Appendix

Final Report of the i3 Evaluation of the Collaboration and Reflection to Enhance Atlanta Teacher Effectiveness (CREATE) Teacher Residency Program

A QUASI-EXPERIMENT IN GEORGIA

February 2021

Andrew P. Jaciw

Audra Wingard

Jenna Zacamy

Li Lin

Sze-Shun Lau

Empirical Education Inc.

Table of Contents

Reference this report: Jaciw, A. P., Wingard, A., Zacamy, J., Lin, L., & Lau, S. (2021). *Final Report of the i3 Evaluation of the Collaboration and Reflection to Enhance Atlanta Teacher Effectiveness (CREATE) Teacher Residency Program: A Quasi-Experiment in Georgia*. (Empirical Education Rep. No. Empirical_GSU-7031-FR1-2021-O.1). Empirical Education Inc. https://www.empiricaleducation.com/create/

Appendix A. Cohort 3, Year 1 Preliminary Findings

INTRODUCTION

Cohort 3 will continue, along with Cohorts 4 through 8, with funding from the Supporting Effective Educator Development (SEED) grant program. Preliminary results from Cohort 3 Year 1 are reported here, as this year was funded through the Investing in Innovation (i3) grant. Unless otherwise stated, all details about the recruitment process and study design for Cohort 3 mirror that of Cohorts 1 and 2 (presented in Chapter 2 of this report).

KEY RESEARCH QUESTIONS

The implementation evaluation investigates the following questions.

1. Were the key components of the CREATE logic model implemented with fidelity?
2. How does the experience of study participants in the comparison group compare to that of the CREATE group, specifically with regard to professional development, mentorship, and collaboration?

Additionally, we conducted a descriptive analysis for Cohort 3 in Year 1 to address the following exploratory research question:

- How does CREATE preservice teacher performance, as measured by edTPA scores, compare with the business-as-usual condition, during their student teacher year?

PARTICIPANT RECRUITMENT

Researchers held recruitment events for Cohort 3 at GSU CEHD in Summer 2017. Initial recruitment efforts resulted in 14 CREATE residents and 41 comparison study participants agreeing to be part of the research. According to our survey data, CREATE residents commenced their practicums in one of eight CREATE schools in APS. Comparison group study participants completed their practicum in 55 different schools across 10 districts over the course of the year. Superintendents and principals from schools and districts of comparison study participants received an informational letter about the research study or a research application (depending on the requirements of the district), allowing them to opt out of participation.

SCHEDULE OF MAJOR MILESTONES

Table A1 lists the major milestones for Cohort 3 during the 2017–18 school year.

TABLE A1. RESEARCH MILESTONES FOR COHORT 3 DURING THE 2017–18 SCHOOL YEAR

Date	Milestone
Summer 2017	Recruited Cohort 3 for research study; participants submitted signed consent forms
Summer–Fall 2017	Deployed baseline survey to eligible Cohort 3 study participants on an ongoing basis as researchers received their consent
November 2017	Sent informational letters or requests for principal permission and/or research applications (where required) to schools/districts of active study participants in Cohort 3

TABLE A1. RESEARCH MILESTONES FOR COHORT 3 DURING THE 2017–18 SCHOOL YEAR

Date	Milestone
November 2017	Deployed first quarterly survey to Cohort 3
November 2017	Submitted first quarterly report
January 2018	Deployed second quarterly survey to Cohort 3
February 2018	Submitted second quarterly report
March 2018	Deployed third quarterly survey to Cohort 3
April 2018	Deployed final quarterly survey to Cohort 3
May 2018	Submitted third quarterly report

COHORT 3 YEAR 1 FIDELITY OF IMPLEMENTATION RESULTS

As we did for Cohorts 1 and 2, we assessed implementation fidelity for Cohort 3 in their first year of CREATE for the following key components: (1) progressive core classroom roles, (2) Critical Friendship (CF) work, and (3) Cognitively-Based Compassion Training (CBCT®) during Year 1.[1] Explanations for each of the components are in the Chapter 2 of the report where FOI is reported for Cohorts 1 and 2, as well as in Appendix D.

We present results that show which indicators within the key program components followed implemented faithfully during the first year of CREATE programming for Cohort 3.

Key Component 1: Progressive Core Classroom Roles

We present data on the following indicators related to progressive core classroom roles.

- Indicator 1: Resident is paired with another Year 1 resident for fall semester (same building)
- Indicator 2: Resident is placed in classroom of veteran educator trained in mentoring
- Indicator 3: Resident is placed in a CREATE school

CREATE administrators expressed that they intended to have 95% or more of residents meet fidelity on at least two of three indicators listed above in Year 1.

The records from CREATE attendance rosters and survey data show that 93% (13/14) met fidelity for 2+ indicators. Therefore, the fidelity threshold at the program level for Component 1 in Year 1 was not met (Table A2).

[1] Component 4 (multiple forms of mentoring) is not reported here. As CREATE programming evolved to meet residents' needs, Cohort 3 residents no longer worked with a School-Based Mentor in year 1 of the residency. Component 5 will be reported with the Cohort 3 Year 2 analysis under the SEED funding.

