes). Schools provided enrollment data for each class by grade, which made it possible to categorize classes as large or small.

Classroom sampling for classroom observations

Observations of instructional practice were conducted in both CLC and non-CLC schools to measure the fidelity with which teachers in CLC schools use the framework's routines and strategies and assess the instructional contrast between CLC and non-CLC schools.

Classroom sampling for observations of instructional practice

To obtain a representative portrait of instructional practice in the study schools, grade 9 and grade 10 classrooms in each school were randomly selected for observation from a sampling frame that included mainstream classes and special education inclusion classes in the four core subject areas (English language arts [ELA], mathematics, science, and social studies) as well as Fusion Reading classes (CLC schools only). The sampling frame was constructed using master schedules obtained from each school in the study and informed by discussions with school administrators as well as documents describing the school's standard grade 9 and grade 10 curricula for the four core subjects and reading. From this sampling frame, core content classrooms were then randomly sampled for observation. In CLC schools, Fusion Reading classes were randomly sampled for observation.[149]

To ensure that classrooms from each of the four core content areas would be observed, the random sampling of core content classrooms was stratified by subject area. The number of classrooms sampled per subject area in each school depended on the number of periods in the school's schedule, but in general a higher number of ELA and social studies classrooms were selected for observation than the other subject areas, and at least two classrooms per subject were sampled at each school. When conducting the analyses in this report, observations are weighted to account for different sampling probabilities across subject areas.

In year 2, sampling was *not* stratified by grade level. As explained in chapter 2, two grade levels had to be observed in year 2, but the total number of classes observed could not be doubled to account for having to observe an additional grade. Thus the sample size for each grade level is too small to conduct separate analyses by grade. For this reason stratifying classroom sampling by grade level was deemed unnecessary.

[149] In fewer than three districts, teachers were given the opportunity to opt out of classroom observations. Thus the sampling frame in these cases consisted of only the consenting teachers' sections in the four core subjects and reading. In these cases the consent rate does not differ by a statistically significant amount among teachers in CLC and non-CLC schools.

Appendix B. Measuring Instructional Practice with the ACE Observation Protocol

The ACE observational protocol was developed to serve two functions for this study:

- To help trained observers record the extent to which specific practices stressed during levels 1, 2, and 3 of Content Literacy Continuum (CLC) professional development are present within instruction
- To help observers record the extent to which quality literacy instruction is demonstrated in the class

This appendix provides information on the development of the protocol, observer qualifications and training, and the use of the protocol during visits to participating schools.

Protocol development

A group composed of members of the CLC evaluation team and two literacy experts developed the protocol. The team was divided in half. Some members were tasked with identifying distinctive features common in the training materials for the Content Enhancement Routines and embedded Learning Strategies; others were tasked with identifying important components of high-quality instruction on content literacy, as found in the research literature. The shared pedagogical practices emphasized in levels 1, 2, and 3 of CLC professional development have been identified in the main text of this report: sequenced instruction, multiple instructional modalities, and coherent instruction. These CLC-emphasized pedagogical practices are used to examine the *fidelity of instruction.* The components of quality literacy-laced instruction were included as measures of *instructional impacts*.

Team members charged with identifying important components of quality content literacy instruction identified 18 such components. A brief explanation and relevant research citations for all 18 components follow:

Setting purpose. Research literature suggests that comprehension is a purposeful act. Readers construct knowledge by setting a purpose, or sometimes multiple purposes, for reading. Purposes for reading can be affected by motivation, cognitive conditions, and learning environments (Cramer, 1970; National Institute of Child Health and Human Development, 2000; Ruddell & Unrau, 2004).

Background knowledge. Research suggests that prior knowledge is a crucial component of effective comprehension. Building and activating students' background knowledge prior to reading enables them to relate and connect what they are reading to what is already known. These connections help consolidate learning into long-term memory (Anderson & Pearson, 1984; Dole, Brown, & Trathen, 1996).

Identifying text structure. Research suggests that efficient readers use text structure as an aid to comprehension. Text structures differ within and across genres. Comprehension improves when

readers are more aware of various text structures (Goldman & Rakestraw, 2000; Pearson & Camperell, 1994).

Predicting content. Some research literature indicates that proficient readers make predictions as they read. They test and revise predictions as they continue reading. Predictions can be facilitated by previewing texts to provide an overview of the text plus expectations for information that can be found in the text (Brown, Palinscar, & Armbruster, 2004; Pearson & Fielding, 1991).

Highlighting essential vocabulary. Research indicates that readers need more than a superficial knowledge of words to fully comprehend text. Effective instruction focuses on selecting essential vocabulary to develop text knowledge and providing conceptually based instruction. Such instruction helps students develop a deeper understanding of the words and facilitates comprehension (Beck, McKeown, & Omanson, 1987; Harmon, 2002; Harmon, Hedrick, & Fox, 2000).

Developing content knowledge. Research suggests that to develop content knowledge, expectations for learning must be clearly stated. Content knowledge becomes what students are expected to know and understand as a result of their learning. This involves learning at both the factual level and the conceptual level (Vacca, 2002; Weaver & Kintsch, 1991).

Determining importance. Research also suggests that proficient readers make decisions about what is important in text at the word, sentence, and text levels. The reader's purpose, as well as expectations and opinions related to the text, have an impact on determining what is important. The key ideas, concepts, and themes found in the text play a role in helping the reader determine importance (Alvermann, Swafford, & Montero, 2004; Harvey & Goudvis, 2000).

Inferring. Some research shows that proficient readers form inferences by using a combination of background knowledge and explicitly stated information from the text. Inferences are reasonable predictions. Readers test, adjust, and revise predictions as they continue reading (Dewitz, Carr, & Patberg, 1987; Pressley, 2002).

Questioning/Clarifying. Proficient readers generate questions before, during, and after reading. They use questions to clarify meaning, locate a specific answer, or consider questions stimulated by the text. Readers use questions to focus on important information; they also understand that they can pose questions to critically evaluate the text (National Institute of Child Health and Human Development, 2000; Raphael & Pearson, 1985).

Synthesizing/Summarizing. Summarizing improves the overall comprehension of text. When readers summarize, they go through the text and distinguish important from unimportant ideas. Synthesizing extends the literal summary of the text to an inferential level. This synthesis includes ideas and themes relevant to the overall meaning from the text (Bean & Steenwyk, 1984; Duke & Pearson, 2002).

Strengthening content knowledge. Learners strengthen content knowledge through consolidation. This involves organizing information to facilitate placement and retrieval within long-term memory (Kintsch, 2004; Vacca, 2002).

Monitoring learning. Monitoring is a self-regulating behavior necessary for optimal learning. Proficient readers monitor their learning at the word and text levels. Effective monitoring enables readers to detect and resolve misunderstandings during and after reading (Graves & Watts-Taffe, 2002; Hacker, 2004).

Reflecting on learning. Readers who reflect on learning not only seek answers to posed questions but also generate new questions based on evidence from the text. Reflection is impacted by characteristics of the learner, the learning environment, and the task at hand (Clark, 2007).

Coherence. Instructional coherence is facilitated by clearly delineated goals for instruction and cohesive lesson designs. This requires active student involvement to maximize learning. Instruction is focused on student learning by intentionally bringing together the sequence of instruction, the use of materials, and student engagement (Bransford, Brown, & Cocking, 2000; Nelson, 2001).

Scaffolded instruction. Scaffolding is a temporary system of support that enables a student to move from dependence on the teacher to independent learning. Stages of scaffolding generally move from teacher modeling to shared instruction to guided learning and finally to independent application. Progression through these stages is continually informed and guided by the needs of each student (Duke & Pearson, 2002; Pressley, 2002).

Discourse. High-quality instructional discourse has a significant impact on student learning. Meaning is situated not only in the text but also in the experiences learners bring to the task. Scaffolding instructional discourse can enable students to connect their experiences to current and future learning (Gee, 2000; Nystrand & Gamoran, 1988).

Engagement. Engagement occurs when learners view a task as doable. Engagement impacts motivation and self-efficacy. Engaged learners exert effort and persistence and have an expectation of success (Guthrie et al., 1996; Wigfield, 1997).

Assessment. Valuable information is derived from effective assessment. This also guides further instruction and learning. Within effective assessment processes, students are aware of the skills and concepts necessary for success. Formal and informal assessment processes, as well as student self-assessment, provide a comprehensive analysis of learning (Black & Wiliam, 1998; Guskey, 2003).

Using the ACE protocol

Twelve of the 18 components were judged by literacy experts to be central to literacy-related instruction. Elements of the remaining 6 components were retained, but because elements of those components overlapped pedagogical practices emphasized in CLC professional development, those elements were reconstituted as part of the indicators for CLC-emphasized pedagogical practices (that is, measures of the fidelity of instruction). For instance scaffolding and engagement are reflected by interactive and scaffolded instruction.

The protocol was originally designed around three dimensions of instruction: *Activating knowledge, Constructing knowledge, and Extending knowledge.* The body of the protocol has three nested levels within each dimension: the *constructs* underlying the dimension, the instruction-related *actions* that demonstrate the construct, and the modality in which the action is applied (*application*).

The dimensions of literacy-related instruction are not included in this report. However, observers' recordings of the presence of instructional actions that reflect CLC-emphasized pedagogical practices were used to assess the fidelity of instruction. The creation of scores

representing the fidelity of instruction to CLC-emphasized pedagogical practices is discussed in appendix H.

Hiring and training of observers

Observer qualifications. During each year of data collection, the project evaluation team recruited observer candidates from geographical areas surrounding the school districts participating in the CLC project. The project evaluation team targeted former educators or former education researchers for this role, with preference given to candidates with experience at the secondary level and in observing instruction. Several members of the project evaluation team also served as classroom observers. These observers also had experience in conducting unobtrusive classroom observations. Sixteen people conducted observations of instruction across the two data collection years.

All observers were required to pass a background check conducted by REL Midwest's parent organization. Several districts also required observers to pass the background check conducted by local vendors and their state police department.

All observers took part in at least one 2-day training session on using the ACE protocol. Follow-up refresher sessions also were conducted via phone and online conferencing.

Training sessions for observers. Observer training sessions were conducted in January 2009 and October 2009. The training sessions lasted two days and were held in a central location. The two literacy experts who created the ACE protocol also led the observer training. The training sessions covered the following topics:

- Background information on the CLC evaluation project and the role of classroom observations within the project
- The components or constructs assessed with the ACE protocol and actions that reflect the protocol
- The logistics of conducting site visits and classroom observations
- The expectations for professional demeanor during site visits
- The process of recording observations on laptops and synchronizing to the central database
- The reimbursement process for travel expenses

At the end of the second day, observers completed the necessary paperwork for employment and were asked to visit a National Institutes of Health website to receive the necessary certification for the Learning Point Associates Institutional Review Board.

Background information on the CLC project. The primary investigator began each training session with some background information on the intervention, the evaluation project, and the participating districts. The research summary document written for IES/NCEE was shared with the observers so that they could become more familiar with this project. The use of classroom observation data for examining the fidelity of instruction and the contrast in literacy instruction was discussed as well.

Learning and using the protocol. The majority of the observer training involved familiarizing observers with the constructs being examined in the ACE protocol (the components listed earlier in this appendix) and the specific actions that they might observe that reflect those components. The facilitators had the observers read over the descriptions of a subset of components and then discussed the types of actions that observers might see that would reflect those components. Distinctions between components and actions also were discussed.

Then the observers were provided with paper copies of the ACE protocol and led through the process of recording the presence of a particular construct (that is, the components) and then recording the action that signified that construct and the modality or application in which that action was observed (for example, graphic organizer, verbal, visual, text, or strategy). The process of observing in 10-minute segments, recording notes, and then observing again (repeated throughout the class period) was discussed as well. After the observers acknowledged an understanding of the process of recording what they witnessed, they were given numerous opportunities to practice observing and recording their observations using 5–10 minute film clips of actual instruction.

After attempting to record observations based on a film clip, the observers were asked to discuss which constructs/components they witnessed (and the supporting actions and applications). If inconsistencies arose between the recordings of the facilitators and the observers, the film clip was presented again, and the facilitators would clarify whether constructs/components and actions were present. Once the observers and the facilitators were in agreement regarding the presence/absence of constructs in a particular clip, the group members would watch another film clip and use the ACE protocol to record their observations. This iterative process was done five times across the two-day training. Following these five trials, the observers were asked to view a final film clip and make ratings for the purpose of determining interrater reliability.

Logistics. The evaluation project staff member responsible for scheduling site visits discussed the procedure for scheduling the site visits, showing up at the school, becoming acquainted with the school's layout, and synchronizing databases. This scheduler was tasked with making arrangements for the observations with each school's liaison, collecting the master schedules from the schools, lining up the observers for that site visit, and communicating the visit-specific instructions to the observers (for example, where to park, where to show up, and the liaison's name). The scheduler also made sure that information that observers required for their site visit (class names, room numbers, teachers' names, and subject area) was prepopulated within the databases. As mentioned in chapter 2, classes to be observed were randomly chosen from the master schedule.[150]

Expectations for professional demeanor. A code of conduct was discussed with the observers. This included a discussion of appropriate attire for site visits, what to say to teachers and leaders about the study and the purpose of the observations, and how to conduct the observations in an unobtrusive manner. Throughout this discussion, the facilitators and the principal investigator stressed that the observers were to refrain from engaging teachers in a discussion about their instruction. If asked about the specific constructs that they were looking for, observers were to mention only "actions related to literacy instruction."

[150] This was true except in one district that required teacher consent. In that district, classes to be observed were chosen randomly from the grade 9 core content and reading teachers (year 1) and grade 9 and grade 10 core content and reading teachers (year 2) who consented to being observed.

Prior to the beginning of each class period, the observers were expected to introduce themselves to the teacher and ask him or her where to sit so as to minimize any disruption. The observers were to say nothing throughout the class period; they were to sit in back of the class and record their observations in 10-minute segments separated by 5 minutes of note taking. At the end of each class period, the observers were to thank the teacher and provide him or her with a slip of paper that gives a cursory description of the study and contact information for the co-primary investigator. The observers were to go to their assigned classes following the school's bell schedule. They were asked to minimize contact with students and faculty between classes.

Recording observations on laptops. A project team member who designed the computerized version of the ACE protocol (programmed into a Microsoft Access database) presented the process for synchronizing the laptops prior to a site visit, opening the program once a class had begun, recording the observations during each segment, recording extra notes, and synchronizing the program after the site visit was completed. The database designer and facilitators also demonstrated to observers how to get to the project's SharePoint site where the observer training manual, an electronic version of the ACE protocol, and additional information were stored.

The observers were given laptops containing the necessary software. Extra batteries were provided in case any observers were unable to plug in their laptops during site visits.

Reimbursement for travel. The facilitators provided observers with information on travel reimbursement. The guidelines established by the U.S. Department of Education and the General Services Administration were used to make travel reimbursements for observers and other staff for this project.

Refresher training. Within a week of the two-day training session, the observers were asked to participate in a conference call during which the facilitators once again reviewed the constructs in the ACE protocol and asked if there were any questions. Then the observers were asked to follow an Internet link that had been emailed to them to view another film clip of instruction and make observations of that film clip. This recording of observations gave observers experience in making recording observations on the laptop and synchronizing their program with the central database. During year 1, facilitators examined the observations, and a subsequent call was conducted during which they discussed what they saw during the film clip. During year 2, these observations served as the basis for interrater agreement calculations.[151]

Two other conference calls were conducted with the observers. The first call was conducted to answer questions and address any concerns that arose during the pilot observations. A second call was conducted at the beginning of year 2 with observers who were rehired from the previous year to refresh their understanding of the ACE protocol and the process for recording observations.

Using the observation protocol

The observation protocol was computerized as a Microsoft Access database on laptop computers that observers used in the classroom. The observers also had paper copies of the protocol on hand should any problems develop with their laptops. Prior to arriving at a school for a site visit

[151] Interrater agreement calculations for year 1 were based on a pilot set of observations conducted in eight of the project schools by pairs of observers. These observers followed the same schedule of classes during their visits to those schools.

(that is, a series of observations), the observers synchronized their databases to a centrally administered master database. The synchronization process allows certain fields within the protocol to be prepopulated for each observer (for example, school name, teachers' names, and the subject area of the class). At the end of the site visit, the observers synchronized their databases again, downloading their data from the day's observations to the central database.

The nesting within the computerized protocol is evident when an observer identifies a construct as being present (for example, setting purpose). When the construct is identified, the observer is prompted to indicate which actions reflective of the construct were apparent in instruction (for the "setting purpose" example, possible actions were "explicitly identifies purpose," "supports students in identifying purpose," and "purpose is implied"). If a specific action is present, the observer is prompted again to specify how the action is applied (that is, oral, writing, strategy, or graphic organizer).

The protocol begins with general questions about the observation session, such as the time period and date of the observation and the given observer's name. Next is a section containing fields for the name of the district, the name of the school, the grade, the instructional focus (that is, ELA, science, social studies, mathematics, or reading), student demographics and characteristics within the class (for example, special education or English language learners [ELLs]), and the way in which students are grouped within the class (for example, whole group or small group). With the exception of class grouping, most fields in this section are prepopulated for observers with information provided by the school (that is, master schedule and student demographics within a class). The body of the protocol is filled in during the observational segments. This includes the identification of constructs, actions, and applications demonstrated by the teacher during the segment.

The end of the protocol contains items that focus on the coherence of instruction. Four items inquire about scaffolded instruction, discourse, engagement, and assessment; these items capture aspects of classroom instruction that reflect specific facets of the CLC program. Scaffolded instruction (modeled, shared, guided, or independent), discourse (none, teacher, teacher-student, student-and-teacher and/or student-and-student, or interactive student-and-student), and assessment (none, student self-assessment, informal oral teacher assessment, informational written teacher assessment, or formal teacher assessment) are based on a categorical scale, while engagement (active student, passive student, minimal, moderate student, or active) is based on an ordinal scale.

ACE Protocol—Investigations of Instructional Technique

Start time _____ End time _____

Date _____ Observer _____

Background

District _____ School _____

Grade _____ Teacher gender ☐ M ☐ F Teacher _____

Total number of students _____ Total boys _____ Total girls _____

Student categories

	All	Some	Number	None
Special Education	☐	☐	_____	☐
English Language Learners	☐	☐	_____	☐
Title I	☐	☐	_____	☐
Other	☐	☐	_____	☐

(explain)_____

Classroom

Instructional focus

☐ Language Arts ☐ Science ☐ Social Science ☐ Math ☐ Reading

Instructional modality

☐ Whole group ☐ Small group ☐ Pairs ☐ Independent ☐ Other _____

Desk arrangement:

☐ Rows ☐ Pairs ☐ Groups ☐ Tables ☐ Other _____

	Y	N
Seating arrangement allows all students to easily view delivery of instruction.	☐	☐
Room arrangement allows for whole group, small group, and independent work.	☐	☐

Materials

	Y	N
Student materials are arranged for easy access.	☐	☐

Lesson materials:

☐ Textbooks ☐ Trade books ☐ Teacher created ☐ Other

Briefly described the materials used:

By the teacher _____

By the students _____

Significant classroom disruption: ☐ _____

CONSTRUCT	ACTION			APPLICATION	
	ACTIVATING KNOWLEDGE				
Setting purpose	Explicitly identifies purpose			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Supports students in identifying purpose			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Purpose is implied			Oral	
				Writing	
				Strategy	
				Graphic organizer	
Building background	Makes connections to previous texts			Oral	
knowledge				Writing	
				Strategy	
				Graphic organizer	
	Makes connections to previous instruction			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Makes connections to students' personal knowledge			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Provide additional information for insufficient background knowledge (before and during learning)			Oral	
				Writing	
				Strategy	
				Graphic organizer	
Identifying text structure	Explicitly identifies text structure			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Supports student identification of text structure			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Text structure is implicitly identified			Oral	
				Writing	
				Strategy	
				Graphic organizer	

Predicting content	Explicitly elicits predictions			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Requests support and/or verification of predictions (discussion)			Oral	
				Writing	
				Strategy	
				Graphic organizer	
Highlighting essential vocabulary	Presents words that are central to understanding the text/lesson			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Introduces words in meaningful contexts			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	CONSTRUCTING KNOWLEDGE				
Determining importance	Establish key ideas and concepts			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Organize essential important learning			Oral	
				Writing	
				Strategy	
				Graphic organizer	
Inferring	Draw conclusions based on connections to text			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Draw conclusions based on connections to class activity			Oral	
				Writing	
				Strategy	
				Graphic organizer	
Questioning / Clarifying	Generate questions to clarify word and/or text meaning			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Generate questions to deepen comprehension			Oral	
				Writing	
				Strategy	
				Graphic organizer	

	Generate questions to critically evaluate text			Oral	
				Writing	
				Strategy	
				Graphic organizer	
Synthesizing / Summarizing	Develop a summary using information from text			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Develop a summary using information from class activity			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Generate synthesis integrating background knowledge with information from text			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Generate synthesis integrating background knowledge with information from class activity			Oral	
				Writing	
				Strategy	
				Graphic organizer	
Strengthening content knowledge	Knowledge consolidated by organizing information			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Consolidating knowledge at the sentence or paragraph level			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Consolidating knowledge at the essay or theme level			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	EXTENDING KNOWLEDGE				
Monitoring learning	Monitor learning to revise and refine vocabulary knowledge			Oral	
				Writing	
				Strategy	
				Graphic organizer	
	Monitor learning to revise and refine content knowledge			Oral	
				Writing	
				Strategy	
				Graphic organizer	
Reflecting on learning	Recalling/reviewing essential learning			Oral	

				Writing	
				Strategy	
				Graphic organizer	
	Extending essential learning			Oral	
				Writing	
				Strategy	
				Graphic organizer	

STRATEGIES

1. Brainstorming	7. Planning/Organizing Routine	13. Response Journals	19. Word Maps – morphology/decoding
2. Bridging Strategy	8. Prediction Strategy	14. Structured Overview	20. 7-Step Vocabulary Process
3. Concept Routine	9. Question-Answer Relationships (QAR)	15. Think Aloud	21. Other
4. FRAME/Framing Routine	10. Question Exploration Routine	16. Vocabulary LINCing Routine	
5. KWL	11. Quick Writes	17. Webbing	
6. Paragraph Writing Strategy	12. Reciprocal Teaching	18. Word Maps – concepts	

COHERENCE

Scaffolded Instruction

Scaffolding not evident	Modeled	Shared	Guided	Independent

Discourse

No discourse	Teacher discourse	Teacher-to-student discourse	Student-and-teacher and/or student-and-student discourse	Interactive student-and-student discourse

Engagement

Active student disengagement	Passive student disengagement	Minimal engagement	Moderate student engagement	Active engagement

Assessment

No assessment	Student self-assessment	Informal oral teacher assessment	Informal written teacher assessment	Formal teacher assessment

Notes:

Appendix C. Characteristics of the Student Analysis Samples

This appendix supplements the description of the student analysis samples presented in chapter 2. As described in that chapter, the target population for the study consists of all grade 9 and grade 10 students enrolled in the study schools. However, data on reading achievement and academic performance were not available for all students because not all parents consented to the release of school records and achievement testing. Moreover, even among students for whom consent was obtained, some did not take the GRADE reading assessment due to absences on the day of testing and other factors. Thus, student impacts in this report are estimated using students who have data on the outcome of interest. Two analysis samples are used:

- **The school records sample.** This sample is used to estimate impacts on students' course performance outcomes (credit accumulation and grade point average [GPA]). It includes all consenting students enrolled in the study schools on the last day of the school year for which course transcript data are available.
- **The GRADE respondent sample.** This sample is used to estimate impacts on reading achievement (reading comprehension and reading vocabulary scores). It includes consenting students enrolled in the study schools in the spring at the time of test administration and who took the GRADE assessment.