TABLE A2. DETAILS OF COMPONENT 1 IN YEAR 1

Indicator	Fidelity threshold	Cohort 3 met fidelity?
Indicator 1: **Resident is paired with another Y1 resident for fall semester (same building)**	0 = Resident is not paired with another Y1 resident 1 = Resident is paired with another Y1 resident for part of the semester 2 = Resident is paired with another Y1 resident for whole semester	12/14 (86%)
Indicator 2: **Resident is placed in classroom of veteran educator trained in mentoring**	0 = Resident is not placed with a veteran educator trained in mentoring 1 = Resident is placed with a veteran educator trained in mentoring	12/14 (86%)
Indicator 3: **Resident is placed in a CREATE school**	0 = Resident is placed in a non-CREATE school 1 = Resident is placed in a CREATE school	14/14 (100%)
Program level	**95% or more of residents meet fidelity on 2+ indicators**	**13/14 (93%)** **Fidelity not met**

Key Component 2: Critical Friendship

We present data on the following Year 1 indicators related to CF.

- Indicator 1: CREATE administrators host 2 or more options for veteran teachers to attend 4-day CF Institute each year
- Indicator 2: Veteran educators attend 4-day CF Institute
- Indicator 4: CREATE residents attend monthly CF meetings.

CREATE planned to host at least two CF Institutes for veteran educators during the 2017–18 school year with 85% or more of the educators who sign up attending 3–4 days. They exceeded this goal and offered four institutes during which 92% of educators attended 3–4 days. Additionally, 100% of CREATE residents met indicator 4 that says residents should attend 8–10 monthly CF meetings. Because CF and CBCT meetings are combined, indicator 4 in component 3 is identical to this one.

Cohort 3, Year 1 meets fidelity for this Component 2 in Year 1 (Table A3).

TABLE A3. DETAILS OF COMPONENT 2 IN YEAR 1

Indicator	Fidelity threshold	2017–18 school year
Indicator 1: **CREATE administrators host 2 or more options for veteran teachers to attend 4-day CF Institute each year**	0 = CREATE administrators host 0-1 institute	CREATE administrators host 4 institutes
	1= CREATE administrators host 2+ institutes	Fidelity met
Indicator 2: **Veteran teachers attend 4-day CF Institute**	0 = Educator attends 1-2 days	
	1 = Educator attends 3-4 days	71/77 (92%)
	Of the veteran educators who attend, 85% or more meet fidelity at the unit level	Fidelity met
Indicator	**Fidelity threshold**	**Cohort 3 met fidelity?**
Indicator 4: **CREATE residents attend monthly CF meetings**	0 = Attend 0-7 meetings	
	1 = Attend 8-10 meetings	14/14 (100%)
	95% or more of residents meet fidelity at the unit level	Fidelity met
Program level	**0 = Fidelity not met for any indicator, or fidelity met for Indicator 1 but no other indicator** **1 = Fidelity was met for Indicator 1 and at least one other indicator**	**Fidelity met**

Key Component 3: Cognitively-Based Compassion CBCT Training

We present data on the following indicators related to CBCT.

- Indicator 1: Program administrators offer at least one CBCT per year for general population of teachers at CREATE schools
- Indicator 2: Residents attend CBCT classes

CREATE planned to host at least one CBCT institutes veteran educators during the 2017–18 school year. They exceeded this goal and offered 3 CBCT institutes.

Cohort 3, Year 1 meets fidelity for Component 3 in Year 1 (Table A4).

TABLE A4. DETAILS OF COMPONENT 3 IN YEAR 1

Indicator	Fidelity threshold	2017–18 school year
Indicator 1. **Program administrators offer at least one Cognitively-Based Compassion Training (CBCT®) per year for general population of teachers at CREATE schools**	0 = CREATE administrators offer 0 courses	CREATE administrators host 3 institutes
	1 = CREATE administrators offer 1 or more courses	Fidelity met
Indicator	**Fidelity threshold**	**Cohort 3 met fidelity?**
Indicator 2. **Residents attend CBCT classes**	0 = Attend 0-7 meetings	14/14 (100%)
	1 = Attend 8-10 meetings	
	95% or more of residents meet fidelity at the unit level	Fidelity met
Program level	**Component score is sum of indicator scores.** **Fidelity not met = score of 0-1** **Fidelity met = score of 2**	**Fidelity met**

DESCRIPTIVE FINDINGS RELATED TO SUPPORT, PERCEIVED SUCCESS IN TEACHING, AND PARTICIPATION IN CREATE PROFESSIONAL LEARNING

In this section, we present descriptive findings from survey data for the first year of Cohort 3 where study participants were asked to respond to questions about how supported they feel at their schools (Table A1), how successful they feel in a variety of professional areas (Table A2), and their level of participation in Together Time meetings (Table A3). Overall, Cohort 3 participants felt supported (with 92% reporting feeling more than moderately or very supported) and successful in teaching (with the lowest level of success reported in balancing work and personal life), and attended Together Time meeting as expected.