This appendix describes the construction and characteristics of the two analysis samples used in this report. The first section describes the construction of the school records sample; the second section provides further information about this sample by examining year-to-year student mobility in the school records sample. The third section describes the construction of the GRADE respondent sample. The fourth section concludes with a comparison of the characteristics of students in Content Literacy Continuum (CLC) and non-CLC schools, in each of the two analysis samples (school records sample and GRADE respondent sample).

In some of the tables in this appendix, we compare the baseline characteristics and achievement of two groups of students (students in CLC schools and non-CLC schools or respondents and nonrespondents). Because many hypothesis tests are conducted in these tables (one for each student characteristic), there is an increased probability of concluding that a particular difference is statistically significant when in fact it is not (a type I error or a false positive).[152] For this reason, an omnibus test is used to test for a systematic or overall difference between the characteristics of students in different groups. This test is reported at the bottom of each table in this appendix. If the omnibus test is not statistically significant, then this means that a statistically significant difference for any given background characteristic may be due to chance.

Finally, as in other parts of the report, the analyses presented in this appendix are based on the 28 schools that participated in both year 1 and year 2. (See appendix M for the characteristics of

[152] In particular, one would expect to see one false positive for every 20 hypothesis tests conducted when $p < .05$ is selected as the criterion for statistical significance.

students in the 33 schools that participated in year 1, as well as first-year impact findings for these study schools.)

Creation of the school records sample

The construction of the school records sample, for each study year and grade level, is depicted in figures C.1–C.3. Each figure starts with the number of students officially enrolled in the 28 schools in the study sample, which is the target population for the CLC study.[153] The figures then demonstrate how the school records sample is created by tracing the reasons why some students in the enrolled target population are not included (or are re-included) in the analysis sample.

These figures show that there are three reasons why the number of students in the school records sample differs from the officially enrolled target population:

1. **Non-consent:** in districts where passive parental consent was required for students' participation in the study, not all parents consented to the release of their child's school records.[154]
2. **Differences in timing:** the timing of enrollment counts may differ between the two data sources: school records data were extracted for students enrolled on *the last day of school* while official enrollment numbers are based on spring enrollment.
3. **Differences in the definition of grade level:** the definition of "grade level" in the school records sample differs from students' official grade level. For reasons explained in chapter 2, retained grade 9 students are included in the grade 10 sample, whereas in the official enrollment numbers, these retained students are counted as being in grade 9.[155]

The first two factors—non-consent and differences in timing—are represented in the box labeled "records not provided" in figures C.1–C.3, and they account for most of the difference between official enrollment and the school records sample. An important question here is whether these factors are similarly relevant in CLC and non-CLC schools. This question is examined more closely in table C.1, which shows the number of students in the school records data, as a percentage of official enrollment, by study year and grade level, for CLC and non-CLC schools. That is, the table compares the samples in the middle set of boxes ("students with school records data") to the top boxes based on enrollment ("official enrollment") in figures C.1–C.3.[156] Note that the rates in this table do not exactly correspond to the rates obtained using the sample

[153] These numbers are based on enrollments reported on state and district websites.

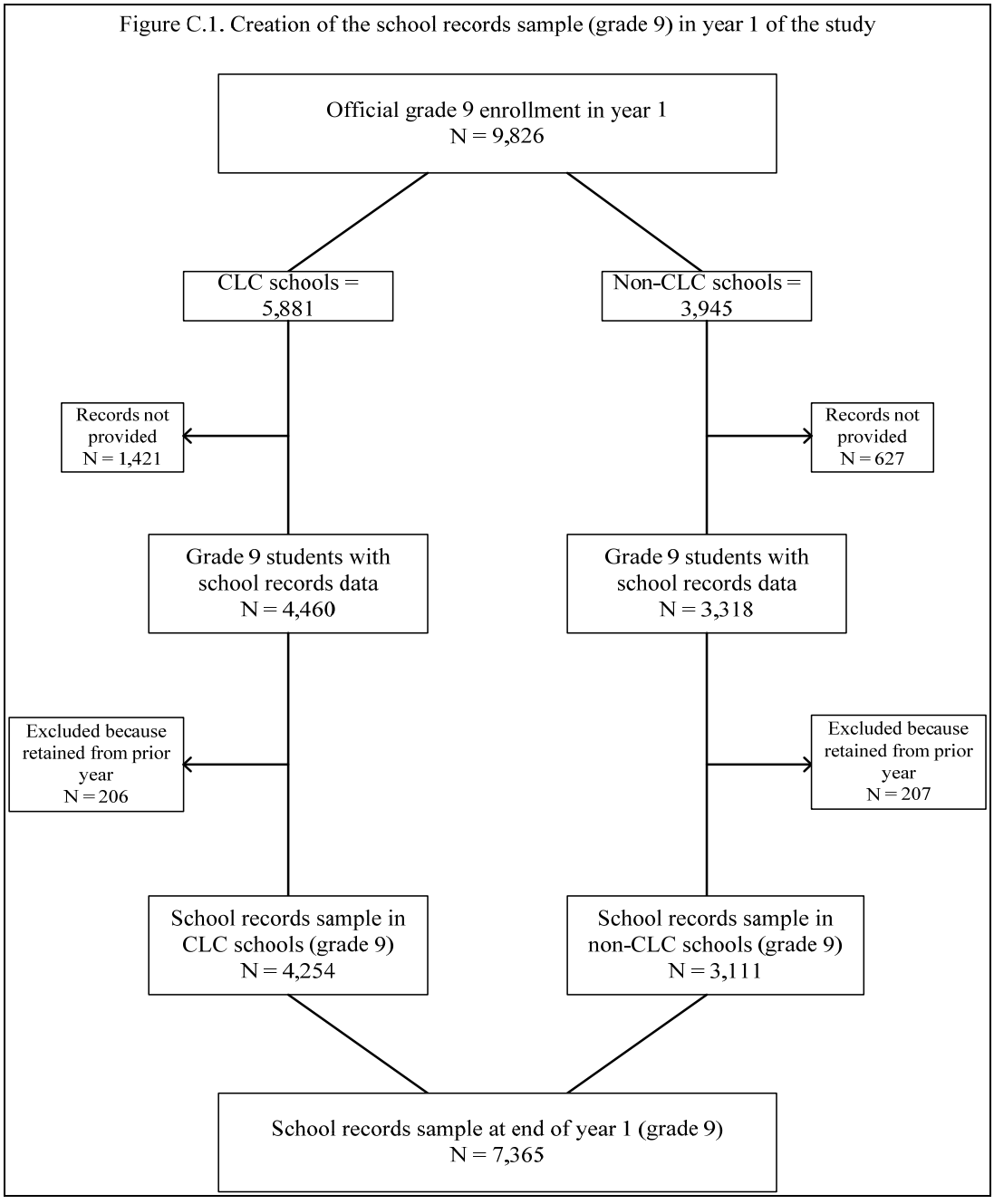
[154] In five school districts in year 1 and four districts in year 2, passive parental consent was required prior to sharing students' school records and administering the GRADE assessment. In these districts, parental notification letters were sent to students' families, alerting them to the research activities and asking them to notify the school if they did not want their child to participate in the study. In the remaining study districts, no parental consent was required (either active or passive).

[155] Another possible reason for the discrepancy between official enrollment and the school records sample is that data extraction by school districts was incomplete. However, because this type of error cannot be assessed, we assume that school districts did in fact send us complete data for all consenting students enrolled at the end of the year.

[156] For the purposes of Table C.1, students in the school records sample are categorized on the basis of their official grade level (that is, retained grade 9 students are grouped with grade 9 students).

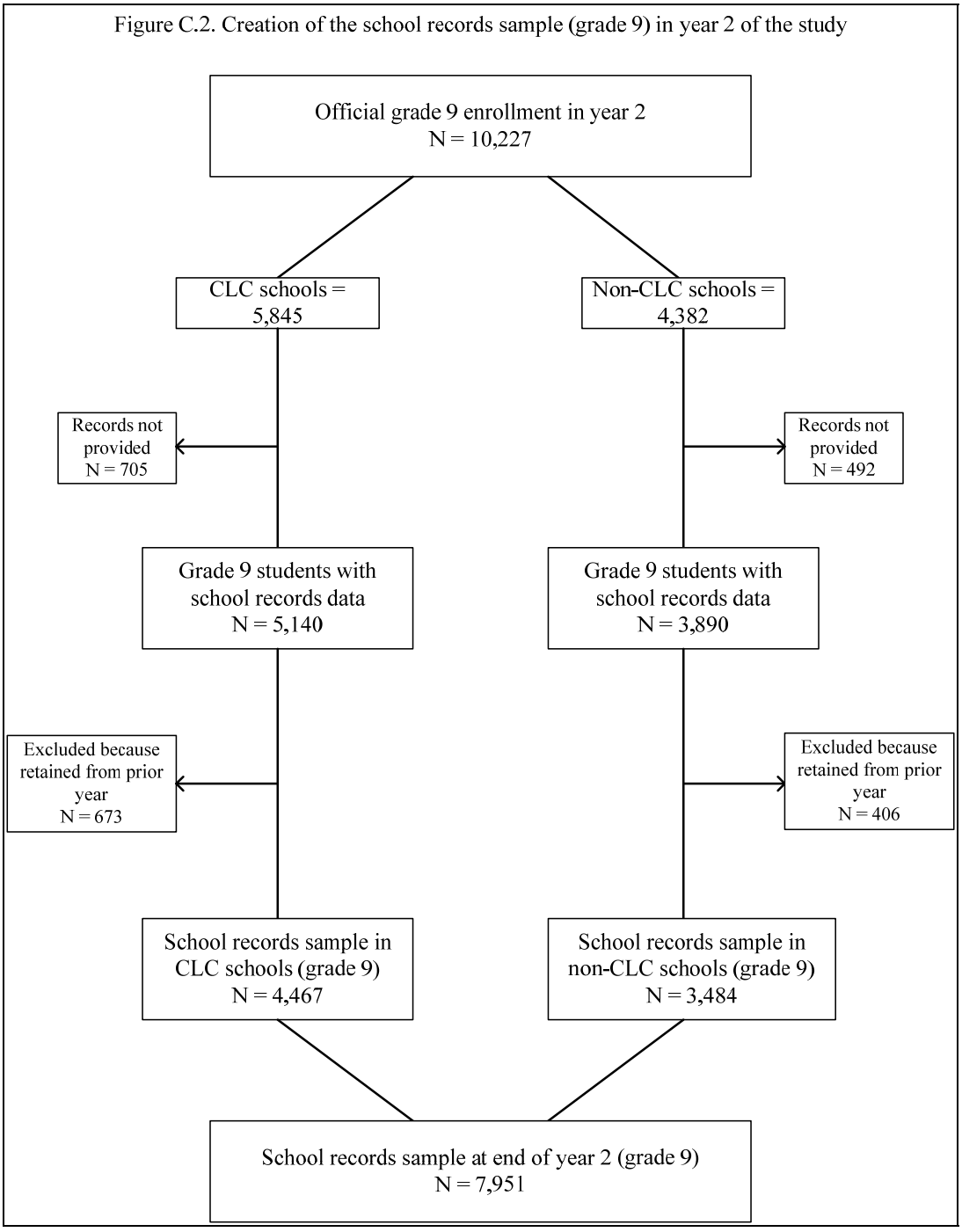
numbers in the figures, because the rates in table C.1 are for the average school in the study (rather than the average student). This is true of all tables comparing response rates in CLC and non-CLC schools in this appendix.[157]

Figure C.1. Creation of the school records sample (grade 9) in year 1 of the study



[157] Specifically, the overall difference between the two groups of schools—as well as mean levels—is weighted in the same way as the impact findings, that is, by the number of CLC schools in each random assignment block (see box 2.1 and appendix E). Thus the findings in this table are for the average school in the sample. When estimating the difference between groups, a two-level model (students within schools) is used to account for clustering and the blocking of random assignment.

Figure C.2. Creation of the school records sample (grade 9) in year 2 of the study



Note: Retained grade 9 students are included in the grade 10 sample.

Figure C.3. Creation of the school records sample (grade 10) in year 2 of the study

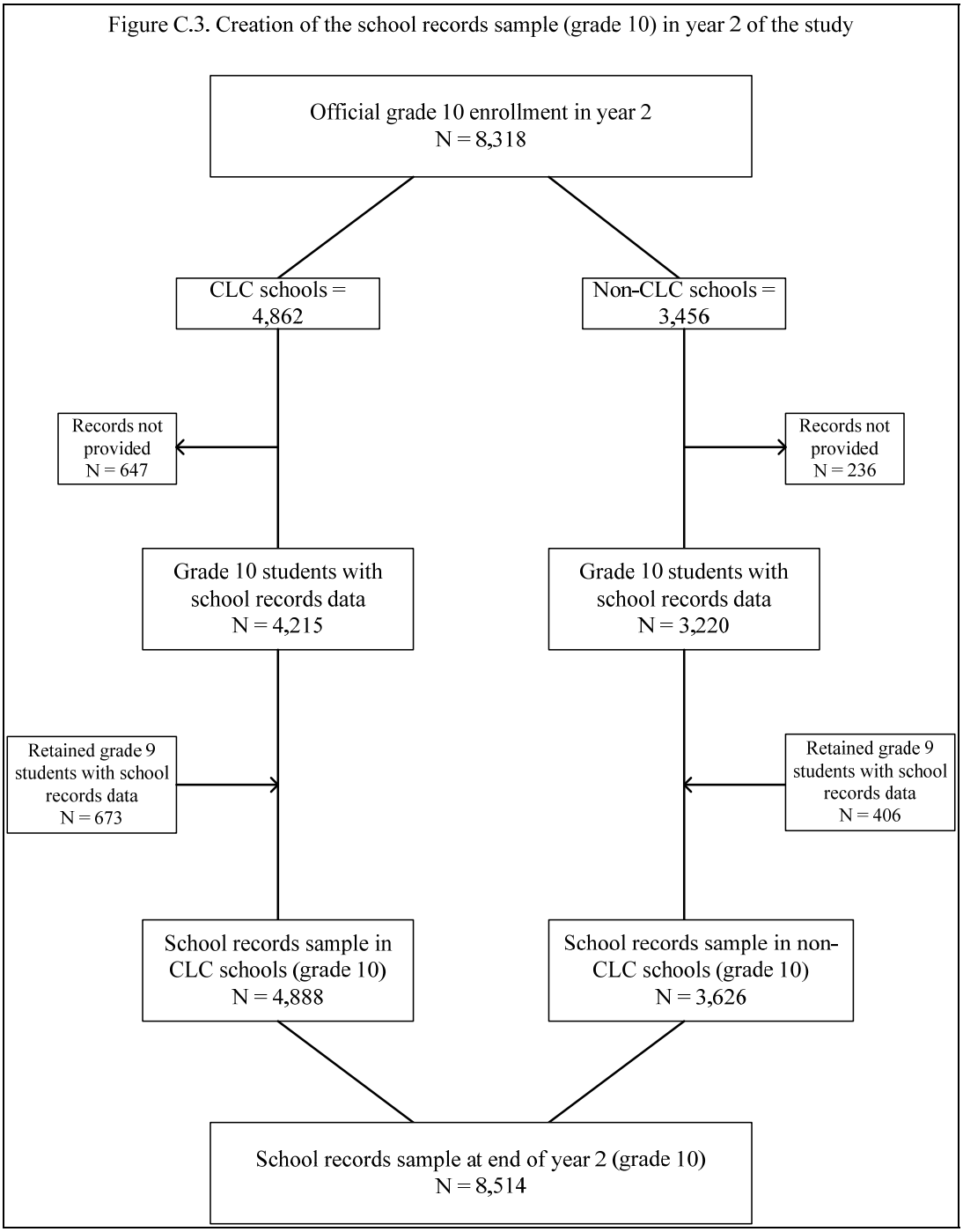


Table C.1. Students with school records data, as a percentage of official enrollment

	CLC group	Non-CLC group	Estimated difference	*p*-value
Year 1				
Grade 9 students with school records data (%)	79.6	88.9	−9.2	0.313
Official enrollment				
Students (total = 9,826)	5,881	3,945		
Schools (total = 28)	15	13		
Year 2				
Grade 9 students with school records data (%)	91.0	91.4	−0.4	0.862
Official enrollment				
Students (total = 10,227)	5,845	4,382		
Schools (total = 28)	15	13		
Grade 10 students with school records data (%)	88.9	94.8	−5.9	0.262
Official enrollment				
Students (total = 8,318)	4,862	3,456		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Students' grade level is the grade reported in their school transcript. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

Source: Calculations from school records data obtained as part of the Content Literacy Continuum study and official school enrollment numbers reported on state and district websites.

As shown in table C.1, in the average CLC school, records data are available for 80 percent of enrolled students in the first year of the study, and 89–91 percent of officially enrolled students in the second year, depending on the grade level. Target population coverage rates do not differ by a statistically significant amount between CLC and non-CLC schools, which indicates that the balance achieved by random assignment is not compromised by differential coverage rates in the two groups of schools.

Based on this table, about 11–20 percent of officially enrolled students are excluded from the analysis. As noted above, these exclusions occur for two reasons: non-consent and timing differences. However, the second factor—timing differences—is simply an accounting issue and does not represent a true exclusion of students in the target population. It is therefore important to disentangle how many students are excluded due to non-consent.

Unfortunately the role of non-consent and timing cannot be directly isolated. However, a sense of their relative importance is obtained by looking at coverage rates in districts where consent was *not* required, because in these districts the only relevant factor is timing differences. These results are presented in table C.2. As seen here, the target population coverage rate in "no

consent" districts ranges from 95–100 percent, depending on the study year and grade level. One may assume that in these districts, the discrepancy between official enrollment and the school records data (ranging from 0–5 percent) is due to differences in the timing of student counts.

Table C.2. Students with school records data, as a percentage of official enrollment, by consent requirements

	Passive consent districts	No consent districts
Year 1		
Grade 9 students with school records data (%)	74.5	94.7
Number of districts (total = 8)	5	3
Number of schools (total = 28)	17	11
Year 2		
Grade 9 students with school records data (%)	82.9	99.8
Grade 10 students with school records data (%)	86.4	96.3
Number of districts (total = 8)	4	4
Number of schools (total = 28)	16	12

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Students' grade level is the grade reported in their school transcript.

Source: Calculations from school records data obtained as part of the Content Literacy Continuum study and official school enrollment numbers reported on state and district websites.

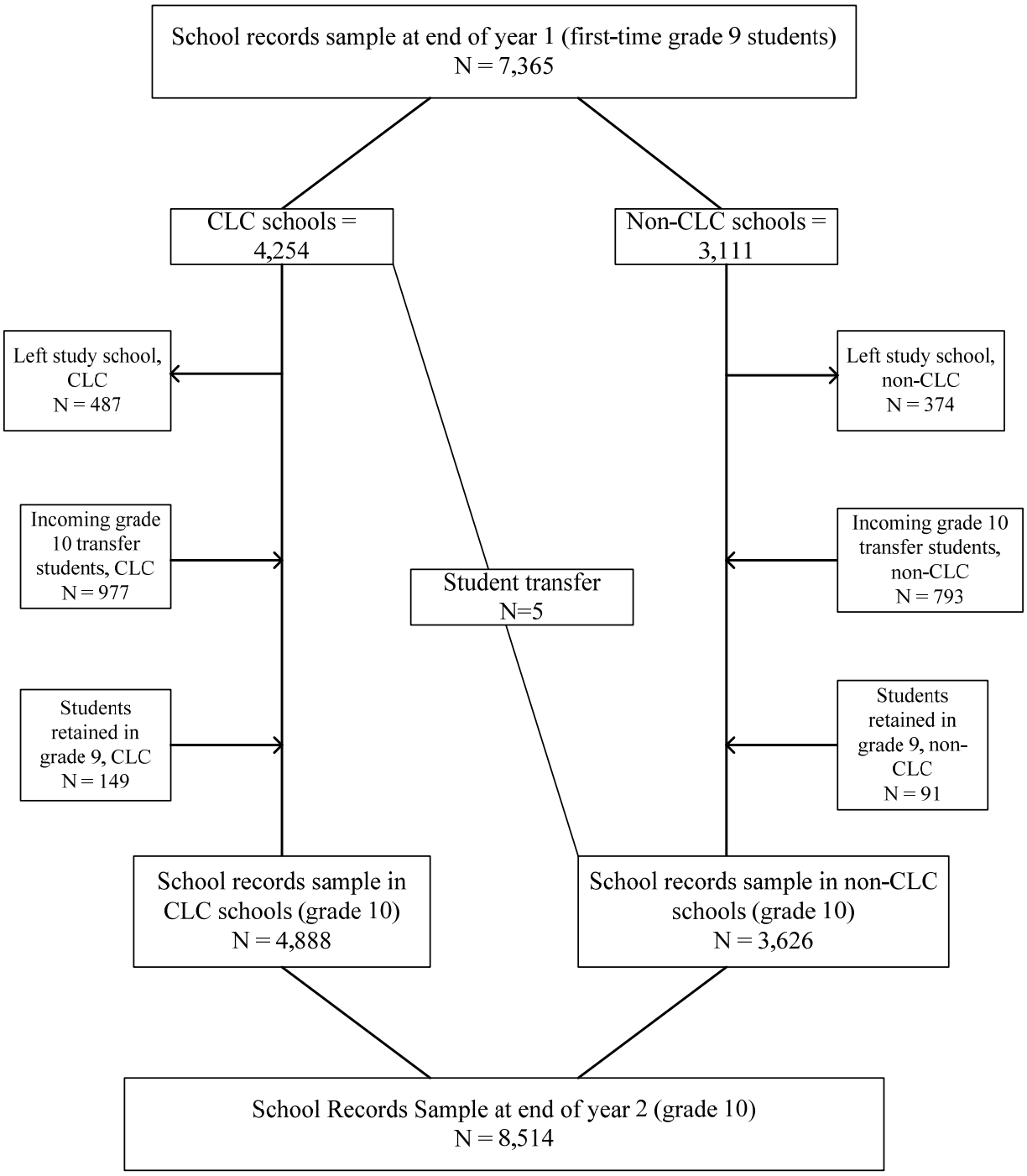
This suggests that the *true* target population coverage rate for the CLC study is somewhat higher than what is shown in Table C.1 or conversely that the true percentage of the target population excluded from the analysis is lower than 11–20 percent. For example, if we assume that discrepancies due to timing differences are the same in passive consent districts as in no-consent districts, then the percentage of the true target population excluded from the sample would be about 8–15 percent. Thus, the school records sample does cover the vast majority of students enrolled in the study schools. This suggests that impact findings based on the school records sample are likely to be generalizable to the full target population.

Mobility and attrition in the school record sample across study years

Because the CLC framework was implemented for two school years, there is overlap between the grade 9 school records sample in year 1 and the grade 10 school records sample in year 2. The two samples are not exactly the same because some students left the study schools after year 1, and other students transferred into the study schools.

Figure C.4. Student mobility from year 1 to year 2 of the study, school records sample

Figure C.4. Student mobility from year 1 to year 2 of the study, school records sample



The process by which the grade 9 sample in year 1 becomes the grade 10 sample in year 2 is mapped in figure C.4. The factors that underlie this process are important for understanding the characteristics of grade 10 students in the school records sample. Three factors are of particular interest: attrition from the study sample, mobility into the study schools (transfers), and grade retention. The importance of these three factors for CLC and non-CLC schools is shown in table C.3:

- **Attrition from the study sample.** Some grade 9 students enrolled in the study schools in year 1 of the study were no longer enrolled in a study school in year 2, either because they dropped out of school or because they transferred to another school district. This mobility out of the study schools affects the number of grade 10 students in the analysis as well as their characteristics, as students who drop out of school are likely to be lower achieving. Attrition from the study schools can be measured by looking at the percentage of grade 9 students in the school records sample in year 1 who were no longer in the sample in year 2. In the average CLC school, attrition from the school records sample from year 1 to year 2 of the study is 13 percent, as shown in table C.3.
- **Mobility into the study schools.** Some grade 10 students enrolled in the study schools in year 2 were not enrolled in a study school in the previous school year—they are transfer students. In CLC schools, these grade 10 transfer students received one year of CLC services, and other grade 10 students received two consecutive years of the framework. Thus student mobility into the study schools dilutes the "dosage" of the CLC intervention among grade 10 students. In the average CLC group school, 19 percent of grade 10 students in the school records sample were new to the school in that year or conversely 81 percent of students were enrolled in the same school the previous year (see table C.3).
- **Grade retention.** As explained earlier and in chapter 2, the sample of grade 10 students also includes students who were retained in the previous school year. Because retained students are likely to be lower achieving than regular grade 10 students, the percentage of retained students in the grade 10 sample can affect the overall achievement level of this group of students. In the average CLC school, 13 percent of students in the grade 10 school records sample were retained from the previous school year (see table C.3).

As also shown in table C.3, attrition, mobility, and retention rates do not differ by a statistically significant amount between CLC and non-CLC schools. Thus, overall, the balance between CLC and non-CLC schools achieved by random assignment is not compromised by differential mobility in/out of the study sample or by differential retention rates.

Creation of the GRADE respondent sample

The construction of the GRADE respondent sample, for each study year and grade level, is depicted in figures C.5–C.7. Each figure starts with the school records sample and then tracks the reasons why some students in the school records sample are not included (or re-included) in the GRADE respondent sample.

Table C.3. Student attrition and mobility between study years, school records sample

	CLC group	Non-CLC group	Estimated difference	*p*-value
Attrition from the study schools				
Grade 9 sample in year 1 who are no longer enrolled in a study school in year 2 (%)	13.1	13.6	−0.5	0.824
Sample size				
Students (total = 7,365)	4,254	3,111		
Schools (total = 28)	15	13		
Mobility into the study schools (transfers)				
Grade 10 sample in year 2 who are new to the study sample (%)	18.6	21.3	−2.7	0.205
Sample size				
Students (total = 8,514)	4,888	3,626		
Schools (total = 28)	15	13		
Mobility across grade levels (retention)				
Grade 10 sample in year 2 who are retained grade 9 students (%)	13.0	12.1	0.8	0.796
Sample size				
Students (total = 8,514)	4,888	3,626		
Schools (total = 28)	15	13		

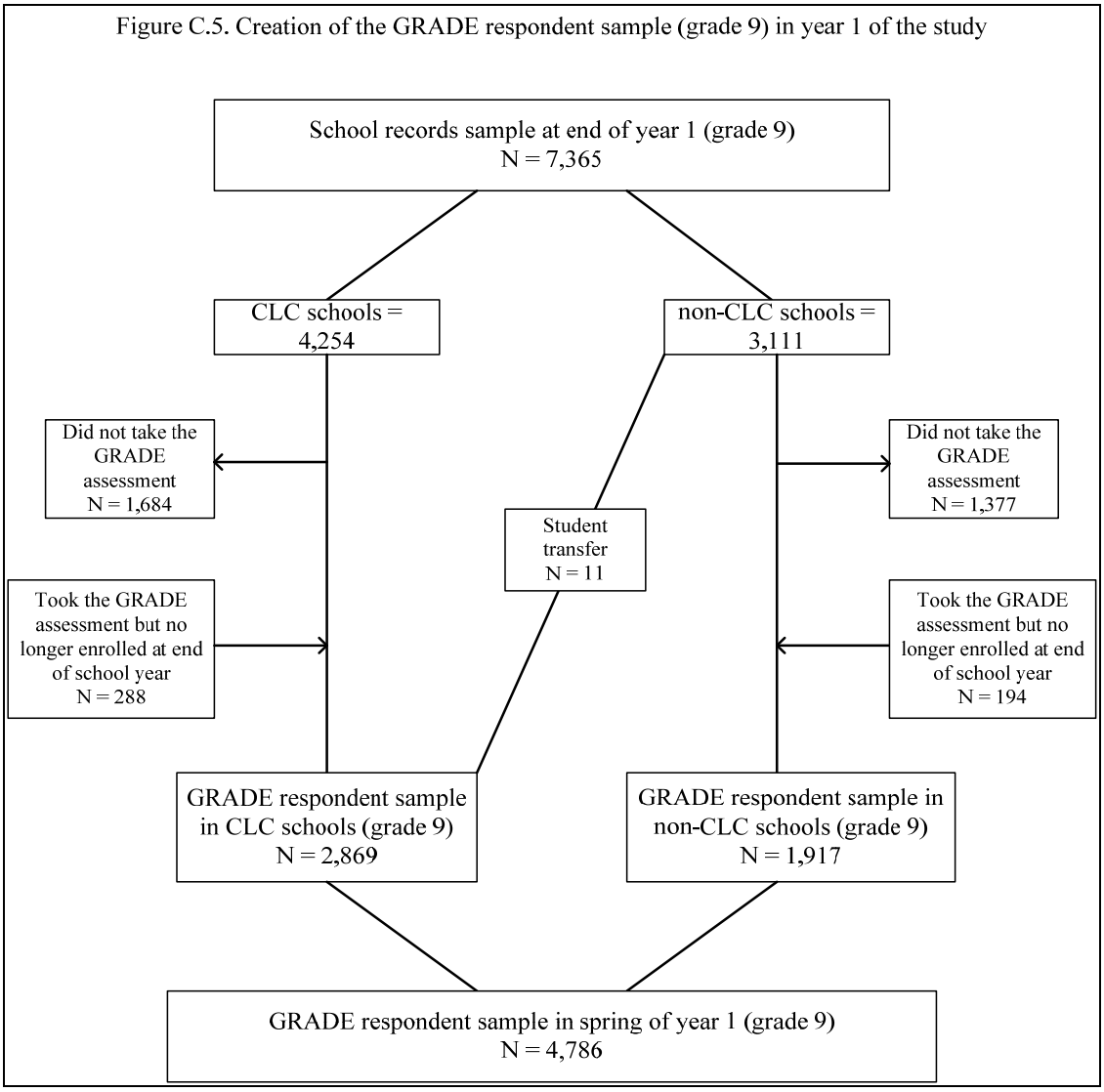
* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. The grade 9 sample includes students who are in grade 9 for the first time; the grade 10 sampled includes grade 10 students and retained grade 9 students. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

Source: Calculations from school records data obtained as part of the Content Literacy Continuum study.

These figures show that the GRADE respondent sample is not a perfect subset of the school records sample due to student mobility. Some students who took the GRADE assessment in the spring (April, May) were no longer enrolled in the district on the last day of the school year. Thus the GRADE respondent sample includes a small number of students who moved out of the school district during the spring and who are not in the school records sample (in the figures, this is the box labeled “took the GRADE assessment but no longer enrolled at end of school year”). The figures also show that some students transferred from a CLC school to a non-CLC school (or vice versa) between the time of GRADE testing and the end of the school year (“student transfer”).

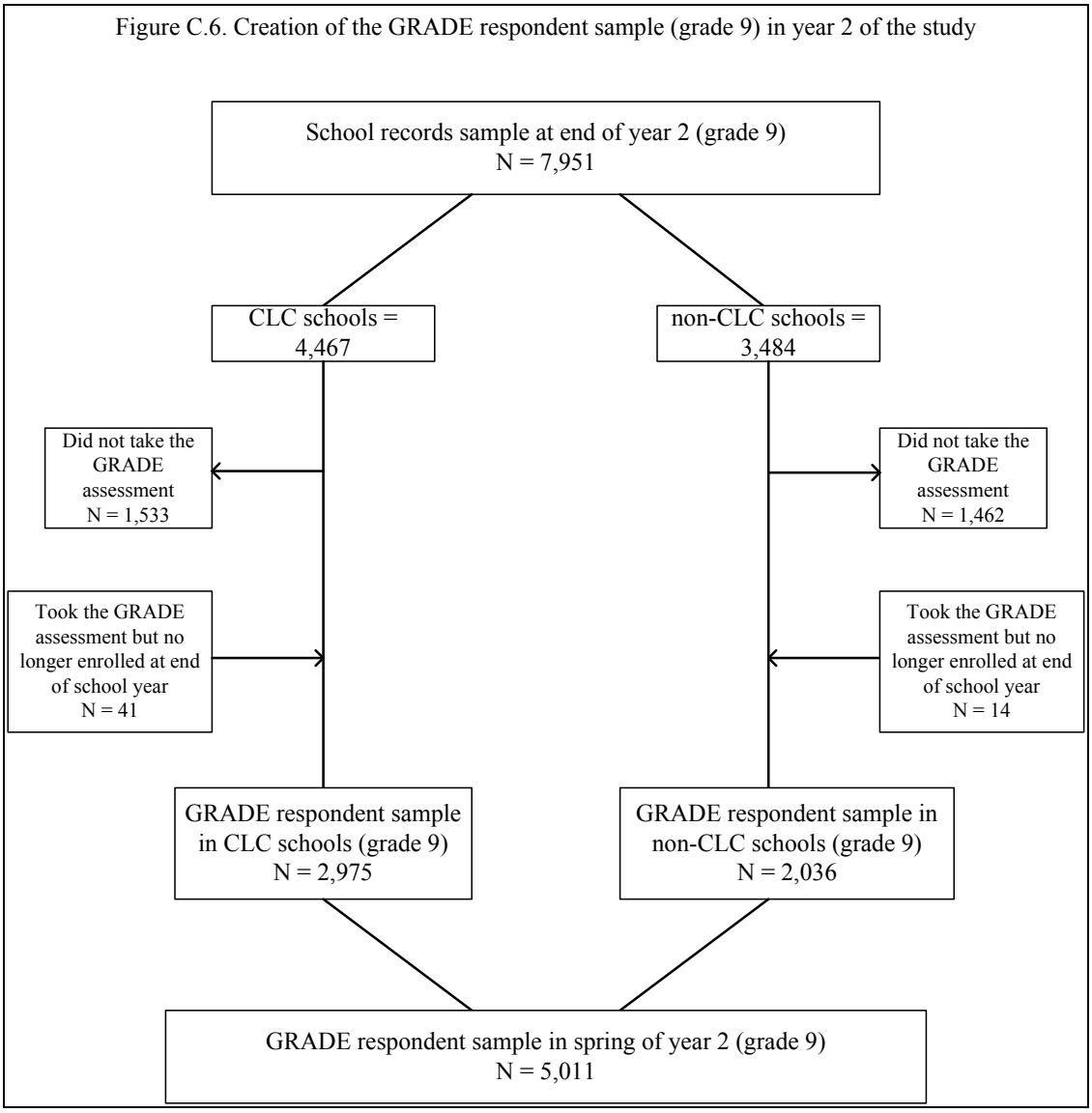
Figure C.5. Creation of the GRADE respondent sample (grade 9) in year 1 of the study



On balance, however, the GRADE respondent sample includes fewer students than the school records sample. The GRADE response rates in CLC and non-CLC schools are compared in table C.4. In the average CLC school, 61 percent of students in the school records sample took the GRADE assessment in year 1; in year 2, 64 percent of grade 9 students took the test, and 57 percent of grade 10 students took the test. GRADE response rates do not differ by a statistically significant amount among CLC schools; this indicates that the balance achieved by random assignment is not compromised by differential GRADE response rates across the two groups of schools.

Response rates less than 100 percent threaten the generalizability of the findings, because students in the analysis sample may not be representative of all students in the study schools (as measured by students in the school records sample). Specifically, impacts on GRADE scores—because they are based on only a subset of students in the school records sample—may not be generalizable to all students.

Figure C.6. Creation of the GRADE respondent sample (grade 9) in year 2 of the study



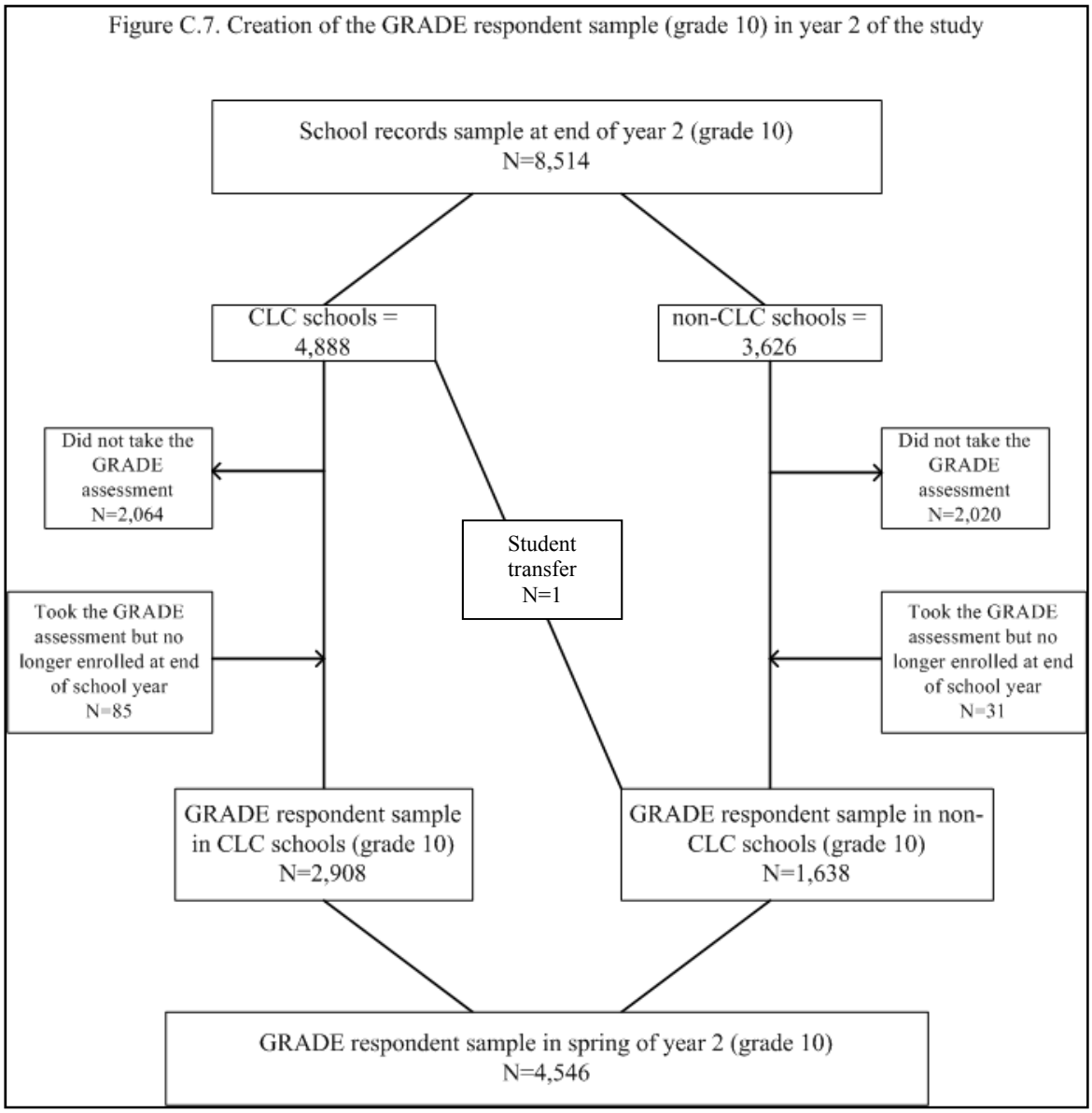
To better understand the external validity of the findings, the first step is to take a closer look at the reasons why the GRADE respondent sample includes fewer students than the school records sample. There are three causes of non-response:

1. **Absenteeism:** there were student absences on the day of GRADE testing. Although makeup testing sessions were scheduled for these students, some were absent for the retest session as well.
2. **Testing exclusions:** many ESL and special education students were not tested because the GRADE cannot be administered to students who require certain types of testing

accommodations (for example, alternate language or large-print test forms, audio recordings, an aide, or a translator).[158]

3. **Sampling of students:** in order to ease testing burden, some districts chose to test random samples of their students rather than all students. Four schools chose this option in year 1, and five schools chose it in year 2.

Figure C.7. Creation of the GRADE respondent sample (grade 10) in year 2 of the study



[158] The only allowable testing accommodation was additional time.

Table C.4. GRADE respondent sample as a percentage of the school records sample

	CLC group	Non-CLC group	Estimated difference	*p*-value
Year 1				
Grade 9 sample				
Students with a GRADE score (%)	60.8	59.2	1.6	0.811
Sample size (school records sample)				
Students (total = 7,365)	4,254	3,111		
Schools (total = 28)	15	13		
Year 2				
Grade 9 sample				
Students with a GRADE score (%)	64.2	65.3	−0.2	0.777
Sample size (school records sample)				
Students (total = 7,951)	4,467	3,484		
Schools (total = 28)	15	13		
Grade 10 sample				
Students with a GRADE score (%)	56.7	50.1	6.6	0.102
Sample size (school records sample)				
Students (total = 8,514)	4,888	3,626		
Schools (total = 28)	15	13		

* p-value ≤ .05.

Note: This table is based on the 28 schools that participated in both years of the study. The grade 9 sample includes students who are in grade 9 for the first time; the grade 10 sample includes grade 10 students and retained grade 9 students. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

Source: Calculations based on school records and reading assessment data collected for the Content Literacy Continuum study. School records were obtained for students enrolled in the study schools at the end of each school year. The GRADE reading assessment was administered in the spring of each study year.

Some of these factors are a more serious threat than others to the generalizability of the impact findings. The last cause—the random sampling of classrooms for testing—is least problematic. Students sampled for GRADE testing should be a representative sample of all students in the study schools, and thus the generalizability of the impact findings should not be affected by the sampling of classrooms. However, the first two causes of nonresponse—student absences and the exclusion of ESL and special education students—are more problematic. Students absent on the day of testing or who are classified as ESL or special education are likely to differ in observed and unobserved ways from students who do not have these characteristics. Therefore the exclusion of these students from the analysis potentially can affect the representativeness of the

GRADE respondent sample and thus the external validity of estimated impacts on reading achievement.[159]

GRADE response rates among students in the school records sample, by student subgroup (ESL, special education, neither) and by whether or not students were sampled for testing are presented in table C.5 to examine the relative importance of these factors.[160] As seen in the table, GRADE response rates are especially low among special education students and in schools where students were sampled for testing, as expected. However, because special education students are few in number—and relatively few schools sampled students for testing—low response rates in these subpopulations play a relatively minor role in explaining GRADE non-response in the overall sample. For example, among students who are *not* in special education or ESL—and who are enrolled in schools where students were *not* sampled for testing—the GRADE response rate is 71 percent in year 1 and 74 percent in year 2. These numbers are about 10 percentage points higher than the overall response rates in table C.4. This indicates that sampling and testing restrictions explain only about a quarter of GRADE non-response. Hence, student absenteeism appears the primary cause of non-response.

Because absenteeism appears to be driving non-response, it is important to examine whether the GRADE respondent sample is representative of students in the study schools (as measured by students in the school records sample), because students absent on the day of testing are likely to be different from students who were present. The background characteristics of students in the school records sample who took the GRADE assessment are compared to the characteristics of students who did not in tables C.6–C.8.[161] As seen in these tables, students who took the GRADE are higher performing on average than students who did not.[162] This difference arises because students who did not take the GRADE—absentees and students requiring special testing accommodations—are likely to be lower achieving. Statistical tests confirm that in terms of their background characteristics, students who wrote the GRADE are systematically different from students who did not take the test. This means that the CLC impacts on reading achievement in this report may not be generalizable to all students enrolled in the study schools in the spring, nor to students who did *not* take the GRADE (absentees, and ESL and special education students requiring special testing accommodations).

[159] If student absences are different in CLC schools compared with non-CLC schools, then student absenteeism also can create an imbalance between the background characteristics and prior achievement of students in CLC and non-CLC schools and compromise the experimental study design. However, as noted earlier, GRADE response rates in CLC and non-CLC schools do not differ by a statistically significant amount, which suggests that student absenteeism was similar in CLC and non-CLC schools.

[160] Unlike the values by program group in table C.4, the values in Table C.5 are not weighted by the number of schools in each block—they are simple student-level means.

[161] In these tables, the sample size for students who took the GRADE test does not exactly match the number of students in the GRADE respondent sample; this discrepancy arises because some students in the GRADE sample are not in the school records sample.

[162] When estimating differences between respondents and non-respondents, a two-level model (students within schools) is used to account for clustering and the blocking of random assignment. Unlike tables that compare CLC and non-CLC schools, the average levels and the overall difference between respondents and non-respondents are *not* weighted by the number of schools, and therefore the findings in these tables are for the average student in the sample.

Table C.5. GRADE response rates for student and school subgroups

	School Subgroups	
	Schools that tested all students	**Schools that sampled**
Year 1 (grade 9)		
Students in school records sample with a GRADE score (%)		
All students	63.4	37.4
Students requiring testing accommodations		
English as a second language (ESL) students	81.3	36.6
Special education students	33.0	31.8
Students not requiring testing accommodations (non-ESL and non-special education)	70.9	39.2
Sample size (school records sample)		
Students (total = 7,365)	5,966	1,399
Schools (total = 28)	24	4
Year 2 (grades 9 and 10)		
Students in school records sample with a GRADE score (%)		
All students	63.6	34.2
Students requiring testing accommodations		
English as a second language (ESL) students	69.9	24.8
Special education students	40.3	20.2
Students not requiring testing accommodations (non-ESL and non-special education)	74.1	40.4
Sample size (school records sample)		
Students (total = 16,465)	12,757	3,708
Schools (total = 28)	23	5

Note: This table is based on the 28 schools that participated in both years of the study. The grade 9 sample includes students who are in grade 9 for the first time; the grade 10 sample includes grade 10 students and retained grade 9 students.