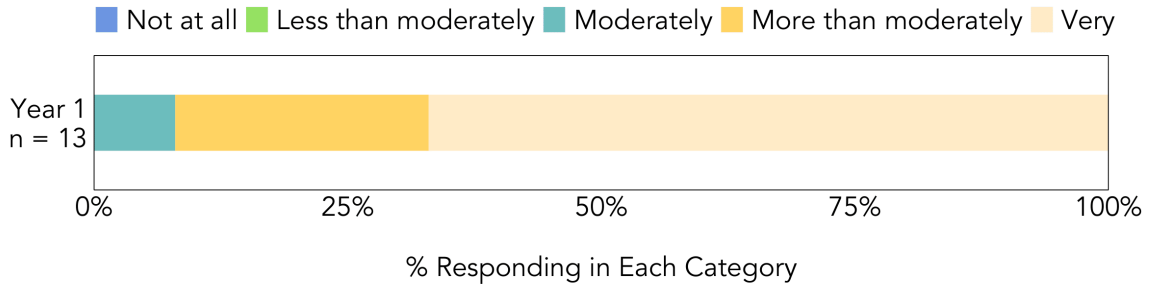

FIGURE A1. LEVELS OF SUPPORT FOR TEACHING

Source: Quarterly survey 4

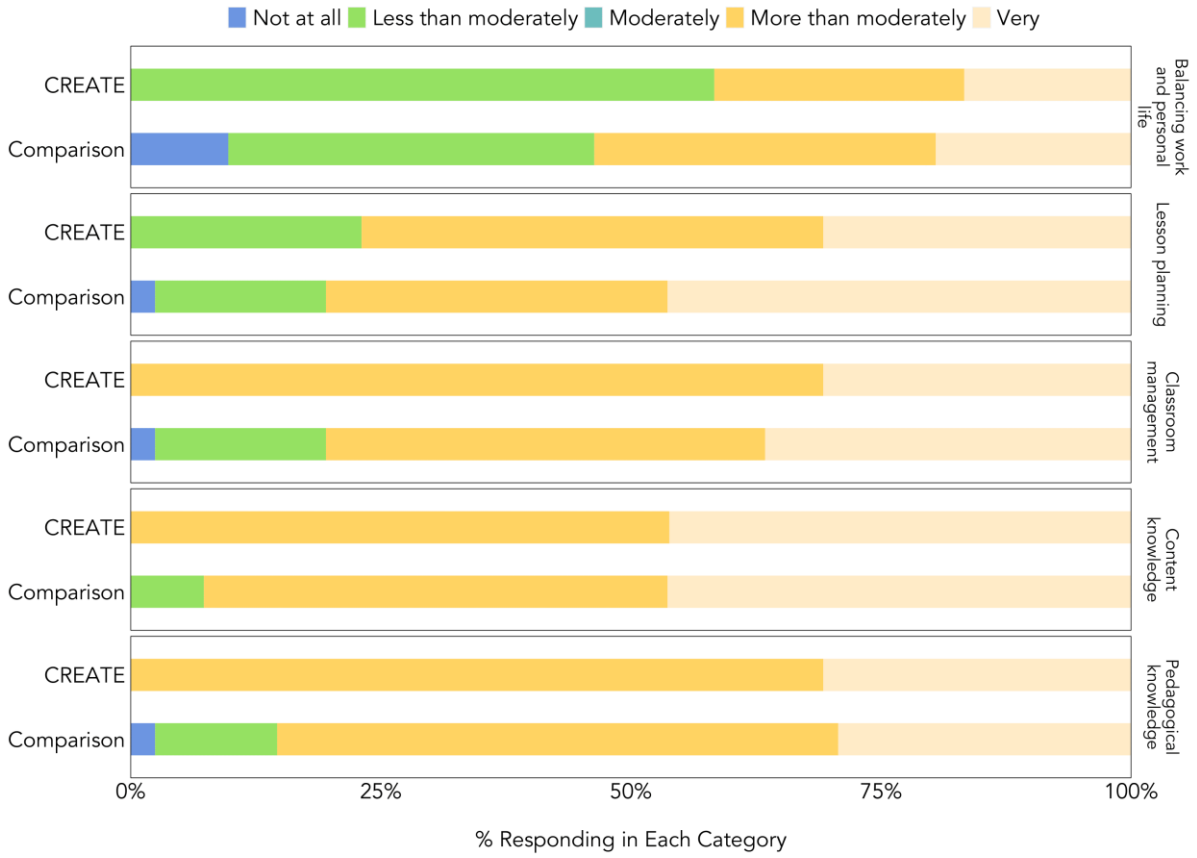

FIGURE A2. LEVELS OF SUCCESS IN VARIOUS ASPECTS OF TEACHING FOR COHORT 3 IN YEAR 1

Note. *N* = 12–13 for the CREATE group; *N* = 41 for the comparison group, depending on the category

Source: Quarterly survey 4