Source: Calculations based on school records and reading assessment data collected for the Content Literacy Continuum study. School records were obtained for students enrolled in the study schools at the end of each school year. The GRADE reading assessment was administered in the spring of each study year.

Characteristics of the analysis samples by treatment group

A comparison of the background characteristics and prior achievement of students in the school records sample, by program group (CLC schools and non-CLC schools), for each study year and grade level, appears in tables C.9–C.11. The characteristics of students in the GRADE respondent sample are shown in similar tables (tables C.12–C.14).

As seen in these tables, there is no systematic difference between students in CLC schools and non-CLC schools with respect to their background characteristics and prior achievement, for either of the two analysis samples (school records sample or GRADE respondent sample) in either grade level in year 2 of the study. There is a systematic difference between the two groups in year 1: the CLC group was composed of 6.7 percent Hispanic students as compared to 3.5 percent in the non-CLC group. This difference is significant at the $p<.01$ level.

In general, this suggests that the student analysis samples preserve the balance achieved with random assignment, and the difference in student outcomes between CLC and non-CLC schools reflects the impact of the CLC framework rather than preexisting differences in students' background characteristics or prior achievement.

Table C.6. Background characteristics of students who did and did not take the GRADE assessment in year 1 of the study, school records sample (grade 9)

Characteristic	Students with a GRADE score	Students without a GRADE score	Estimated difference	*p*-value
Average age (years)	14.6	14.9	−0.2	0.000*
Overage for grade (%)[a]	22.0	39.4	−17.3	0.000*
Free/reduced-price lunch eligible (%)	64.9	71.7	−6.9	0.000*
Race/ethnicity (%)				
Hispanic	5.0	5.6	−0.6	0.286
Black, non-Hispanic	42.6	44.9	−2.3	0.004*
White, non-Hispanic	48.3	44.1	4.2	0.000*
Other	4.1	5.5	−1.4	0.018*
Gender (% male)	49.5	56.8	−7.3	0.000*
English language learner (%)	11.0	13.1	−2.1	0.003*
Special education (%)	10.7	34.6	−23.9	0.000*
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	68.3	44.6	23.7	0.000*
Mathematics	61.2	42.1	19.1	0.000*
Test of systematic difference between groups[b] ($\chi 2 = 622.9$)				0.000*
Sample size[c]				
Students (total = 7,365)	4,304	3,061		
Schools	28	28		

* *p*-value ≤ .05.

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 23.2 percent (math) and 24.0 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators; the percentage ranges from 0.0 percent (age) to 1.6 percent (English language learner status) for grade 9 students in the first year.

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table C.7. Background characteristics of students who did and did not take the GRADE assessment in year 2 of the study, school records sample (grade 9)

Characteristic	Students with a GRADE score	Students without a GRADE score	Estimated difference	*p*-value
Average age (years)	14.7	14.9	−0.3	0.000*
Overage for grade (%)[a]	24.6	39.1	−14.5	0.000*
Free/reduced-price lunch eligible (%)	66.2	68.7	−2.6	0.016*
Race/ethnicity (%)				
Hispanic	4.5	6.1	−1.6	0.003*
Black, non-Hispanic	51.7	51.7	0.0	0.954
White, non-Hispanic	38.9	36.3	2.6	0.002*
Other	4.9	6.0	−1.1	0.053
Gender (% male)	49.8	56.6	−6.8	0.000*
English language learner (%)	7.9	12.6	−4.6	0.000*
Special education (%)	11.0	32.4	−21.5	0.000*
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	62.3	41.3	21.0	0.000*
Mathematics	60.2	39.8	20.4	0.000*
Test of systematic difference between groups[b] ($\chi 2 = 638.6$)				0.000*
Sample size[c]				
Students (total = 7,951)	4,956	2,995		
Schools	28	28		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 31.0 percent (math) and 31.7 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.4 percent (age) to 2.2 percent (special education status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table C.8. Background characteristics of students who did and did not take the GRADE assessment in year 2 of the study, school records sample (grade 10)

Characteristic	Students with a GRADE score	Students without a GRADE score	Estimated difference	*p*-value
Average age (years)	14.6	15.1	−0.4	0.000*
Overage for grade (%)[a]	23.1	47.6	−24.5	0.000*
Free/reduced-price lunch eligible (%)	64.1	71.6	−7.4	0.000*
Race/ethnicity (%)				
Hispanic	3.8	6.2	−2.4	0.000*
Black, non-Hispanic	45.5	47.4	−1.9	0.009*
White, non-Hispanic	46.7	41.2	5.5	0.000*
Other	4.0	5.3	−1.3	0.015*
Gender (% male)	49.5	57.8	−8.2	0.000*
English language learner (%)	9.6	14.2	−4.6	0.000*
Special education (%)	12.0	29.7	−17.7	0.000*
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	69.2	47.3	22.0	0.000*
Mathematics	62.4	41.0	21.4	0.000*
Test of systematic difference between groups[b] ($\chi 2 = 522.3$)				0.000*
Sample size[c]				
Students (total = 8,514)	4,430	4,084		
Schools	28	28		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are based on grade 10 students and retained grade 9 students. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 50.3 percent (math) and 51.5 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.7 percent (age) to 23.2 percent (English language learner status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table C.9. Background characteristics of students in year 1 of the study, school records sample (grade 9)

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Average age (years)	14.8	14.8	0.0	0.880
Overage for grade (%)[a]	30.7	31.6	−0.8	0.527
Free/reduced-price lunch eligible (%)	69.1	69.5	−0.4	0.870
Race/ethnicity (%)				
Hispanic	6.9	3.4	3.4	0.000*
Black, non-Hispanic	52.5	52.7	−0.2	0.964
White, non-Hispanic	35.9	39.2	−3.3	0.378
Other	4.7	4.7	0.1	0.976
Gender (% male)	51.4	52.9	−1.5	0.301
English language learner (%)	8.5	5.7	2.8	0.227
Special education (%)	18.3	22.6	−4.2	0.050*
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	58.3	55.2	3.1	0.322
Mathematics	50.8	48.9	1.9	0.615
Test of systematic difference between groups[b]	($\chi 2 = 13.2$)			0.212
Sample size[c]				
Students (total = 7,365)	4,254	3,111		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 23.2 percent (math) and 24.0 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.0 percent (age) to 1.6 percent (English language learner status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table C.10. Background characteristics of students in year 2 of the study, school records sample (grade 9)

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Average age (years)	14.8	14.8	0.0	0.792
Overage for grade (%)[a]	32.5	31.6	0.9	0.782
Free/reduced-price lunch eligible (%)	68.9	66.2	2.7	0.182
Race/ethnicity (%)				
Hispanic	6.5	5.3	1.2	0.043*
Black, non-Hispanic	52.0	53.4	−1.3	0.752
White, non-Hispanic	34.9	36.2	−1.3	0.664
Other	6.6	5.2	1.4	0.496
Gender (% male)	52.8	49.8	3.0	0.092
English language learner (%)	8.1	7.1	1.0	0.657
Special education (%)	18.3	19.7	−1.4	0.384
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	55.3	53.1	2.2	0.430
Mathematics	51.1	50.7	0.5	0.866
Test of systematic difference between groups[b]	($\chi 2 = 16.7$)			0.082
Sample size[c]				
Students (total = 7,951)	4,467	3,484		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 31.0 percent (math) and 31.7 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.4 percent (age) to 2.2 percent (special education status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table C.11. Background characteristics of students in year 2 of the study, school records sample (grade 10)

Characteristic	CLC group	Non-CLC group	Estimated Difference	*p*-value
Average age (years)	14.9	14.9	0.0	0.542
Overage for grade (%)[a]	35.4	37.3	−1.9	0.478
Free/reduced-price lunch eligible (%)	69.2	68.9	0.3	0.906
Race/ethnicity (%)				
Hispanic	7.1	3.9	3.2	0.000*
Black, non-Hispanic	51.8	52.9	−1.1	0.817
White, non-Hispanic	36.3	38.5	−2.3	0.574
Other	4.9	4.7	0.2	0.945
Gender (% male)	52.1	54.3	−2.2	0.289
English language learner (%)	9.1	6.1	2.9	0.244
Special education (%)	18.5	23.4	−4.9	0.026*
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	58.6	54.6	4.1	0.151
Mathematics	51.1	48.8	2.3	0.505
Test of systematic difference between groups[b]	($\chi 2 = 15.4$)			0.119
Sample size[c]				
Students (total = 8,514)	4,888	3,626		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the table are for grade 10 students and retained grade 9 students. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 50.3 percent (math) and 51.5 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.7 percent (age) to 23.2 percent (English language learner status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table C.12. Background characteristics of students in year 1 of the study, GRADE respondent sample (grade 9)

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Average age (years)	14.7	14.7	0.0	0.316
Overage for grade (%)[a]	25.8	26.5	−0.7	0.580
Free/reduced-price lunch eligible (%)	68.0	68.1	−0.1	0.970
Race/ethnicity (%)				
Hispanic	6.7	3.5	3.2	0.002*
Black, non-Hispanic	52.2	52.8	−0.6	0.889
White, non-Hispanic	37.1	39.1	−2.0	0.606
Other	4.0	4.7	−0.7	0.475
Gender (% male)	49.3	50.5	−1.2	0.570
English language learner (%)	7.7	6.0	1.7	0.173
Special education (%)	12.0	12.1	−0.1	0.966
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	64.3	61.7	2.5	0.478
Mathematics	54.2	55.2	−1.0	0.795
Test of systematic difference between groups[b] ($\chi 2 = 18.3$)				0.050*
Sample size[c]				
Students (total = 4,786)	2,869	1,917		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 21.9 percent (math) and 22.1 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.9 percent (age) to 2.0 percent (free lunch status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table C.13. Background characteristics of students in year 2 of the study, GRADE respondent sample (grade 9)

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Average age (years)	14.7	14.7	0.0	0.997
Overage for grade (%)[a]	27.7	27.1	0.6	0.801
Free/reduced-price lunch eligible (%)	72.2	68.7	3.5	0.162
Race/ethnicity (%)				
Hispanic	3.7	2.9	0.9	0.220
Black, non-Hispanic	58.8	61.8	−3.0	0.337
White, non-Hispanic	32.2	31.1	1.1	0.665
Other	5.3	5.3	0.0	1.000
Gender (% male)	49.8	47.4	2.4	0.317
English language learner (%)	4.4	2.8	1.6	0.419
Special education (%)	9.7	11.9	−2.1	0.416
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	59.2	55.4	3.7	0.328
Mathematics	55.3	52.6	2.7	0.462
Test of systematic difference between groups[b] ($\chi 2 = 15.3$)				0.121
Sample size[c]				
Students (total = 5,011)	2,975	2,036		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 26.4 percent (math) and 26.7 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.1 percent (age) to 2.6 percent (special education status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table C.14. Background characteristics of students in year 2 of the study, GRADE respondent sample (grade 10)

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Average age (years)	14.7	14.7	0.0	0.182
Overage for grade (%)[a]	25.9	28.0	−2.1	0.343
Free/reduced-price lunch eligible (%)	71.2	72.5	−1.4	0.600
Race/ethnicity (%)				
Hispanic	4.7	2.9	1.7	0.225
Black, non-Hispanic	59.5	59.5	0.0	0.999
White, non-Hispanic	33.0	33.9	−0.9	0.832
Other	2.8	3.7	−0.9	0.335
Gender (% male)	49.0	50.3	−1.3	0.605
English language learner (%)	6.0	3.8	2.2	0.365
Special education (%)	11.2	14.5	−3.3	0.144
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	63.9	61.3	2.5	0.561
Mathematics	54.4	54.3	0.1	0.974
Test of systematic difference between groups[b] ($\chi 2 = 14.0$)				0.175
Sample size[c]				
Students (total = 4,546)	2,908	1,638		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the table are for grade 10 students and retained grade 9 students. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 35.5 percent (math) and 35.8 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.1 percent (age) to 15.1 percent (English language learner status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Appendix D. Characteristics of the Sampled and Observed Classrooms

As described in chapter 2, grade 9 and grade 10 classrooms in the study schools were randomly selected for observation as part of the implementation research for this study.[163] Observations of core content classrooms then were used to derive measures of instructional practice and to estimate the service contrast in instruction between Content Literacy Continuum (CLC) and non-CLC schools (see chapter 4).

This appendix supplements chapter 2 by presenting additional information on the characteristics (subject and grade distribution) of *sampled* and *observed* core content classrooms. The first section compares the characteristics of sampled classrooms in CLC schools and non-CLC schools to ascertain whether random assignment resulted in two comparable groups of classes. The second section compares the characteristics of sampled classrooms that were observed to those of sampled classrooms that were *not* observed to better understand the external validity of the instructional contrast findings. As in other parts of the report, the analyses presented in this appendix are based on the 28 schools that participated in both years of the study. Year 1 observations were conducted in more than 89 percent of the schools; in year 2, observations were conducted in all schools.

The tables compare the characteristics of two groups of classrooms: (1) classes in CLC schools and non-CLC schools and (2) classes observed and not observed. A t-test is used to test whether the groups differ with respect to the percentage of classrooms in each subject area and grade level. However, because many hypothesis tests are conducted in these tables (one for each characteristic), there is an increased probability of concluding that a particular difference is statistically significant when in fact it is not (a type I error or a false positive).[164] For this reason, an omnibus test is used to test for a *systematic* or overall difference between the characteristics of classes in different groups. This test is reported at the bottom of each table in this appendix. An omnibus test that is not statistically significant means that a statistically significant difference for any given characteristic in the tables may be due to chance.

Characteristics of sampled classrooms

The distribution of sampled classrooms across subject areas and grade levels in CLC and non-CLC schools is shown in table D.1.[165] As seen in this table, there is no systematic difference between sampled classrooms in CLC schools and non-CLC schools with respect to subject area or grade level. This is true for both year 1 and year 2.

Recall from chapter 2, however, that in year 2, a higher percentage of sampled classrooms were ultimately observed in CLC schools compared to non-CLC schools. This means that among classrooms that were observed (the instructional sample), the balance achieved by random

[163] See appendix A for a description of the sampling process.

[164] In particular, one would expect to see one false positive for every 20 hypothesis tests conducted.

[165] These distributions are weighted to account for different sampling probabilities across subject areas.

assignment may be compromised. This in turn limits the study's ability to infer that the difference in instructional practice between CLC and non-CLC schools represents the effect of

Table D.1. Characteristics of sampled core content classrooms, by program group

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Year 1 (grade 9)				
Subject area (%)				
English language arts	24.7	22.2	2.4	0.646
Social studies	23.9	24.4	−0.5	0.921
Science	24.3	25.3	−0.9	0.863
Mathematics	27.1	28.1	−1.0	0.860
Test of systematic difference between groups ($\chi 2 = 4.97$)				0.174
Sample size				
Classrooms (total = 299)	164	135		
Schools (total = 27)	15	12		
Year 2 (grades 9 and 10)				
Subject area (%)				
English language arts	23.3	21.0	2.3	0.651
Social studies	23.2	24.9	−1.6	0.756
Science	25.3	26.4	−1.0	0.846
Mathematics	28.2	27.8	0.4	0.949
Grade level (%)				
grade 9	54.1	62.7	−8.6	0.219
grade 10	45.9	37.3	8.6	0.219
Test of systematic difference between groups[a] ($\chi 2 = 2.8$)				0.591
Sample size				
Classrooms (total = 312)	168	144		
Schools (total = 28)	15	13		

**p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. In year 1, observations were conducted in 27 schools; in year 2, observations were conducted in all schools. Observations are weighted to account for different sampling probabilities across subject areas. Rounding may cause slight discrepancies in calculating sums and differences.

a. An omnibus chi-squared test was used to determine whether there is a systematic difference between schools in the CLC group and the non-CLC group, with respect to the characteristics included in this table.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

CLC on instruction because differences in instructional practice may be due to preexisting differences in teachers' characteristics and ability.

Characteristics of observed and not observed classrooms

As noted previously and explained in chapter 2, not all classrooms sampled for observation were observed during the site visits. Therefore it is important to examine whether the characteristics of classrooms that were observed differ from those of classrooms that were not observed. This affects the extent to which the instructional contrast findings in chapter 4 are generalizable to classrooms that could not be included in the analysis.

The characteristics of core content classrooms in the instructional sample (that is, sampled classes that were observed) are compared to the characteristics of sampled classrooms that were not observed in year 1 and year 2 in tables D.2 and D.3.[166] As seen in these tables, there is a systematic difference between observed and unobserved classrooms in year 2. This indicates that the year 2 instructional contrast findings may not be generalizable to classrooms that were not observed.

Table D.2. Characteristics of observed and not observed core content classrooms, year 1 of the study (grade 9)

Characteristic	Observed	Not observed	Estimated difference	*p*-value
Subject area (%)				
English language arts	22.8	24.7	−1.9	0.727
Social studies	24.6	22.8	1.9	0.731
Science	26.4	21.2	5.2	0.348
Mathematics	26.2	31.3	−5.2	0.365
Test of systematic difference between groups[a] ($\chi 2 = 1.1$)				0.767
Sample size				
Classrooms (total = 299)	213	86		
Schools (total =27)	27	27		

**p*-value ≤ .05

Note: This table is based on the 27 schools where classroom observations were conducted (of the 28 schools that participated in both years of the study). Observations are weighted to account for different sampling probabilities across subject areas.

a. An omnibus chi-squared test was used to determine whether there is a systematic difference between classrooms that were observed and not observed, with respect to the characteristics included in this table.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

[166] Observations are weighted to account for different sampling probabilities across subject areas.

Table D.3. Characteristics of observed and not observed core content classrooms, year 2 of the study (grades 9 and 10)

Characteristic	Observed	Not observed	Estimated difference	*p*-value
Subject area (%)				
English language arts	23.3	4.5	18.7	0.063
Social studies	23.6	27.8	−4.2	0.688
Science	25.4	34.4	−9.0	0.399
Mathematics	27.7	33.3	−5.6	0.608
Grade level (%)				
Grade 9	56.5	78.5	−22.0	0.069
Grade 10	43.5	21.5	22.0	0.069
Test of systematic difference between groups[a]		($\chi 2 = 9.8$)		0.045*
Sample Size				
Classrooms (total = 312)	295	17		
Schools (total = 28)	28	28		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Observations are weighted to account for different sampling probabilities across subject areas.

a. An omnibus chi-squared test was used to determine whether there is a systematic difference between classrooms that were observed and not observed, with respect to the characteristics included in this table.

Source: Classroom observations conducted for the Content Literacy Continuum study in the spring of each study year.

Appendix E. Technical Notes for the Impact Analysis

This appendix describes the statistical models used to estimate impacts on student and classroom outcomes in this report. The first section describes the model used to estimate the effect of the Content Literacy Continuum (CLC) on students' school records outcomes, which include credit accumulation (chapters 5 and 6), grade point average (GPA) (chapter 6), and attendance and course-taking patterns (appendix O).

The second section discusses the model used to estimate impacts on students' reading achievement (chapter 5 and 6), which is slightly different because in some districts students were sampled for GRADE testing. The third section discusses the model used to estimate the instructional service contrast presented in chapter 4.

Impact on course performance outcomes

Impacts on school records outcomes (credits earned, grade point average, and attendance) in this report are estimated by fitting a two-level model to the school records sample, separately for each study year and grade level.

Level 1: students-within-schools. Level 1 describes the relationship between students' course performance outcomes and their background characteristics:

$$Y_{ij} = \pi_j + \sum_{s,k} \phi_{sk} X_{sij} D_k + \sum_s \omega_s M_{sij} + e_{ij} \quad \text{(1a)}$$
$$YY$$

where

Y_{ij} = School records outcome of student i in school j (credit accumulation, GPA, or attendance).

X_{ij} = A set of S student-level characteristics for student i in school j measured prior to students' first exposure to the CLC framework. These covariates reduce within-school and between-school variation in the outcome measure, thereby increasing the precision of the impact estimates. The effect of each student characteristic is allowed to vary across random assignment blocks (D_k, defined below).[167]

[167] As explained in chapter 2, the following covariates are included in the statistical model to adjust for random differences between the characteristics of students in CLC and non-CLC schools: whether students were overage at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, English as a second language (ESL) status, special education status, racial/ethnic group, and gender. The decision about which covariates to include in the model was made based on the expectation that they would be important predictors of academic outcomes. Prior to starting the impact analysis, a model was estimated in which student outcomes were regressed against the set of student background covariates; this analysis confirmed that all characteristics are statistically significant predictors of the student outcomes in this report. See appendix K for information on the explanatory power of these covariates. Note that the covariates are not grand-mean centered; grand-mean centering (or not) would not have affected the estimated treatment effect, which is the parameter of interest.

M_{ij} = A set of S missing indicators for each of the student-level characteristic, coded 1 if missing and 0 otherwise.[168]

e_{ij} = A random error term for student i from school j assumed to be independently and identically distributed across students within schools (that is, the within-school residual).

Therefore,

π_j = Average of outcome Y of students at school j.

Level 2: schools. Given that random assignment occurs at the school level, CLC impacts are estimated at the school level. Thus, level 2 examines the difference between the school-level adjusted outcomes—π_j—of students in CLC and non-CLC schools, controlling for random assignment blocks:

$$\pi_j = \sum_K \delta_k D_k + \sum_K \beta_k T_j * D_k + \mu_j \quad \text{(1b)}$$

where

T_j = Indicator of CLC group membership. This indicator equals 1 if school j was randomly assigned to implement the CLC framework and 0 otherwise.

D_k = Random assignment block indicators, which equal 1 if student i is in random assignment block k and 0 otherwise.[169] These blocks are included in the model to capture a central feature of the research design in which random assignment was conducted in blocks. These blocks also account for all variation in average student outcomes across blocks; thus the only sources of variation in Y in this model are among schools within a random assignment block and among students within schools.

μ_j = A random error term for school j assumed to be independently and identically distributed across schools (that is, the between-school residual).[170]

---

[168] Missing information on the student characteristics X was imputed using a dummy variable approach. This approach consists of (1) imputing a value of "zero" for the missing values in each of the covariates, (2) creating a dichotomous indicator of missingness for each covariate, and (3) including these indicators alongside the imputed covariates in the statistical model (Puma et al., 2009). The percentage of missing data is highest for grade 8 test scores: 23.2 percent (math) and 24.0 percent (reading) for grade 9 students in the first year; 31.0 percent (math) and 31.7 percent (reading) for grade 9 students in the second year; and 50.3 percent (math) and 51.5 percent (reading) for grade 10 students in the second year. The percentage of missing data is lowest for demographic characteristics and educational indicators; the percentage ranges from 0.0 percent (age) to 1.6 percent (English language learner status) for grade 9 students in the first year; from 0.4 percent (age) to 2.2 percent (special education status) for grade 9 students in the second year; and from 0.7 percent (age) to 23.2 percent (English language learner status) for grade 10 students in the second year. There is more missing data for grade 10 students because historical school records information is not available for students who are new to the school in grade 10.

[169] There are 9 random assignment blocks for the 28 schools that participated in both year 1 and year 2; there are 10 blocks for the 33 schools that participated in year 1. See chapter 2 for a discussion of blocking.

[170] School-level characteristics were not included at this level because (1) there are too few remaining degrees of freedom, (2) the block indicators explain much of the variation in Y between schools, and (3) the student-level covariates (X) also explain some of the variation between schools.

Therefore,

β_k = The difference between the school-level average of outcome *Y* in the CLC schools and the non-CLC schools in block *k* (that is, the impact of the CLC framework on outcome *Y* in block *k*).

In practice, this two-level model is estimated by substituting equation 1b into equation 1a and then fitting equation 1*:

$$Y_{ij} = \sum_K \delta_k D_k + \sum_K \beta_k T_j D_k + \sum_{S,K} \phi_{sk} X_{sij} D_k + \sum_s \omega_s M_{sij} + u_j + e_{ij} \quad (1^*)$$
$$YY$$

The average impact of the CLC framework across school districts—$\overline{\beta}$—is then obtained by weighting the block-level impacts—β_k—by the number of CLC schools in the block. Thus $\overline{\beta}$ is a fixed-effects estimate of the impact of the CLC framework for the average student in the average CLC school in the study sample. For this reason the average estimate cannot be used to make statistical inferences about the impact of CLC in some larger population of schools. This "fixed effects" approach to obtaining a pooled impact estimate is used because the school districts in the study were selected purposefully and are not a random sample of districts from a larger target population.[171]

Impact on reading achievement

For the GRADE reading assessment, some districts chose to test a random sample of classrooms rather than all students.[172] In these districts, the analysis must account for not only cluster random assignment (at the school level) but also cluster sampling (at the classroom level). Thus, in this report, impacts on the GRADE are estimated using a hybrid approach:

- In districts where students were not sampled for GRADE testing, the impact of CLC is estimated by fitting a *two-level* model (equation 1*) to the GRADE respondent sample.
- In districts where students were sampled, a *three-level* model is used to estimate the impact of CLC, where students are nested within classrooms, which are nested within schools (see equation 2* below). Observations in these districts also are weighted to account for different sampling probabilities across classrooms.
- To obtain the overall estimated impact of CLC—$\overline{\beta}$—estimated impacts for each block (β_k) then were pooled by weighting each block by the number of CLC schools in the block.

[171] Another option for pooling the results would be to use a random-effects approach (with block random effects). This approach attempts to estimate the impact of the program for the broader population of sites represented by the study sample (as opposed to the fixed-effects approach, which restricts its inferences to the sites in the study sample). To date, given the typically small number of sites (districts) for most social experiments, it has been common practice to use fixed-effect models for pooling experimental findings. In this study, for example, the small number of districts/blocks does not provide enough information about how true impacts vary across districts to support generalizations with adequate precision.

[172] Classroom sampling was used in four schools in year 1 and five schools in year 2; see appendix A for details.

The three-level model used in the districts where cluster sampling was implemented is as follows:

Level 1: students-within-classrooms. As before, level 1 describes the relationship between students' course performance outcomes and their background characteristics:

$$Y_{igj} = \alpha_{gj} + \sum_{s,k} \phi_{sk} X_{sigj} D_k + \sum_s \omega_s M_{sigj} + e_{igj} \quad \text{(A)}$$

where

Y_{igj} = Reading achievement outcome of student *i* in classroom *g* in school *j* (reading comprehension or reading vocabulary).

X_{igj} = A set of *S* student-level characteristics for student *i* in classroom *g* in school *j* measured prior to students' first exposure to the CLC framework (same characteristics as in equation 1*).[173]

e_{igj} = A random error term for student *i* in classroom *g* in school *j* assumed to be independently and identically distributed across students within classrooms (that is, the within-classroom residual).

Therefore,

α_{gj} = Average outcome for students in classroom *g* at school *j*.

Level 2: classrooms-within-schools. Level 2 describes the average outcomes of students at a given school:

$$\alpha_{gj} = \lambda_j + \nu_{gj} \quad \text{(B)}$$

where

ν_{gj} = A random-error term for classroom *g* in school *j* assumed to be independently and identically distributed across classrooms within schools (that is, the within-school residual).

Therefore,

λ_j = Average outcomes of classrooms in school *j*.

Level 3: schools. Level 3 examines the difference between school-level adjusted outcomes—λ_j—of students in CLC and non-CLC schools, controlling for random assignment blocks:

[173] In this case, these covariates explain within-classroom, between-classroom, and between-school variation in the outcome measure.

$$\lambda_j = \sum_K \delta_k D_k + \sum_K \beta_k T_j * D_k + \mu_j \quad \text{(C)}$$

where

T_j = Indicator of CLC group membership.

D_k = Random assignment block indicators, which equal 1 if student *i* is in random assignment block *k* and 0 otherwise.[174]

μ_j = A random error term for school *j* assumed to be independently and identically distributed across schools (that is, the between-school residual).

Therefore,

β_k = The difference between the school-level average of outcome *Y* in the CLC schools and the non-CLC schools in block *k* (that is, the impact of the CLC framework on outcome *Y* in block *k*).

In practice, this three-level model is estimated by substituting equation C and equation B into equation A and then fitting equation 2*:

$$Y_{igj} = \sum_K \delta_k D_k + \sum_K \beta_k T_j D_k + \sum_{S,K} \phi_{sk} X_{sigj} D_k + \sum_s \omega_s M_{sigj} + u_j + v_{gj} + e_{ij} \quad \text{(2*)}$$

Instructional contrast

In chapter 4, the instructional contrast is estimated using a three-level model fitted to classrooms in the instructional sample. The statistical model is similar to equation 2*, with the following exceptions:

- The three levels in the analysis are (1) classrooms nested within teachers, (2) teachers nested within schools, and (3) schools.
- The level 1 model does not include covariates (*X*).[175]
- Observations are weighted to account for stratified sampling by subject area (that is, some subject areas were oversampled).

[174] GRADE classroom sampling was conducted in two random assignment blocks in year 1 and in three random assignment blocks in year 2 (see chapter 2 and appendix A).

[175] Prior to estimating the impact model, an investigation was conducted to determine whether the precision of the instructional contrast estimates might be improved by controlling for the subject area of the classroom in the model (that is, ELA, science, social studies, and mathematics). This was done by including a set of indicators for subject area in equation 2* but omitting the treatment indicator. It was found that these indicators were not statistically significant predictors of the instructional outcomes, which means that the statistical power of the analysis would not be improved by including them in the model.

Appendix F. Statistical Power and Minimum Detectable Effect Size

This appendix supplements the discussion of statistical power in chapter 2. First, it describes the calculation of the minimum detectable effect size (MDES) for impacts on student outcomes. It also presents the MDES for the instructional contrast estimates presented in chapter 4.

The discussion refers to two related concepts that often are used to convey statistical power: the minimum detectable effect (MDE) and the MDES. The MDE is the smallest true program impact that can be detected, given random sampling and estimation error.[176] The MDES is the MDE scaled as an effect size; in other words, it is the MDE divided by the standard deviation of the outcome of interest. Effect sizes are used widely for measuring the impacts of educational programs and are defined in terms of the underlying population standard deviation of student achievement. For example, an MDES of 0.20 indicates that an impact estimator can reliably detect a program-induced increase in student achievement that is equal to or greater than the 0.20 standard deviation of the existing student distribution.

Impact on student outcomes

The MDE and MDES for impacts on student outcomes are calculated as follows:

$$MDE = M_{J-2B} * s.e.(\bar{\beta}) \quad (1a)$$

$$MDES = M_{J-2B} * \frac{s.e.(\bar{\beta})}{\sigma} \quad (1b)$$

where

$s.e.(\bar{\beta})$ = The standard error of the impact estimate.

σ = The standard deviation used to calculate effect sizes (in this case, the standard deviation of the outcome among students in the non-CLC group of schools; see appendix L for these standard deviations).

B = The number of random assignment blocks (10).

J = The number of schools in the study (28 schools that participated in both year 1 and year 2).

[176] The MDE is defined as the smallest true program impact that would have an 80 percent chance of being detected (have 80 percent power) using a two-tail hypothesis test at the 5 percent level of statistical significance.

M_{J-B} = The degrees of freedom multiplier, which is calculated to be 3.1 in this study, assuming a two-tailed test with a statistical power level of 0.80 and a statistical significance level of 0.05.[177]

The MDE and the MDES for impacts on student outcomes (reading achievement, credit accumulation, and GPA) are presented in chapter 2. As discussed in that chapter, the MDES is larger for some grade levels and study years than others because the standard error of their impact estimates is larger. (See appendix J for the standard errors used to calculate the MDE for the primary outcomes.) In turn three factors affect the size of the standard error: the intraclass correlation or ICC, the between-school R^2, and the within-school R^2.[178] All else equal, the MDES decreases when the ICC is smaller and when the two R^2s are larger. With a large ICC, a high between-school R^2 is especially important for detecting impacts of reasonable magnitude.[179] Appendix K presents the values of these model fit parameters for the key student outcomes in this study.

[177] There are 8 degrees of freedom for estimating impacts on student outcomes (28 schools, 10 block indicators, 10 treatment indicators). See Bloom (2005) for details on how the multiplier is calculated in a group randomized experiment.
[178] The ICC is the proportion of the total variation in the student outcome that is between schools (as opposed to within schools). The between-school R^2 is the proportion of between-school variation that is explained by the student-level covariates included in the model (students' background characteristics and prior achievement) and the school-level covariates in the model (random assignment block indicators); see appendix E for a description of these covariates. The within-school R^2 is the proportion of within-school variation that is explained by the student-level covariates.
[179] See Bloom, Richburg-Hayes, and Black (2007) for a discussion.

Appendix G. Description of Frequently Introduced Content Enhancement Routines and Learning Strategies

This appendix presents an overview of the most commonly introduced Content Enhancement Routines (level 1) and Learning Strategies (level 2) by the Content Literacy Continuum (CLC) schools.[180] The descriptions provide context for the routines and strategies presented in chapter 3.

Content enhancement routines

The *Framing Routine* is used to transform abstract main ideas and key topics into a concrete representation that helps students think and talk about the key topic and essential related information.

The *Unit Organizer Routine* is used to plan units and then introduce and maintain the big ideas in units and show how units, critical information and concepts are related.

The *Concept Mastery Routine* is used to define, summarize, and explain a major concept and where it fits within a larger body of knowledge.

The *Course Organizer Routine* is used to plan courses around essential learning and critical concepts. The teacher uses the routine to introduce the course and the rituals that will be used throughout the course. The teacher then uses this framework throughout the year to maintain the big ideas and rituals.

Learning strategies

The *7-Step Vocabulary Strategy* is a strategy for helping students understand new vocabulary by breaking down a word to deduce meaning in seven steps.

The *Paragraph Writing Strate*gy is a strategy for organizing ideas related to a topic; planning the point of view and verb tense to be used in the paragraph; planning the sequence in which ideas will be expressed; and writing a variety of topic, detail, and clincher sentences.

The *Sentence Writing Strategy* program is composed of two parts: *Fundamentals in the Sentence Writing Strategy* and *Proficiency in the Sentence Writing Strategy.* Together these components constitute a strategy for recognizing and writing 14 sentence patterns with four types of sentences: simple, compound, complex, and compound-complex.

[180] Descriptions of the Content Enhancement Routines and Learning Strategies are stated as found on The University of Kansas Center for Research on Learning (KU-CRL) website: http://kucrl.org

Appendix H. Explanation of Shared Pedagogical Practices Obtained through Classroom Observations Using the ACE Protocol

As described in the description of the Content Literacy Continuum (CLC) framework in chapter 2, the three pedagogical practices emphasized in levels 1, 2, and 3 of CLC professional development are sequencing of instruction (Cue-Do-Review), instructing through multiple modalities (for example, written, verbal, or visually/graphically), and using interactive discourse and scaffolding with students. Observation items, constructs, and actions were written into the ACE protocol so observers could identify when components of sequenced instruction, multiple instructional modalities, and interactive instruction were present. The scores for the three CLC-emphasized pedagogical practices were constructed post hoc by creating composites across protocol items and segments. A complete listing of both shared pedagogical measures and the manner in which shared pedagogy was later classified is in table H.1.

Sequenced instruction

This CLC-emphasized pedagogical practice was designed to measure the extent to which teachers implement a deliberate instructional sequence (such as Cue-Do-Review) within their teaching. For instance, within the Content Enhancement Routine Cue phase, the teacher employs a routine and explains the purpose of the routine and how students are to participate in the routine. In the Do phase, teachers use at least one action related to a Content Enhancement Routine or Learning Strategy. In the Review phase, teachers ensure that given content has been learned and that a Content Enhancement Routine and/or process of learning about the topic or routine has been carried out. For ideal implementation, these phases would be carried out in sequence. The presence of each phase within a classroom observation is coded 1 (0 for absence). If these phases occur in sequence, the observation of that classroom also is coded 1 (0 otherwise). The codes for these four components are added together to create an aggregate score for this CLC-emphasized pedagogical practice (that is, scores can range from 0 to 4). Reliability (coefficient alpha) for sequenced instruction was calculated to be .65.

Multiple instructional modalities

Across levels 1, 2, and 3 of CLC professional development sessions, site coordinators emphasize the importance of instructing through multiple modalities. For instance, teachers are told to specifically name the Content Enhancement Routines or Learning Strategies they are using. Teachers are encouraged to use graphic organizers along with instruction. They also are expected to use a blend of oral and written discourse. The use of CLC-emphasized pedagogical practices is especially apparent when teachers incorporate all three components into their instruction (a named Content Enhancement Routine or Learning Strategy, a graphic organizer, and a blend of oral and written discourse).

The presence or absence of each component was recorded by observers. The presence of each component within a classroom period was coded 1; absence of the component was coded 0. A fourth component signifies the presence of the three other components within a class period (coded 1 if present or 0 if absent). Aggregate scores for multiple instructional modalities represented the sum across the four components, thereby creating scores for an observed classroom period that can range from 0 to 4. Reliability for this four-item pedagogical construct was .71.

Interactive and scaffolded instruction

Levels 1, 2, and 3 of CLC professional development also emphasize teacher-student co-construction of knowledge through interactive discourse and transitioning instruction from teacher-mediated to student-mediated activity.

The presence of interactive discourse at any time during a class period was recorded by observers and coded 1; the absence of interactive discourse was coded 0. The presence of interactive discourse in every observed segment within a class period was later coded 1 (0 if interactive discourse was not present in every observational segment). Likewise the presence of scaffolded instruction within any observation segment within a class period was coded 1 for presence and 0 for absence. Scaffolded instruction present in all segments within a class period was also coded 1 or 0 (signifying presence or absence, respectively). An aggregated score for interactive and scaffolded instruction was created by adding together these four components. Reliability for this four-item pedagogy measure was .69.

Table H.1. Constructing instructional shared pedagogical measures with items from the ACE protocol

	Component of CLC pedagogical practice	Coding of pedagogical practice	Alignment with ACE protocol
Sequenced instruction	**1. Teachers initiate an instructional sequence by establishing its purpose (that is, the Cue phase from the Cue-Do-Review sequence).** During this phase, teacher names a Content Enhancement Routine or Learning Strategy to be used, explains how it will help students learn, and specifies what students need to do to participate in routine.	Coded 1 if at least one action is checked under **setting purpose** within a single class period.	**Setting purpose** construct
	2. Teachers perform at least one action aligned with Content Enhancement Routines or Learning Strategies. This reflects the Do phase from Cue-Do-Review.	Coded 1 if at least one action is checked under any of the following **CLC-aligned constructs** during a single class period: • Building background knowledge • Predicting content • Highlighting essential vocabulary • Determining importance • Questioning/clarifying • Summarizing/synthesizing	The following **constructs aligned with CLC routines/strategies** (*indicates Content Enhancement Routines and Learning Strategies taught in at least one-third of the CLC schools during the 2008/09 school year, according to the KU-CRL): • Building background knowledge (Concept Anchoring, Concept Mastery*, Concept Comparison) • Predicting content (Course/Unit/Lesson Organizer*, Prediction Strategy) • Highlighting essential vocabulary (LINCS Vocabulary Routine, 7-Step Vocabulary Process, Bridging Strategy) • Determining importance (FRAME*, Concept Anchoring, Concept Mastery*, Concept Comparison) • Questioning/clarifying (Question Exploration, Clarifying Routine, Bridging Strategy, Word Mapping Strategy) • Summarizing/synthesizing (FRAME*, Concept Anchoring, Concept Mastery*, Concept Comparison, Summarization

			Strategy)
	Component of CLC pedagogical practice	**Coding of pedagogical practice**	**Alignment with ACE protocol**
Sequenced instruction	**3. Teachers conclude an instructional sequence by checking students' understanding (that is, the Review phase from the Cue-Do-Review sequence).** During this phase, teacher checks and bolsters students' understanding of the routine/topic and/or the process of learning about the routine/topic.	Coded 1 if at least one action is checked under **monitoring learning or reflecting on learning** constructs within a single class period.	**Monitoring learning** and **reflecting on learning** constructs
	4. Teachers complete an entire instructional sequence (such as Cue-Do-Review) in order. According to KU-CRL, the entire Cue-Do-Review sequence should be observed at least once, and possibly more than once, within an ideal CLC classroom (unless, for example, students are taking a test). Levels 2 and 3 instructional subsequences (similar to Cue-Do-Review) also are typically completed within a single class period.	Coded 1 if all of the following are observed within a single class period: • **Evidence of Cue phase** • **Evidence of Do phase** in same/later observation segment as Cue • **Evidence of Review phase** in same/later observation segment as Do	**See shared pedagogical features 1, 2, and 3**

	Component of CLC pedagogical practice	Coding of pedagogical practice	Alignment with ACE protocol
Multiple instructional modalities	**5. Teachers specifically name the Content Enhancement Routines or Learning Strategies they are using.** CLC materials explicitly train teachers to do this to help students remember strategies and apply them independently.	Coded 1 if at least one of the following **named routines or strategies** is checked within a single class period: • FRAME/Framing Routine • Concept Routine • Planning/Organizing Routine • Vocabulary LINCing Routine • Question Exploration Routine • Bridging Strategy • Prediction Strategy • 7-Step Vocabulary Process • Paragraph Writing Strategy • Word Maps—Decoding	**Strategies list** (under applications)
	6. Teachers use graphic organizers. Graphic organizers are at the core of Content Enhancement Routines and Learning Strategies and help teachers display new content in an organized manner.	Coded 1 if at least one **graphic organizer** application is checked anytime within a single class period.	**Graphic organizer** application
	7. Teachers use oral and written discourse. Oral and written discourse is explicitly encouraged by CLC training materials and KU-CRL's fidelity checklists.	Coded 1 if, under applications, both the **oral application and the writing application** are checked for either of the following: • a single observation segment (possibly across multiple constructs) • a single construct (possibly across multiple observation segments)	**Oral and writing** applications

	Component of CLC pedagogical practice	Coding of pedagogical practice	Alignment with ACE protocol
	8. Teachers combine all four applications (named strategy, graphic organizer, oral discourse and written discourse) within a single instructional experience. Content Enhancement Routines typically blend named strategies, graphic organizers, and oral and written interactions within a single instructional experience.	Coded 1 if a **named strategy, graphic organizer, oral application and writing application** (shared pedagogical features 5, 6, and 7) are present within either of the following: • a single observation segment (possibly across multiple constructs) • a single construct (possibly across multiple observation segments)	**Strategies list, graphic organizer, oral application and written** applications
Interactive and scaffolded instruction	**9. Teachers use interactive discourse.** Interactive discourse is heavily emphasized within Content Enhancement Routines; students and teachers interactively co-construct knowledge. Interactive discourse also is emphasized within the Fusion Reading program.	Coded 1 if **teacher-to-student discourse, student-and-teacher discourse, or student-and-student discourse** is checked in at least half of the different instructional segments within a single class.	**Discourse** coherence item
	10. Teachers use interactive discourse throughout entire class. KU-CRL confirmed that ideal CLC classrooms should exhibit more co-construction and more pervasive student discourse.	Coded 1 if **teacher-to-student discourse, student-and-teacher discourse, or student-and-student discourse** is checked in all observed instructional segments within a single class.	**Discourse** coherence item
	11. Teachers use different phases of scaffolded instruction. Teachers are expected to help students become increasingly independent when using Content Enhancement Routines. Fusion teachers also are expected to scaffold instruction by providing initial guidance and support and encouraging students to work more independently.	Coded 1 if **at least two different phases of scaffolding** are checked within a single class period. Scaffolding phases include modeled, shared, guided, and independent instruction.	**Scaffolding** coherence item

	Component of CLC pedagogical practice	Coding of pedagogical practice	Alignment with ACE protocol
	12. Teachers use scaffolded instruction throughout the entire class. KU-CRL generally expects to see evidence of scaffolding throughout an ideal CLC class.	Coded 1 if **scaffolding is present across the entire class** (that is, modeled, shared, guided, or independent instruction) and checked for each observed instructional segment within a single class. The same phase or multiple phases of scaffolding can be present across observed instructional segments.	**Scaffolding** coherence item

Appendix I. Interrater Reliability for Shared Pedagogical Features

Interrater reliability was calculated to determine whether the multi-item indicators that were constructed to reflect the emphasized pedagogical practices of the Content Literacy Continuum (CLC) produced similar scores across observers. Interrater agreement shows the consistency of interpretation of target behaviors across observers. Poor interrater agreement produces unwanted error and jeopardizes the precision of estimates.

Interrater reliability estimates for year 1 are based on pilot observations in eight schools conducted by pairs of observers in spring 2009. Each pair of observers sat in on the same classes in the same schools. Estimates of interrater reliability for year 2 were calculated by examining ratings made by all observers after viewing a film clip following the training in fall 2009.[181] In both years, the interrater reliability estimates are based on constructed variables (i.e., fidelity features) and not the coded variables themselves (see appendix B for the observation protocol that includes all constructs, actions, and applications and appendix H for the criteria for coding each fidelity feature).

This appendix summarizes the procedure used to calculate interrater reliability for year 1 and year 2. Estimates for interrater reliability then are presented for each year of data collection.

Calculating interrater reliability for year 1

Each class period consisted of between two and four observational segments of roughly 10–15 minutes each. Each observer first rated each 10-15 minute segment according to whether the twelve fidelity features were present. Then, for each rater the scores at the segment level were aggregated to the class period level to provide the data to calculate agreement percentages.[182] At this point, each rater had a set of ratings on each of the twelve fidelity features and the ratings were compared among raters to obtain the agreement percentages.[183] In practice if two raters agreed that a feature was present (or absent) during a class period (for example, the Cue phase), this indicated agreement.[184] Conversely if the two raters disagreed about whether the Cue phase was apparent, this indicated disagreement. Interrater agreement was calculated for each component.

[181] Thus, reliability estimates are based on data collected as separate exercise, rather than from data collected during actual observations for this study.

[182] A caveat to aggregating across segments to obtain class period scores is that for fidelity features 7 and 8, a point was awarded at the class period level for the presence of an application either across segments within a given construct or across constructs within a given segment.

[183] Other indicators of interrater reliability, such as Cohen's kappa or coefficient alpha, are preferable to percent agreement because they account for "chance" agreement. However, the very low incidence of some behaviors and constructs within classrooms or the film clip made the kappa and alpha statistics impossible to calculate (for example, creating a denominator for kappa calculation of 0). The study team therefore chose to use percent agreement and add cautions to readers that these reliability indices represent upper bound estimates.

[184] There are multiple ways that an observer's ratings can be classified as showing the presence of a fidelity feature. For instance, for the Cue phase to be present, an observer can witness any of three actions (e.g., Explicitly identifies purpose, Supports students in identifying purpose, and Purpose is implied). An agreement would be evident if two raters witnessed two different actions because any of the three actions is coded as the presence of the Cue phase.

To examine interrater agreement for the indicator level (representing each CLC-emphasized pedagogical practice: sequenced instruction, multiple instructional modalities, and interactive and scaffolded instruction), observers' scores for the root components were added together. For each observer pair, percent agreement at this broader indicator level represents the match in the sum for observer 1 versus the sum for observer 2. A match between observer 1's aggregated score (ranging from 0 to 4) and observer 2's aggregated score was classified as agreement; otherwise it classified as disagreement. The percent agreement represents the agreements across all observer pairs for each class that was observed. Thus the process of adding across components produces reliability estimates for these aggregates that are smaller (that is, less reliable) than are the reliability estimates for their root components.

Calculating interrater reliability for year 2

Interrater reliability for year 2 was calculated based on all observers' ratings from a single film clip of instruction. The single film clip was divided into three segments of 10 minutes each for a total length of the video of 35 minutes. Fidelity features were calculated across segments. The same video was rated by all observers. The video clip used for determining interrater reliability showed a teacher facilitated discussion of Shakespeare's *Othello* within an actual high school classroom.

Interrater reliability for year 2 was calculated differently from year 1. Reliability estimates were calculated by comparing each observer's ratings with the ratings of every other rater for a given component/construct for a CLC-emphasized pedagogical practice. For example, an agreement between observer 1 and observer 2 that the Cue phase was present would indicate agreement. If observer 3 then did not perceive the presence of the Cue phase, then there would be a disagreement between that trio of observers. This process would continue until observer 1 was compared with all observers, observer 2 was compared with all observers, and so forth. In the three observer example just described, the Cue phase would have 33 percent agreement because there would be one agreement (between observer 1 and observer 2) but two disagreements (between observers 1 and 3 and observers 2 and 3). To obtain the overall percent agreement for each shared pedagogical feature, the number of agreements was divided by the number of agreements plus disagreements.

To calculate percent agreement for the broader indicators of CLC-emphasized pedagogical practices, each observer's ratings for the components of each indicator were added together, and the process of matching against aggregated scores of other observers was performed again. The percent agreement for each construct was the number of agreements divided by the number of agreements plus disagreements. As with the calculation of percent agreement among observer pairs for year 2, the process of aggregating across components resulted in smaller reliability coefficients for the aggregated indicators than for each of the root components.

Interrater reliability estimates for year 1

The percent agreement for 12 components and 3 broader pedagogical practices is shown in table I.1. The second column shows the percent agreement for all observers. The reliability estimates range from 57 percent agreement for the Cue and Review phases to 96 percent agreement for the

code indicating whether teachers combine all four applications for multiple instructional modalities (that is, used a named strategy, graphic organizer, oral discourse and written discourse). Most of the percent agreement for components is 70 percent or greater. The percent agreement for the aggregated pedagogical practices is as follows: 43 percent for sequenced instruction, 43 percent for multiple instructional modalities, and 39 percent for interactive and scaffolded instruction. The average percent agreement across these aggregated practices is 42 percent.

Table I.1. Percent agreement for 12 shared pedagogical features and 3 shared pedagogical constructs for spring 2009 data observations

Shared pedagogical features and constructs	Percent agreement
Sequenced instruction	**0.43**
Cue phase	0.57
Do phase	0.78
Review phase	0.57
Entire Cue-Do-Review sequence	0.70
Multiple instructional modalities	**0.43**
Named strategies	0.74
Graphic organizer	0.78
Oral and written discourse	0.74
Combine all strategies	0.96
Interactive instruction	**0.39**
Interactive discourse	0.61
Interactive discourse for entire class	0.65
Scaffolded instruction	0.87
Scaffolded instruction for entire class	0.74
Average	**0.42**

Note: Shared pedagogical constructs are in listed in bold.

Interrater reliability estimates for year 2

The interrater reliability estimates for year 2 of data collection are shown in table I.2. These estimates are based on ratings of 10 observers of the same instructional videotape. The percent agreement for the root components ranged from 47 percent to 100 percent. The percent agreement for the broader indicators of shared pedagogical practices was 36 percent for sequenced instruction, 47 percent for multiple instructional modalities, and 38 percent for interactive and scaffolded instruction.

Table I.2. Percent agreement for 12 shared pedagogical features and 3 shared pedagogical constructs for fall 2009 data observations

Shared pedagogical features and constructs	Percent agreement
Sequenced instruction	**0.36**
Cue phase	0.47
Do phase	0.80
Review phase	0.80
Entire Cue-Do-Review sequence	1.00
Multiple instructional modalities	**0.47**
Named strategies	1.00
Graphic organizer	1.00
Oral and written discourse	0.47
Combine all strategies	1.00
Interactive instruction	**0.38**
Interactive discourse	1.00
Interactive discourse for entire class	0.80
Scaffolded instruction	0.47
Scaffolded instruction for entire class	0.53
Average	**0.40**

Note: Shared pedagogical constructs are in listed in bold. Observers received additional training prior to conducting these year 2 observations from which the fidelity of instruction is based.

Summary of interrater reliability estimates

As noted throughout this appendix and appendix H, there exist a number of concerns which suggest that readers be skeptical of the observation-based findings in this report. First, inter-rater reliability checks were performed prior to actual observations (i.e., during pilot observations in year 1 and immediately following training in year 2), rather than during the actual observations conducted in study schools. Second, estimates of instructional fidelity are based on just one set of observations of classrooms in schools per year. Additional observations would provide more reliable estimates. Third, inter-rater reliabilities were calculated on aggregates across segments and activities. Finally, reliability estimates are based on percent agreement, rather than statistics that account for chance agreement. Thus, the reliabilities cited in this appendix should be considered "upper bound" estimates.

Appendix J. Robustness of Estimated Impacts on Student Outcomes

This appendix presents the results of sensitivity analyses that were conducted to examine whether the primary impact findings presented in chapter 5 are robust to the specification of the statistical model used for the impact analysis. The first sensitivity analysis looks at the robustness of the findings to the weighting scheme used when pooling the impact findings across random assignment blocks, while the second sensitivity analysis looks at the sensitivity of the findings to controlling for students' background characteristics and prior achievement. The tables also present the standard error and confidence interval for the primary impact findings presented in chapter 5.

Alternate weighting schemes for pooling the impact findings

The first sensitivity analysis examines whether the primary impact findings are robust to alternate methods of weighting the block-specific impact estimates when pooling the results across random assignment blocks. As explained in chapter 2 and appendix E, the overall (or pooled) impact of the Content Literacy Continuum (CLC) framework is obtained by estimating the impact for each random assignment block and then taking a weighted average of the block-specific impact estimates, where the weight for a given block is the proportion of CLC-group schools that are located in that block. Given this weighting scheme, the pooled impact estimate represents the effect of the CLC framework for the average school in the CLC group.

To test the sensitivity of the primary impact findings to this weighting scheme, we estimated the pooled impact of the CLC intervention using two alternative strategies for weighting the bloc-specific impact estimates (each of which yields a different interpretation for the pooled impact estimate):

- **Weight by precision.** In this approach, the impact estimate for a given random assignment block is weighted based on its precision relative to the other block-specific estimates.[185] Thus blocks whose impact is more precisely estimated (that is, those with a smaller standard error) are weighted more heavily in the overall impact estimate. Here the interpretation of the pooled impact estimate represents (approximately) the effect of the CLC framework for the average student in the study.[186]
- **Equal weight.** In this weighting scheme, each block-specific estimate is weighted equally when calculating the overall impact estimate. Thus the estimated impact for school districts with two study schools has the same weight in the pooled result as do districts with a larger number of study schools. Here the pooled impact estimate represents the effect of the CLC framework for the average block in the study.

[185] Precision is defined as the inverse of the variance of the impact estimate.

[186] In a cluster randomized experiment, precision is a function of both the number of schools and the number of students. Therefore the interpretation of the pooled impact estimate is actually somewhere between the effect for the average student and the average school in the study.

The sensitivity of the primary impact findings in this study to the choice of block weighting scheme is examined in tables J.1 and J.2. The sensitivity of estimated impacts on GRADE reading comprehension scores in year 2 is looked at in table J.1; impacts on credit accumulation in year 2 are looked at in table J.2. For comparative purposes, the corresponding impact estimates presented in chapter 5 (which are weighted based on the number of CLC schools in each block) also are shown in the tables.

Conclusions about the effectiveness of CLC are not sensitive to the choice of weighting scheme. Although the estimated effect of CLC on credit accumulation is negative and statistically significant when blocks are weighted by precision, the statistical significance of this result does not hold when adjustments are made for multiple hypothesis testing.[187] Therefore we can conclude that after two years of implementation, CLC did not improve reading scores or worsen credit accumulation by a statistically significant amount for the average school, student, or random assignment block in the study.

Unadjusted impact estimates

As explained in chapter 2 and appendix E, the statistical model used to estimate impacts controls for several measures of students' background characteristics and prior achievement. It is not strictly necessary to control for these characteristics because random assignment should ensure that students in CLC and non-CLC schools are similar in expectation with respect to their characteristics and prior achievement. However, by including highly predictive student characteristics in the model, it is possible to improve the precision of the impact estimates.

To examine the extent to which estimated impacts on the primary outcomes are sensitive to the inclusion of these covariates, the statistical model was reestimated *without* controlling for students' background characteristics and prior achievement.[188] The unadjusted impact of the programs on GRADE reading comprehension scores in year 2 is presented in table J.3; the unadjusted estimates of impacts on credit accumulation in year 2 is presented in table J.4. For comparative purposes, the adjusted impact estimates presented in chapter 5 also are included in the tables.

[187] As explained in chapter 2, as a protection against false positives, *p*-values greater than 0.05 for impacts on the four primary outcomes must be adjusted for multiple hypothesis testing. This is done using the procedure described in Hochberg and Benjamini (1990).

[188] These sensitivity tests still include random assignment blocks as fixed effects to account for the way in which random assignment was conducted.

Table J.1. Sensitivity of the impact estimates to the weighting strategy used to pool impact findings across blocks, impacts on reading comprehension scores in year 2 of the study, GRADE respondent sample

Weighting strategy	Estimated impact[a]	Standard error	Effect size	*p*- value
Grade 9 sample				
Blocks weighted by number of CLC schools[b]	0.89	0.75	0.06	0.262
Blocks weighted by precision[c]	−0.55	1.09	−0.03	0.623
Blocks weighted equally[d]	0.02	0.86	0.00	0.978
Sample size				
Students (total = 5,011)	2,975	2,036		
Schools (total = 28)	15	13		
Grade 10 sample				
Blocks weighted by number of CLC schools	1.59	1.17	0.10	0.203
Blocks weighted by precision	0.67	0.86	0.04	0.446
Blocks weighted equally	1.42	1.21	0.09	0.267
Sample size				
Students (total = 4,546)	2,908	1,638		
Schools (total = 28)	15	13		

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100 and its standard deviation is 15.

b. Block-specific impact estimates are weighted by the relative number of CLC schools in each block. These are the impact estimates reported in the main body of the report.

c. Block-specific impact estimates are weighted by the relative precision of each impact estimate (inverse of the variance of the impact estimate).

d. Block-specific impacts estimates are equally weighted.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table J.2. Sensitivity of the impact estimates to the weighting strategy used to pool impact findings across blocks, impacts on credit accumulation in year 2 of the study, school records sample

Weighting strategy	Estimated impact[a]	Standard error	Effect size	*p*-value
Grade 9 sample				
Blocks weighted by number of CLC schools[b]	−1.60	0.75	−0.17	0.058
Blocks weighted by precision[c]	−1.59	0.64	−0.17	0.023*
Blocks weighted equally[d]	−1.70	0.85	−0.18	0.075
Sample size				
Students (total = 7,951)	4,467	3,484		
Schools (total = 28)	15	13		
Grade 10 sample				
Blocks weighted by number of CLC schools	0.39	1.08	0.02	0.726
Blocks weighted by precision	0.33	1.00	0.02	0.749
Blocks weighted equally	0.39	1.24	0.02	0.760
Sample size				
Students (total = 8,514)	4,888	3,626		
Schools (total = 28)	15	13		

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned by students is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

b. Block-specific impact estimates are weighted by the relative number of CLC schools in each block. These are the impact estimates reported in the main body of the report.

c. Block-specific impact estimates are weighted by the relative precision of each impact estimate (inverse of the variance of the impact estimate).

d. Block-specific impacts estimates are equally weighted.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

As expected, the magnitude of adjusted and unadjusted impact estimates is similar, but the standard error of the unadjusted estimates is larger. Conclusions about program effectiveness are the same regardless of whether or not estimates are adjusted.

Table J.3. Adjusted and unadjusted estimates, impacts on reading comprehension scores in year 2 of the study, GRADE respondent sample

Regression adjustments	Estimated impact[a]	Standard error	*p*-value	Lower 95% confidence interval	Upper 95% confidence interval	Effect size
Grade 9 sample						
Adjusted for student characteristics[b]	0.89	0.75	0.262	−0.78	2.56	0.06
Not adjusted for student characteristics[c]	1.25	1.10	0.281	−1.19	3.69	0.08
Sample size						
Students (total = 5,011)	2,975	2,036				
Schools (total = 28)	15	13				
Grade 10 sample						
Adjusted for student characteristics[b]	1.59	1.17	0.203	−1.01	4.19	0.10
Not adjusted for student characteristics[c]	2.01	1.49	0.207	−1.31	5.32	0.12
Sample size						
Students (total = 4,546)	2,908	1,638				
Schools (total = 28)	15	13				

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100 and its standard deviation is 15.

b. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender.

c. Impact estimates are regression-adjusted for the blocking of random assignment only.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table J.4. Adjusted and unadjusted estimates, impacts on credit accumulation in year 2 of the study, school records sample

Regression adjustments	Estimated impact[a]	Standard error	*p*-value	Lower 95% confidence interval	Upper 95% confidence interval	Effect size
Grade 9 sample						
Adjusted for student characteristics[b]	−1.60	0.75	0.058	−3.26	0.06	−0.17
Not adjusted for student characteristics[c]	−1.65	0.79	0.063	−3.40	0.11	−0.17
Sample size						
Students (total = 7,951)	4,467	3,484				
Schools (total = 28)	15	13				
Grade 10 sample						
Adjusted for student characteristics[b]	0.39	1.08	0.726	-2.01	2.78	0.02
Not adjusted for student characteristics[c]	0.74	1.14	0.533	−1.80	3.28	0.04
Sample size						
Students (total = 8,514)	4,888	3,626				
Schools (total = 28)	15	13				

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned by students is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

b. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender.

c. Impact estimates are regression-adjusted for the blocking of random assignment only.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Appendix K. Model Fit Information

This appendix presents information about the fit of the models used to estimate impacts on student outcomes in this report. Model fit information for estimated impacts on GRADE reading achievement scores in year 2 of the study is presented in table K.1; information for estimated impacts on course performance outcomes in the second year is presented in table K.2.

Since a multilevel model is used to estimate impacts (students within schools), there are two levels of outcome variance in the analysis: variation between schools and variation between students within schools. For this reason several pieces of information about model fit are presented in the tables in this appendix:

- **Intraclass correlation (ICC):** this is the proportion of the total variation in the student outcomes that is between schools (as opposed to within schools). For the reading achievement outcomes, the ICC ranges from 0.14 to 0.22, depending on the outcome, grade level, and study year. For the course performance outcomes, it ranges from 0.10 to 0.15.
- **Between-school R^2:** this is the proportion of between-school variation that is explained by the student-level covariates included in the model (students' background characteristics and prior achievement) and the school-level covariates in the model (random assignment block indicators).[189] For the reading achievement outcomes, the between-school R^2 ranges from 0.64 to 0.96. For the course performance outcomes, it ranges from 0.56 to 0.89.
- **Within-school R^2:** this is the proportion of within-school variation that is explained by the student-level covariates. For the reading achievement outcomes, the within-school R^2 ranges from 0.21 to 0.27. For the course performance outcomes, it ranges from 0.12 to 0.33.

As noted in appendix F, these three factors affect the minimum detectable effect size (MDES) of the study. All else equal, the MDES decreases when the ICC is smaller, and when the two R^2s are larger. With a large ICC, a high between-school R^2 is especially important for detecting impacts of reasonable magnitude.

[189] The student-level covariates are whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, gender, and dummy indicators of missing data for each of these characteristics.

Table K.1. Model fit information for estimated impacts on reading achievement, GRADE respondent sample

Outcome[a]	Intraclass correlation	Between-school R^2	Within-school R^2
Year 1			
Grade 9 sample			
Reading achievement (standard score)			
Reading comprehension	0.16	0.64	0.27
Reading vocabulary	0.14	0.96	0.27
Year 2			
Grade 9 sample			
Reading achievement (standard score)			
Reading comprehension	0.19	0.96	0.24
Reading vocabulary	0.22	0.95	0.23
Grade 10 sample			
Reading achievement (standard score)			
Reading comprehension	0.19	0.86	0.23
Reading vocabulary	0.17	0.88	0.21

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. The intraclass correlation is the proportion of the total outcome variation that is between schools (as opposed to between students within schools). The between-school R^2 is the proportion of between-school variation that is explained by the random assignment blocks and the student-level covariates (that is, whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender). The within-school R^2 is the proportion of within-school variation that is explained by the student-level covariates.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 10 students and Form B to grade 10 students).

Table K.2. Model fit information for estimated impacts on course performance, school records sample

Outcome	Intraclass correlation	Between-school R^2	Within-school R^2
Year 1			
Grade 9 sample			
Credits earned in core subject areas (%)[a]	0.11	0.56	0.12
GPA in core subject areas[b]	0.10	0.70	0.15
Year 2			
Grade 9 sample			
Credits earned in core subject areas (%)	0.13	0.72	0.14
GPA in core subject areas	0.13	0.79	0.17
Grade 10 sample			
Credits earned in core subject areas (%)	0.15	0.89	0.33
GPA in core subject areas	0.12	0.77	0.15

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. The intraclass correlation is the proportion of the total outcome variation that is between schools (as opposed to between students within schools). The between-school R2 is the proportion of between-school variation that is explained by the random assignment blocks and the student-level covariates (that is, whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, E SL status, special education status, racial/ethnic group, and gender). The within-school R2 is the proportion of within-school variation that is explained by the student-level covariates.

a. The cumulative number of credits earned is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

b. GPA in core subject areas is based on a four-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Appendix L. Descriptive Statistics for Student Outcomes and Impact Model Covariates

This appendix presents unadjusted student-level descriptive statistics for the key student outcomes and impact model covariates used in the analysis:

- Student-level means for the key outcome measures (reading achievement and course performance) are presented in tables L.1 and L.2. As explained in chapter 2 (box 2.1), the mean outcome levels presented in the impact tables are weighted by the number of CLC schools in each random assignment block, and they also are regression-adjusted to account for the blocking of random assignment and students' baseline characteristics. In contrast the values in tables L.1 and L.2 are unadjusted student-level means.[190] The means in these tables are for both program groups together (CLC and non-CLC schools). Unadjusted means by program group are not shown because the comparison of outcomes across program groups should be based on the *adjusted* means only, as the latter account for the blocked experimental design.
- The standard deviation of the key student outcome measures in this study, overall and by program group, for the reading achievement and course performance outcomes respectively appear in tables L.3 and L.4. The values in the "non-CLC schools" row are used to calculate the effect sizes presented in the impact tables. Standard deviations also are presented for the CLC group and both groups pooled together.
- Student-level standard deviations for the student characteristics used as covariates in the impact model (see appendix E for a description of the model) are presented in tables L.5 and L.6. For the unadjusted student-level mean of these characteristics, see chapter 2. For adjusted means by program group, see appendix C.

[190] Means for GRADE scores are still weighted to account for the stratified sampling of students for GRADE testing in some districts (see appendix A); however, they are not weighted by the number of schools.

Table L.1. Student-level means for reading achievement outcomes, GRADE respondent sample

Outcome[a]	Year 1	Year 2	
	Grade 9 sample	Grade 9 sample	Grade 10 sample
Reading comprehension			
Average standard score	92.2	91.9	97.4
Corresponding grade equivalent	*6.7*	*6.6*	*8.0*
Corresponding percentile	*30*	*29*	*42*
Reading vocabulary			
Average standard score	94.7	94.3	101.0
Corresponding grade equivalent	*8.0*	*7.9*	*9.8*
Corresponding percentile	*35*	*35*	*51*
Sample size			
Students	4,786	5,011	4,546
Schools	28	28	28

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. The values presented in this table are student-level means. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities.

a. The national average for standard scores is 100 and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A).

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table L.2. Student-level means for course performance outcomes, school records sample

Outcome	Year 1	Year 2	
	Grade 9 sample	Grade 9 sample	Grade 10 sample
Credits earned in core subject areas (%)[a]	23.3	23.3	41.0
English language arts	19.8	20.1	35.8
Social studies	26.5	26.3	45.5
Science	27.4	27.6	47.6
Math	21.9	21.5	38.8
Sample size			
Students	7,365	7,951	8,514
Schools	28	28	28
GPA in core subject areas[b]	1.9	1.9	1.8
English language arts	2.0	2.0	2.0
Social studies	2.0	2.0	1.9
Science	1.8	1.8	1.8
Math	1.8	1.8	1.7
Sample size			
Students	7,315	7,917	8,209
Schools	28	28	28

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. The values presented in this table are student-level means.

a. The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject specific) required for graduation in a student's district.

b. GPA in core subject areas is based on a four-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum Study.

Table L.3. Student-level standard deviations for reading achievement outcomes, GRADE respondent sample

Outcome[a]	Year 1	Year 2	
	Grade 9 sample	Grade 9 sample	Grade 10 sample
Reading comprehension			
CLC schools	15.4	15.7	16.6
Non-CLC schools	15.7	16.1	16.2
All schools	15.5	15.8	16.4
Reading vocabulary			
CLC schools	15.7	16.4	15.9
Non-CLC schools	16.1	16.2	15.9
All schools	15.9	16.3	15.9
Sample size			
Students	4,786	5,011	4,546
Schools	28	28	28

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. The values presented in this table are student-level standard deviations.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A).

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to ninth-grade students and Form B to grade 10 students).

Table L.4. Student-level standard deviation for course performance outcomes, school records sample

Outcome	Year 1	Year 2	
	Grade 9 sample	Grade 9 sample	Grade 10 sample
Credits earned in core subject areas (%)[a]			
CLC schools	9.8	10.2	20.1
Non-CLC schools	9.3	9.5	20.6
All schools	9.6	9.9	20.3
Sample size			
Students	7,365	7,951	8,514
Schools	28	28	28
GPA in core subject areas[b]			
CLC schools	1.1	1.1	1.1
Non-CLC schools	1.1	1.1	1.1
All schools	1.1	1.1	1.1
Sample size			
Students	7,315	7,917	8,209
Schools	28	28	28

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. The values presented in this table are student-level standard deviations.

a. The cumulative number of credits earned is scaled as a percentage of the total number of credits (core or subject-specific) required for graduation in a student's district.

b. GPA in core subject areas is based on a four-point scale: A+/A/A- = 4.0; B+/B/B- = 3.0; C+/C/C- = 2.0; D+/D/D- = 1.0; F = 0.0.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum Study.

Table L.5. Student-level standard deviations for impact model covariates, GRADE respondent sample

Characteristic	Year 1	Year 2	
	Grade 9 sample	Grade 9 sample	Grade 10 sample
Overage for grade (%)[a]			
CLC schools	0.41	0.42	0.41
Non-CLC schools	0.43	0.45	0.44
All schools	0.42	0.43	0.42
Free/reduced-price lunch eligible (%)			
CLC schools	0.47	0.46	0.48
Non-CLC schools	0.48	0.49	0.49
All schools	0.48	0.47	0.48
Race/ethnicity (%)			
Hispanic			
CLC schools	0.22	0.20	0.19
Non-CLC schools	0.20	0.22	0.19
All schools	0.21	0.21	0.19
Black, non-Hispanic			
CLC schools	0.50	0.50	0.50
Non-CLC schools	0.50	0.50	0.50
All schools	0.50	0.50	0.50
White and other			
CLC schools	0.50	0.50	0.50
Non-CLC schools	0.50	0.49	0.50
All schools	0.50	0.50	0.50
Gender (% male)			
CLC schools	0.50	0.50	0.50
Non-CLC schools	0.50	0.50	0.50
All schools	0.50	0.50	0.50
English language learner (%)			
CLC schools	0.33	0.30	0.33
Non-CLC schools	0.27	0.22	0.22
All schools	0.31	0.27	0.29
Special education (%)			
CLC schools	0.31	0.30	0.31

Characteristic	Year 1	Year 2	
	Grade 9 sample	Grade 9 sample	Grade 10 sample
Non-CLC schools	0.32	0.33	0.35
All schools	0.32	0.31	0.33
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)			
Reading/ELA			
CLC schools	0.47	0.48	0.46
Non-CLC schools	0.47	0.49	0.46
All schools	0.47	0.48	0.46
Mathematics			
CLC schools	0.49	0.49	0.49
Non-CLC schools	0.49	0.49	0.48
All schools	0.49	0.49	0.49
Sample size[b]			
Students	4,786	5,011	4,546
Schools	28	28	28

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Background characteristics are measured in grade 8 or at the start of grade 9. The values presented in this table are student-level standard deviations.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. Due to missing values, the number of students varies by characteristic. The sample size reported here is for the full analysis sample. For the impact analysis, missing values are imputed with zero and indicators of missing information (one for each characteristic) are included in the regression model. The percentage of missing data is highest for grade 8 test scores: 21.9 percent (math) and 22.1 percent (reading) for grade 9 students in the first year; 26.4 percent (math) and 26.7 percent (reading) for grade 9 students in the second year; and 35.5 percent (math) and 35.8 percent (reading) for grade 10 students in the second year. The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.9 percent (age) to 2.0 percent (free-lunch status) for grade 9 students in the first year; from 0.1 percent (age) to 2.6 percent (special education status) for grade 9 students in the second year; and from 0.1 percent (age) to 15.1 percent (English language learner status) for grade 10 students in the second year.

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table L.6. Student-level standard deviations for impact model covariates, school records sample

Characteristic	Year 1	Year 2	
	Grade 9 sample	Grade 9 sample	Grade 10 sample
Overage for grade (%)[a]			
CLC schools	0.44	0.45	0.47
Non-CLC schools	0.46	0.46	0.48
All schools	0.45	0.46	0.47
Free/reduced-price lunch eligible (%)			
CLC schools	0.46	0.46	0.46
Non-CLC schools	0.48	0.49	0.48
All schools	0.47	0.47	0.47
Race/ethnicity (%)			
Hispanic			
CLC schools	0.22	0.22	0.22
Non-CLC schools	0.22	0.24	0.22
All schools	0.22	0.23	0.22
Black, non-Hispanic			
CLC schools	0.50	0.50	0.50
Non-CLC schools	0.50	0.50	0.49
All schools	0.50	0.50	0.50
White and other			
CLC schools	0.49	0.49	0.50
Non-CLC schools	0.50	0.50	0.50
All schools	0.50	0.50	0.50
Gender (% male)			
CLC schools	0.50	0.50	0.50
Non-CLC schools	0.50	0.50	0.50
All schools	0.50	0.50	0.50
English language learner (%)			
CLC schools	0.31	0.30	0.32
Non-CLC schools	0.28	0.30	0.29
All schools	0.29	0.30	0.31
Special education (%)			
CLC schools	0.38	0.38	0.38

Characteristic	Year 1	Year 2	
	Grade 9 sample	Grade 9 sample	Grade 10 sample
Non-CLC schools	0.41	0.39	0.41
All schools	0.39	0.38	0.40
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)			
Reading/ELA			
CLC schools	0.49	0.49	0.49
Non-CLC schools	0.49	0.49	0.49
All schools	0.49	0.49	0.49
Mathematics			
CLC schools	0.50	0.50	0.50
Non-CLC schools	0.50	0.49	0.50
All schools	0.50	0.50	0.50
Sample size[b]			
Students	7,365	7,951	8,514
Schools	28	28	28

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Background characteristics are measured in grade 8 or at the start of grade 9. The values presented in this table are student-level standard deviations.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. Due to missing values, the number of students varies by characteristic. The sample size reported here is for the full analysis sample. For the impact analysis, missing values are imputed with zero and indicators of missing information (one for each characteristic)are included in the regression model. The percentage of missing data is highest for grade 8 test scores: 23.2 percent (math) and 24.0 percent (reading) for grade 9 students in the first year; 31.0 percent (math) and 31.7 percent (reading) for grade 9 students in the second year; and 50.3 percent (math) and 51.5 percent (reading) for grade 10 students in the second year. The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.0 percent (age) to 1.6 percent (English language learner status) for grade 9 students in the first year; from 0.4 percent (age) to 2.2 percent (special education status) for grade 9 students in the second year; and from 0.7 percent (age) to 23.2 percent (English language learner status) for grade 10 students in the second year.

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Appendix M. Impact Findings and Sample Characteristics for All Study Schools in Year 1

This appendix presents first-year impact findings for all 33 schools that participated in year 1. As explained in chapter 2, 33 schools agreed to implement the CLC framework in year 1 of the study, but not all of these schools were able to continue in year 2. To make it possible to compare impact findings across study years, the first-year impact findings presented in chapter 6 of this report are based on the 28 schools that were able to participate in both years of the study. This appendix examines whether conclusions about the CLC framework in year 1 differ when all 33 schools are included in the analysis for that year.

As context for interpreting these impact findings, the appendix begins by providing information on the characteristics of these 33 schools prior to random assignment, and examines the characteristics of the ninth-grade students in the analysis samples in these schools in year 1.[191] The appendix then presents first-year impact findings based on all 33 schools that participated in the first year of the study.

School characteristics

The characteristics of all 33 schools that participated in year 1, in the school year prior to random assignment (2006/07), are presented in table M.1. The characteristics of the 28 schools that participated in both years to the characteristics of the 5 schools that left the study in year 2 also are compared in the table.[192] As seen here, the characteristics of the 5 schools that withdrew in year 2 are systematically different from the characteristics of schools that continued in year 2. Most notably the 5 schools that withdrew have a higher proportion of minority students.

As also shown in table M.2, for this group of 33 schools, there is no systematic difference between the characteristics of CLC and non-CLC schools prior to random assignment. This confirms that the two groups of schools were statistically equivalent in expectation before the start of the study, and the non-CLC group provides a counterfactual for what would have happened in the CLC schools had they not implemented the intervention.

Student sample characteristics

The characteristics of grade 9 students in the school records sample and the GRADE respondent sample in year 1 of the study, based on all 33 schools, are presented in tables M.3 and M.4. As seen in these tables, there is no systematic difference between students in CLC and non-CLC schools with respect to their background characteristics and prior achievement, in either the GRADE respondent sample or the school records sample. Therefore differences between the

[191] In the first year of the study, the CLC framework was implemented in grade 9 only.

[192] As explained in chapter 2, five schools were no longer in the year 2 sample either because they withdrew from the study after year 1 due to competing school reform priorities or were closed as a result of district restructuring.

academic performance and reading achievement of students in CLC and non-CLC schools can be attributed to the effect of the intervention.

Table M.1. Characteristics of year 1 and year 2 study schools before random assignment (2006/07)

Characteristic	Year 1 schools	Year 2 schools (A)	Schools that left study (B)	Estimated difference (A) − (B)	*p*-value for estimated difference
Title I status (% of schools)	54.5	60.7	20.0	40.7	0.098
Free/reduced-price lunch eligible students (school average % of students)	57.4	56.9	60.1	−3.2	0.671
Race/ethnicity (school average % of students)					
Hispanic	4.5	4.3	5.5	−1.2	0.734
Black, non-Hispanic	55.5	49.2	90.5	−41.2	0.036*
White, non-Hispanic	36.3	42.2	3.3	38.9	0.024*
Other	3.2	3.7	0.6	3.1	0.269
Male (school average % of students)	50.7	50.9	49.5	1.5	0.285
Average total school enrollment	1314	1387	905	481	0.055
Percentage of students in grade 9	30.9	31.2	29.3	1.9	0.616
Percentage of students in grade 10	25.8	25.4	28.4	−3.0	0.081
Percentage of students in grade 11	21.7	21.6	22.0	−0.4	0.843
Percentage of students in grade 12	18.4	18.0	20.3	−2.3	0.433
Average promoting power (%)	63.1	62.0	69.3	−7.3	0.638
Test of systematic difference between groups[a] ($\chi 2 = 28.1$)					0.005*
Number of schools	33	28	5		

* *p*-value ≤ .05

Note: This table provides information on the characteristics of schools in the school year prior to random assignment (2006–20/07). Values in the "Year 1 study schools" column are for the 33 schools that participated in the first year of the study. Values in the "Year 2 study schools" column are for the 28 schools that participated in both years of the study. Values in the "Schools that Left Study" column are for the 5 schools that participated in the first year of the study but not the second. Rounding may cause slight discrepancies in calculating sums and differences.

a. An omnibus chi-squared test was used to determine whether there is a systematic difference between schools in the CLC group and the non-CLC group, with respect to the characteristics included in this table.

Source: U.S. Department of Education, National Center for Education Statistics, The NCES Common Core of Data (CCD), "Public Elementary/Secondary School Universe Survey Data," 2006/07 and 2003/04.

Table M.2. Characteristics of year 1 study schools before random assignment (2006/07), by treatment status

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*-value
Title I status (% of schools)	58.8	52.9	5.9	0.253
Free/reduced-price lunch eligible students (school average % of students)	57.6	57.6	0.0	0.987
Race/ethnicity (school average % of students)				
Hispanic	4.4	4.3	0.1	0.942
Black, non-Hispanic	56.6	55.7	0.9	0.843
White, non-Hispanic	35.5	36.1	−0.5	0.879
Other	2.9	3.3	−0.4	0.659
Male (school average % of students)	50.2	51.3	−1.1	0.140
Average total school enrollment	1394	1212	182	0.152
Percentage of students in grade 9	29.3	32.9	−3.6	0.114
Percentage of students in grade 10	26.4	25.3	1.1	0.316
Percentage of students in grade 11	22.1	21.3	0.7	0.397
Percentage of students in grade 12	19.3	17.3	2.0	0.147
Average promoting power (%)	70.0	55.6	14.4	0.076
Average number of full-time teachers[a]	81	75	7	0.322
Test of systematic difference between groups[b] ($\chi 2 = 12.6$)				0.476
Number of schools (total = 33)	17	16		

* *p*-value ≤ .05

Note: This table is based on the 33 schools that participated in the first year of the study, in the school year prior to random assignment (2006/07). Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. Data on the number of full-time teachers are available for not quite all of the study schools.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between schools in the CLC group and the non-CLC group, with respect to the characteristics included in this table.

Source: U.S. Department of Education, National Center for Education Statistics (NCES) Common Core of Data (CCD), Public Elementary/Secondary School Universe Survey Data, 2006/07 and 2003/04.

Table M.3. Background characteristics of students in year 1 of the study, school records sample (grade 9), year 1 study schools

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*- value
Average age (years)	14.8	14.8	0.0	0.817
Overage for grade (%)[a]	33.4	34.3	−0.8	0.498
Free/reduced-price lunch eligible (%)	69.3	69.1	0.2	0.939
Race/ethnicity (%)				
Hispanic	6.3	4.5	1.7	0.395
Black, non-Hispanic	57.7	56.1	1.6	0.726
White, non-Hispanic	31.8	35.0	−3.3	0.283
Other	4.3	4.3	−0.1	0.975
Gender (% male)	51.1	52.4	−1.3	0.335
English language learner (%)	7.5	7.0	0.6	0.852
Special education (%)	18.3	21.7	−3.4	0.072
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	55.2	52.8	2.4	0.445
Mathematics	47.6	47.1	0.4	0.906
Test of systematic difference between groups[b] ($\chi 2 = 10.5$)				0.396
Sample size[c]				
Students (total = 8,392)	4,649	3,743		
Schools (total = 33)	17	16		

* p-value ≤ .05

Note: This table is based on the 33 schools that participated in the first year of the study. Values in the tables are for students who are in grade 9 for the first time. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 24.6 percent (math) and 25.4 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.0 percent (age) to 1.4 percent (English language learner status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Table M.4. Background characteristics of students in year 1 of the study, GRADE respondent sample (grade 9), year 1 study schools

Characteristic	CLC group	Non-CLC group	Estimated difference	*p*- value
Average age (years)	14.7	14.7	0.0	0.702
Overage for grade (%)[a]	28.5	28.9	−0.4	0.726
Free/reduced-price lunch eligible (%)	68.7	67.5	1.2	0.694
Race/ethnicity (%)				
Hispanic	6.3	5.4	0.9	0.788
Black, non-Hispanic	57.4	54.9	2.5	0.657
White, non-Hispanic	32.8	35.1	−2.4	0.474
Other	3.6	4.4	−0.8	0.342
Gender (% male)	48.5	49.6	−1.2	0.511
English language learner (%)	6.9	8.1	−1.2	0.763
Special education (%)	11.1	10.6	0.5	0.825
Proficiency on grade 8 state assessments (% exceeding state accountability threshold)				
Reading/ELA	61.2	58.8	2.3	0.508
Mathematics	51.0	53.2	−2.2	0.552
Test of systematic difference between groups[b] ($\chi 2 = 11.1$)				0.352
Sample size[c]				
Students (total = 5,415)	3,113	2,302		
Schools (total = 33)	17	16		

* *p*-value ≤ .05

Note: This table is based on the 33 schools that participated in the first year of the study. Values in the tables are for students who are in grade 9 for the first time. Background characteristics are measured in grade 8 or at the start of grade 9. Difference estimates are regression-adjusted, controlling for the blocking of random assignment. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

b. An omnibus chi-squared test was used to determine whether there is a systematic difference between the CLC group and the non-CLC group, based on the characteristics included in this table.

c. Due to missing values, the number of students included varies by characteristic. The sample size reported here is for the full analysis sample. The percentage of missing data is highest for grade 8 test scores: 22.5 percent (math) and 22.7 percent (reading). The percentage of missing data is lowest for demographic characteristics and educational indicators: the percentage ranges from 0.9 percent (age) to 1.9 percent (free-lunch status).

Source: Calculations from historical school records data obtained as part of the Content Literacy Continuum study.

Student impacts

The estimated impact of the CLC framework on the two primary outcomes—reading comprehension and credit accumulation—in the first year of the study for all 33 schools is presented in tables M.5 and M.6. First-year impact findings for these 33 schools are similar to findings based on the 28 schools that participated in both year 1 and year 2:

- **Impact on reading comprehension.** It cannot be concluded that the CLC framework improved students' reading comprehension in year 1 (*p*-value = 0.164). The first-year effect size for all 33 schools (0.10) is similar in magnitude to the effect size for the 28 schools that participated in both years of the study (0.13).
- **Impact on credit accumulation.** Nor can it be concluded that the CLC framework helped students earn more credits in year 1 (*p*-value = 0.798). The first-year effect size for all 33 schools is –0.02, compared with −0.04 for the 28 schools that participated in both years.

Table M.5. Impacts on reading comprehension in year 1 study schools (grade 9 students), GRADE respondent sample

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Reading comprehension					
Average standard score	92.1	90.5	1.6	0.10	0.164
Corresponding grade equivalent	*6.6*	*6.2*			
Corresponding percentile	*29*	*25*			
Sample size					
Students (total = 5,415)	3,113	2,302			
Schools (total = 33)	17	16			

* *p*-value ≤ .05

Note: This table is based on the 33 schools that participated in the first year of the study. Findings in the table are for students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100 and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table M.6. Impacts on credit accumulation in year 1 study schools (grade 9 students), school records sample

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Credits earned in core subject areas (%)	22.7	22.9	-0.2	–0.02	0.798
English language arts	19.3	19.7	–0.4	–0.04	0.635
History	25.7	26.3	–0.6	–0.05	0.571
Science	26.7	25.7	0.9	0.06	0.431
Math	21.1	21.8	–0.6	–0.05	0.448
Sample size					
Students (total = 8,392)	4,649	3,743			
Schools (total = 33)	17	16			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Appendix N. Additional Impact Findings: Credits Attempted, Successful Credit Completion, and Attendance

As reported in chapter 5, the effect of the Content Literacy Continuum (CLC) on students' credit accumulation is not statistically significant. In this appendix, we look at impacts on two determinants of credits earned: (1) the number of credits attempted, and (2) student attendance.[193] Overall, we find that that CLC did not have a statistically significant effect on these two outcomes.

Credits attempted

The estimated impact of CLC on credits attempted, in the first and second year of the study, is presented in tables N.1 and N.2, respectively. The outcome measure here is defined as the cumulative number of credits in core classes (English language arts [ELA], social studies, science and mathematics) attempted by students, as a percentage of core credits required for graduation. As these tables show, the CLC framework did not have a statistically significant impact on the number of credits attempted by students.

Tables N.3 and N.4 present the estimated impact of CLC on credits earned *as a percentage of credits attempted*.[194] Scaling credits earned by the number of credits attempted implicitly controls for any difference in course-taking patterns between students in CLC and non-CLC schools. (In contrast the primary credit accumulation measure used in the report is defined as credits earned *as a percentage of the credits required for graduation*, which is a measure that does *not* control for the number of credits attempted.) One can see that the impact estimates in tables N.3 and N.4 (which control for credits attempted) are almost exactly the same in magnitude as the impact estimates in tables 5.2 and 6.2 (which do not control for credits attempted).

[193] The primary factor affecting credit accumulation is academic performance in the course, as measured by course grade. However, in some districts, student attendance also is used to determine whether a student gets a course credit; thus some students whose marks are higher than an F may not earn a credit if their attendance does not meet some minimum threshold.

[194] This outcome is not defined for students who did not attempt any core credits (for example, students who took only electives). This explains why the sample size for this analysis is slightly smaller than that for the analyses of credit accumulation in chapters 5 and 6.

Table N.1. Impacts on cumulative credits attempted, school records sample (grade 9), year 1

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Credits attempted in core subject areas (%)	29.2	29.1	0.1	0.02	0.848
English language arts	24.6	24.7	–0.1	–0.02	0.678
History	32.1	32.2	–0.1	–0.02	0.789
Science	34.2	33.8	0.4	0.05	0.630
Math	28.5	28.3	0.2	0.03	0.319
Sample size					
Students (total = 7,365)	4,254	3,111			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Table N.2. Impacts on cumulative credits attempted, school records sample, year 2

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
Credits attempted in core subject areas (%)	28.8	29.1	–0.2	–0.05	0.530
English language arts	23.9	24.5	–0.5	–0.11	0.013*
History	31.7	32.1	–0.3	–0.05	0.556
Science	34.1	34.0	0.1	0.01	0.874
Math	28.4	28.5	0.0	0.00	0.989
Sample size					
Students (total = 7,951)	4,467	3,484			
Schools (total = 28)	15	13			
Grade 10 sample					
Credits attempted in core subject areas (%)	51.9	51.7	0.2	0.01	0.769
English language arts	44.4	43.9	0.5	0.04	0.299
History	55.7	56.4	–0.7	–0.04	0.474
Science	60.4	60.1	0.4	0.02	0.698
Math	51.4	51.0	0.4	0.02	0.578
Sample size					
Students (total = 8,514)	4,888	3,626			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits attempted is scaled as a percentage of the number of credits (core or subject specific) required for graduation in a student's district.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Table N.3. Impact on cumulative credits earned as a percentage of credits attempted, school records sample (grade 9), year 1

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Credits earned in core subject areas (%)	77.3	79.0	–1.7	–0.06	0.556
English language arts	79.1	80.5	–1.4	–0.04	0.710
History	79.5	82.9	–3.4	–0.10	0.347
Science	77.5	76.5	1.0	0.03	0.765
Math	74.3	76.6	–2.3	–0.06	0.509
Sample size					
Students (total = 7,335)	4,237	3,098			
Schools (total = 28)	15	13			

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences. The analysis is based on students in the school records sample who attempted at least one credit in a core subject area.

a. The cumulative number of credits earned is divided by the number of credits attempted (core or subject specific).

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Table N.4. Impact on cumulative credits earned as a percentage of credits attempted, school records sample, year 2

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
Credits earned in core subject areas (%)	75.7	80.5	–4.8	–0.17	0.079
English language arts	78.2	84.5	–6.3	–0.20	0.046*
History	78.0	83.4	–5.4	–0.17	0.084
Science	74.5	77.8	–3.3	–0.09	0.348
Math	72.2	77.7	–5.5	–0.15	0.117
Sample size					
Students (total = 7,930)	4,463	3,467			
Schools (total = 28)	15	13			
Grade 10 sample					
Credits earned in core subject areas (%)	76.8	76.1	0.7	0.02	0.741
English language arts	79.4	79.0	0.5	0.01	0.811
History	79.3	78.7	0.5	0.02	0.848
Science	75.7	74.9	0.7	0.02	0.822
Math	74.0	73.2	0.7	0.02	0.717
Sample size					
Students (total = 8,500)	4,879	3,621			
Schools (total = 28)	15	13			

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences. The analysis is based on students in the school records sample who attempted at least one credit in a core subject area.

a. The cumulative number of credits earned is divided by the number of credits attempted (core or subject specific).

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Attendance rate

The estimated impact of CLC on student attendance in each year of the study is presented in tables N.5 and N.6. These findings are based on the subset of students in the school records sample for whom attendance data are available. The attendance rate is defined as the number of days attended during the school year as a percentage of the number of days enrolled.

As seen in the tables, CLC did not have an impact on student attendance. Although attendance was higher in CLC schools than non-CLC schools, this difference is not statistically significant. Thus it cannot be concluded that the CLC intervention improved student attendance, in either grade level.[195]

Table N.5. Impacts on attendance, school records sample, year 1 (grade 9)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Attendance rate	86.0	84.4	1.6	0.10	0.460
Days absent	22.9	25.2	−2.3	−0.09	0.504
Days present	142.5	138.8	3.7	0.12	0.347
Sample size					
Students (total = 7,331)	4,235	3,096			
Schools (total = 28)	15	13			

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences. The analysis is based on students in the school records sample for whom attendance data is available.

a. The attendance rate is defined relative to the number of days that the student was enrolled in the district.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

[195] Impacts on attendance do not differ by a statistically significant amount across study years (for grade 9 students) or across grade levels (in year 2).

Table N.6. Impacts on attendance, school records sample, year 2

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Grade 9 sample					
Attendance rate	87.9	87.0	0.9	0.07	0.481
Days absent	19.4	21.3	−1.9	−0.09	0.341
Days present	145.4	144.0	1.4	0.04	0.616
Sample size					
Students (total = 7,946)	4465	3481			
Schools (total = 28)	15	13			
Grade 10 sample					
Attendance rate	87.1	85.3	1.8	0.11	0.124
Days absent	21.0	23.7	−2.7	−0.10	0.182
Days present	143.1	139.9	3.2	0.10	0.086
Sample size					
Students (total = 8,263)	4,743	3,520			
Schools (total = 28)	15	13			

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Findings for grade 9 are based on students who are in grade 9 for the first time; findings for grade 10 are based on grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences. The analysis is based on students in the school records sample for whom attendance data is available.

a. The attendance rate is defined relative to the number of days that the student was enrolled in the district.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Appendix O. Impacts by Student Subgroup

This appendix presents the results of secondary analyses that investigate the impact of the CLC framework for different subgroups of students. As noted in chapter 6, impacts were estimated for student subgroups based on three predictors of reading achievement:

- **Reading proficiency.** Whether or not students had scored below the proficiency cutoff on the grade 8 reading/English language arts (ELA) assessment in their state.
- **Overage for grade.** Whether or not students were overage when entering grade 9.
- **Special education.** Whether or not students were classified as being eligible for special education services at the start of grade 9.

It was hypothesized that impacts would be greater for the subgroups most likely to be struggling readers—students who were not proficient on the grade 8 reading assessment, students who were overage at the start of grade 9, and special education students, because these students are likely to receive a more intensive dose of CLC services and supports.[196]

To minimize the number of analyses, the subgroup analyses focus on the primary outcomes in this report: reading comprehension and credit accumulation in the second year of the study. Estimated impacts on GRADE reading comprehension scores are presented in tables O.1–O.4; subgroup impacts on credit accumulation appear in tables O.5–O.10.[197] The top panel of each table presents impacts for the subgroup most likely to receive intensive CLC services (not proficient, overage, special education), while the lower panel presents the estimated impact for other students.

In general, findings from the analysis indicate that the CLC framework was not more effective for any particular group of students. While the estimated impact for one subgroup is statistically significant,[198] differences in impacts between subgroups are not statistically significant for any of the outcomes. Thus it cannot be concluded that CLC was more effective for students likely to receive more intensive CLC services (that is, lower performing, overage, or special education students).

[196] Students whose second language is English (English language learners or ELLs) also are more likely to be struggling readers and to receive intensive CLC supports. However, it was not possible to look at impacts for ELL students because, in some schools in the study, there were no ELLs.

[197] As explained in chapter 6, impacts on reading comprehension could not be examined by special education status because few special education students took the GRADE assessment. Students requiring special testing accommodations did not take the GRADE assessment. Thus, in some districts, the GRADE respondent sample does not include any students classified as English as a second language (ESL) or special education.

[198] The estimated impact of CLC on the credit accumulation of students who are proficient on the grade 8 reading assessment is negative and statistically significant.

Table O.1. Impact on reading comprehension in year 2 of the study, by students' proficiency level, GRADE respondent sample (grade 9)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Not proficient on grade 8 ELA state test[b]					
Reading comprehension					
Average standard score	83.8	83.2	0.57	0.04	0.479
Corresponding grade equivalent	*4.7*	*4.6*			
Corresponding percentile	*13*	*12*			
Sample size					
Students (total = 1,492)	850	642			
Schools (total = 28)	15	13			
Proficient on grade 8 ELA state test					
Reading comprehension					
Average standard score	98.2	98.0	0.18	0.01	0.970
Corresponding grade equivalent	*8.3*	*8.2*			
Corresponding percentile	*45*	*44*			
Sample size					
Students (total = 2,462)	1,473	989			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			0.39	0.02	0.509

* p-value $\leq .05$

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

b. Proficiency is based on the cut-off defined by the state.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table O.2. Impact on reading comprehension in year 2 of the study, by students' proficiency level, GRADE respondent sample (grade 10)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Not proficient on grade 8 ELA state test[b]					
Reading comprehension					
Average standard score	86.8	87.5	−0.71	−0.04	0.572
Corresponding grade equivalent	*5.3*	*5.5*			
Corresponding percentile	*18*	*20*			
Sample size					
Students (total = 1,039)	666	373			
Schools (total = 28)	15	13			
Proficient on grade 8 ELA state test					
Reading comprehension					
Average standard score	102.7	102.3	0.40	0.02	0.717
Corresponding grade equivalent	*10.4*	*10.3*			
Corresponding percentile	*55*	*54*			
Sample size					
Students (total = 2,308)	1,470	838			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			−1.11	−0.07	0.509

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the table are for grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

b. Proficiency is based on the cut-off defined by the state.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table O.3. Impact on reading comprehension in year 2 of the study, by overage status, GRADE respondent sample (grade 9)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Overage for grade when entered high school[b]					
Reading comprehension					
Average standard score	88.0	86.9	1.04	0.06	0.250
Corresponding grade equivalent	*5.6*	*5.3*			
Corresponding percentile	*21*	*18*			
Sample size					
Students (total = 1,240)	685	555			
Schools (total = 28)	15	13			
Not overage for grade when entered high school					
Reading comprehension					
Average standard score	93.8	93.1	0.69	0.04	0.749
Corresponding grade equivalent	*7.0*	*6.9*			
Corresponding percentile	*33*	*32*			
Sample size					
Students (total = 3,771)	2,290	1,481			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			0.35	0.02	0.509

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, E SL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

b. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table O.4. Impact on reading comprehension in year 2 of the study, by overage status, GRADE respondent sample (grade 10)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Overage for grade when entered high school[b]					
Reading comprehension					
Average standard score	92.0	90.6	1.46	0.09	0.633
Corresponding grade equivalent	*6.6*	*6.2*			
Corresponding percentile	*29*	*25*			
Sample size					
Students (total = 1,062)	629	433			
Schools (total = 28)	15	13			
Not overage for grade when entered high school					
Reading comprehension					
Average standard score	98.2	97.0	1.17	0.07	0.528
Corresponding grade equivalent	*8.3*	*7.8*			
Corresponding percentile	*45*	*41*			
Sample size					
Students (total = 3,484)	2,279	1,205			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			0.29	0.02	0.509

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the table are for grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Observations in districts that chose to test random samples of students are weighted to account for different sampling probabilities. Rounding may cause slight discrepancies in calculating sums and differences.

a. The national average for standard scores is 100, and its standard deviation is 15. The grade equivalent and percentile are those associated with the average standard score as indicated in the GRADE Teacher's Scoring and Interpretive Manual (Level H, Grade 9, Spring Testing, Form A). No statistical tests or arithmetic operations were performed on these reference points.

b. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

Source: Calculations from the GRADE assessment administered as part of the Content Literacy Continuum study. Level H of the assessment was administered in the spring of each study year (Form A to grade 9 students and Form B to grade 10 students).

Table O.5. Impact on credit accumulation in year 2 of the study, by students' proficiency level, school records sample (grade 9)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Not proficient on grade 8 ELA state test[b]					
Credits earned in core subject areas (%)	20.9	22.3	−1.41	−0.15	0.183
Sample size					
Students (total = 2,524)	1,435	1,089			
Schools (total = 28)	15	13			
Proficient on grade 8 ELA state test					
Credits earned in core subject areas (%)	24.7	26.5	−1.75	−0.18	0.033*
Sample size					
Students (total = 3,512)	1,947	1,565			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			0.33	0.04	0.790

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is divided by the number of credits attempted (core or subject specific).

b. Proficiency is based on the cut-off defined by the state.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Table O.6. Impact on credit accumulation in year 2 of the study, by students' proficiency level, school records sample (grade 10)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Not proficient on grade 8 ELA state test[b]					
Credits earned in core subject areas (%)	42.5	40.7	1.82	0.09	0.222
Sample size					
Students (total = 2,193)	1,264	929			
Schools (total = 28)	15	13			
Proficient on grade 8 ELA state test					
Credits earned in core subject areas (%)	49.3	50.8	–1.58	–0.08	0.366
Sample size					
Students (total = 3,428)	1,981	1,447			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			3.40	0.17	0.149

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the table are for grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is divided by the number of credits attempted (core or subject specific).

b. Proficiency is based on the cut-off defined by the state.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Table O.7. Impact on credit accumulation in year 2 of the study, by overage status, school records sample (grade 9)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Overage for grade when entered high school[b]					
Credits earned in core subject areas (%)	19.6	21.4	−1.72	−0.18	0.116
Sample size					
Students (total = 2,340)	1,305	1,035			
Schools (total = 28)	15	13			
Not overage for grade when entered high school					
Credits earned in core subject areas (%)	23.4	24.9	−1.47	−0.16	0.070
Sample size					
Students (total = 5,611)	3,162	2,449			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			−0.24	−0.03	0.848

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is divided by the number of credits attempted (core or subject specific).

b. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Table O.8. Impact on credit accumulation in year 2 of the study, by overage status, school records sample (grade 10)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Overage for grade when entered high school[b]					
Credits earned in core subject areas (%)	34.3	34.0	0.36	0.02	0.806
Sample size					
Students (total = 2,897)	1,591	1,306			
Schools (total = 28)	15	13			
Not overage for grade when entered high school					
Credits earned in core subject areas (%)	44.7	44.3	0.37	0.02	0.730
Sample size					
Students (total = 5,617)	3,297	2,320			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			-0.01	0.00	0.994

* p-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the table are for grade 10 students and retained grade 10 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, special education status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is divided by the number of credits attempted (core or subject-specific).

b. A student is defined as overage for grade if he or she turned 15 before the start of grade 9.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Table O.9. Impact on credit accumulation in year 2 of the study, by special education status, school records sample (grade 9)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Special education status prior to entering high school[b]					
Credits earned in core subject areas (%)	21.9	22.5	−0.55	−0.06	0.638
Sample size					
Students (total = 1,402)	764	638			
Schools (total = 28)	15	13			
Not special education status prior to entering high school					
Credits earned in core subject areas (%)	22.6	24.3	−1.70	−0.18	0.051
Sample size					
Students (total = 6,376)	3,612	2,764			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			1.15	0.12	0.418

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the tables are for students who are in grade 9 for the first time. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is divided by the number of credits attempted (core or subject-specific).

b. Special education status is based on students' status in the school year before they entered the study sample.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

Table O.10. Impact on credit accumulation in year 2 of the study, by special education status, school records sample (grade 10)

Outcome[a]	CLC group	Non-CLC group	Estimated impact	Effect size	*p*-value
Special education status when entered high school[b]					
Credits earned in core subject areas (%)	43.3	41.7	1.61	0.08	0.380
Sample size					
Students (total = 1,339)	708	631			
Schools (total = 28)	15	13			
Not special education status when entered high school					
Credits earned in core subject areas (%)	46.0	45.7	0.25	0.01	0.866
Sample size					
Students (total = 5,579)	3,292	2,287			
Schools (total = 28)	15	13			
Difference in estimated impact between subgroups			1.36	0.07	0.564

* *p*-value ≤ .05

Note: This table is based on the 28 schools that participated in both years of the study. Values in the table are for grade 10 students and retained grade 9 students. Impact estimates are regression-adjusted, controlling for the blocking of random assignment and for random baseline differences between students in the CLC and non-CLC schools with respect to the following variables: whether students were overage for grade at the start of grade 9, students' grade 8 state test scores in reading and math, eligibility for free or reduced-price lunch, ESL status, racial/ethnic group, and gender. Rounding may cause slight discrepancies in calculating sums and differences.

a. The cumulative number of credits earned is divided by the number of credits attempted (core or subject specific).

b. Special education status is based on students' status in the school year before they entered the study sample.

Source: Calculations based on school records data provided by school districts for the Content Literacy Continuum study.

References

ACT, Inc. (2009). The condition of college readiness 2009. Retrieved from http://www.act.org/research/policymakers/pdf/TheConditionofCollegeReadiness.pdf

Aladjem, D. K., LeFloch, K. C., Zhang, Y., Kurki, A., Boyle, A., Taylor, J. E., et al. (2006). *Models matter—The final report of the national longitudinal evaluation of comprehensive school reform* (ERIC Document Reproduction Service No. ED499198). Washington, DC: American Institutes for Research.

Allensworth, E. M., & Easton, J. Q. (2007). What matters for staying on-track and graduating in Chicago Public High Schools: A close look at course grades, failures, and attendance in the freshman year. Chicago: Consortium on Chicago School Research at the University of Chicago. Retrieved from http://ccsr.uchicago.edu/publications/07%20What%20Matters%20Final.pdf

Alvermann, D. E., Swafford, J., & Montero, M. K. (2004). *Content area literacy instruction for the elementary grades*. Boston: Allyn & Bacon.

American Guidance Service. (2001a). *Group reading assessment and diagnostic evaluation: Teacher's scoring and interpretive manual, level H.* Circle Pines, MN: Author.

American Guidance Service. (2001b). Group reading assessment and diagnostic evaluation: Technical manual. Circle Pines, MN: Author.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–291). Mahwah, NJ: Erlbaum.

Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis. Which high schools produce the nation's dropouts? Where are they located? Who attends them?* (CRESPAR Report 70). Baltimore, MD: Center for Research on the Education of Students Placed at Risk.

Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior, 16*(4), 297–306.

Beck, I., McKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 147–163). Hillsdale, NJ: Erlbaum.

Biancarosa, C., & Snow, C. E. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education. Retrieved from

http://www.all4ed.org/files/archive/publications/ReadingNext/ReadingNext.pdf

Billmeyer, R. (1996). Teaching reading in the content areas: If not me, then who? Teacher's manual. Longmont, CO: Sopris West.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments* (pp. 115–172). New York: Russell Sage.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59.

Borman, G. D., Hewes, G. M, Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125–230.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (2nd ed.). Washington, DC: National Academies Press.

Brown, A. L., Palinscar, A. S., & Armbruster, B. B. (2004). Instructing comprehension-fostering activities in interactive learning situations. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 780–809). Newark, DE: International Reading Association.

Clark, D. R. (2007). Learning through reflection (website). Retrieved from http://www.nwlink.com/~donclark/learning/reflecting.html

Common Core State Standards Initiative. (2010). Common core state standards. Retrieved from http://www.corestandards.org/the-standards

Cramer, R. L. (1970). Setting purposes and making predictions: Essential to critical reading. *Journal of Reading, 13*(4), 259–262, 300.

Deshler, D. (2010, December). *Virginia CLC project: 6 years of increasing literacy and learning 2005–2011*. Presentation given to the Virginia Department of Education, Richmond, VA.

Deshler, D. D., & Schumaker, J. B. (2006). *Teaching adolescents with disabilities: Accessing the general education curriculum.* New York: Corwin Press.

Desimone, L. (2000). *Making comprehensive school reform work* (Urban Diversity Series No. 112) (ERIC Document Reproduction Service No.1637). East Lansing, MI: Clearinghouse on Urban Education, Institute for Urban and Minority Education.

Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72*(3), 433–479.

Dewitz, P., Carr, E. M., & Patberg, J. P. (1987). Effects of inference training on comprehension and compehension monitoring. *Reading Research Quarterly, 22*(1), 99–122.

Dole, J. A., Brown, K. J., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly, 31*(1), 62–88.

Duke, N. K., & Pearson, P. D., (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. Samuels (Eds.), *What research has to say about reading instruction* (pp. 205–242). Newark, DE: International Reading Association.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco: Jossey-Bass.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A. Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008–4030). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://www.mdrc.org/publications/499/full.pdf

Gee, J. P. (2000). Discourse and sociocultural studies in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 195–207). Mahwah, NJ: Erlbaum.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/rti_math_pg_042109.pdf

Goldman, S. R., & Rakestraw, J. A. (2000). Structural aspects of constructing meaning from text. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311–335). Mahwah, NJ: Erlbaum.

Graves, M. F., & Watts-Taffe, S. M. (2002). The place of word-consciousness in a research-based vocabulary program. In A. E. Farstrup & S. Samuels (Eds.), *What research has to say about reading instruction* (pp. 140–165). Newark, DE: International Reading Association.

Greenleaf, C., & Schoenbach, R. (2004). Building capacity for responsive teaching of reading in the academic disciplines: Strategy inquiry designs for middle and high school teachers' professional development. In D. Strickland & M. L. Kamil (Eds.), *Improving reading achievement through professional development* (pp. 97–127). Norwood, MA: Christopher-Gordon.

Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, *60*(5), 6–11.

Guthrie, J. T., Van Meter, P., McCann, A., Wigfield, A., Bennett, L., Poundstone, C. C., et al. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly, 31*(3), 306–332.

Hacker, D. J. (2004). Self-regulated comprehension during normal reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 755–779). Newark, DE: International Reading Association.

Harmon, J. (2002). Teaching independent word learning strategies to struggling readers. *Journal of Adolescent and Adult Literacy*, *45*(7), 606–615.

Harmon, J., Hedrick, W. B., & Fox, E. A. (2000). A content analysis of vocabulary instruction in social studies textbooks for grades 4–8. *The Elementary School Journal*, *100*(3), 253–271.

Harvey, S., & Goudvis, A. (2000). *Strategies that work: Teaching comprehension to enhance understanding.* Portland, ME: Stenhouse.

Heller, R., & Greenleaf, C. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement.* Washington, DC: Alliance for Excellent Education.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177.

Hochberg, Y., & Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine*, *9*, 811–818.

Hock, M. F., Brasseur, I. F., & Deshler, D. D. (2008). *Fusion Reading.* Lawrence, KS: University of Kansas Center for Research on Learning.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE 2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U. S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/adlit_pg_082608.pdf

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 1270–1328). Newark, DE: International Reading Association.

Lenz, B. K., Ehren, B. J., & Deshler, D. D. (2005). *The content literacy continuum: A school reform framework for improving adolescent literacy for all students.* Lawrence: University of Kansas Center for Research on Learning.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

National Association of Secondary School Principals. (2005). *Creating a culture of literacy: A guide for middle and high school principals.* Reston, VA: Author.

National Center for Education Statistics. (2009). *The nation's report card: Reading 2009* (NCES 2010–458). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Center for Education Statistics. (2010). *The nation's report card: Grade 12 reading and mathematics 2009 national and pilot state results* (NCES 2011–455). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

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: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2011). *Gates-MacGinitie Reading Tests® (GMRT®) Fourth Edition Forms S and T* - Paper-Pencil*. Rolling Meadows, IL: Riverside Publishing. Retrieved from http://www.riversidepublishing.com/products/gmrt/details.html

NCAA Eligibility Center. (2009). *NCAA freshman-eligibility standards quick reference sheet.* Indianapolis, IN: Author. Retrieved from http://www.eligibilitycenter.org/ECWR2/NCAA_EMS/pdf/Quick_Reference_Sheet_for_IE_Standards-6-18-09.pdf

Nelson, G. D. (2001). Choosing content that's worth knowing. *Educational Leadership, 59*(2), 12–16.

Nystrand, M., & Gamoran, A. (1988). *A study of instruction as discourse.* Madison: Wisconsin Center for Education Research.

Pearson, P. D., & Camperell, K. (1994). Comprehension of text structures. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 448–468). Newark, DE: International Reading Association.

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup & S. Samuels (Eds.), *What research has to say about reading instruction* (pp. 291–309). Newark, DE: International Reading Association.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Raphael, T. E., & Pearson, P. D. (1985). Increasing students' awareness of sources of information for answering questions. *American Educational Research Journal*, *22*(2), 217–235.

Ruddell, R. B., & Unrau, N. J. (2004). Reading as a meaning-construction process: The reader, the text, and the teacher. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 1462–1521). Newark, DE: International Reading Association.

Schochet, P. Z. (2008). *Technical methods report*: *Guidelines for multiple testing in impact evaluations* (NCEE 2008–4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pdf/20084018.pdf

Schumaker, J. B., & Deshler, D. D. (2003). Designs for applied educational research. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 482–495). New York: Guilford.

Slavin, R., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly, 43*, 290–322.

Somers, M.-A., Corrin, W., Sepanik, S., Salinger T., Levin, J., & Zmach, C. (2010). *The enhanced reading opportunities study final report: The impact of supplemental literacy courses for struggling ninth-grade readers* (NCEE 2010-4021). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

University of Kansas Center for Research on Learning. (2008). *Adolescent literacy: Addressing the challenge: Learning strategies.* Lawrence, KS: Author.

U.S. Department of Education Office of Elementary and Secondary Education. (2010). *Summary of 2006 Striving Readers projects.* Washington, DC: Author. Retrieved from http://www2.ed.gov/programs/strivingreaders/crosssumtables.pdf

Vacca, R. T. (2002). Making a difference in adolescents' school lives: Visible and invisible aspects of content area reading. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 184–204). Newark, DE: International Reading Association.

Weaver, C. A., & Kintsch, W. (1991). Expository text. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 230–245). Mahwah, NJ: Erlbaum.

Wigfield, A. E. (1997). Children's motivations for reading and reading engagement. In J. T. Guthrie & A. E. Wigfield (Eds.), *Reading engagement: Motivating readers through integrated instruction* (pp. 14–33). Newark, DE: International Reading Association.


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA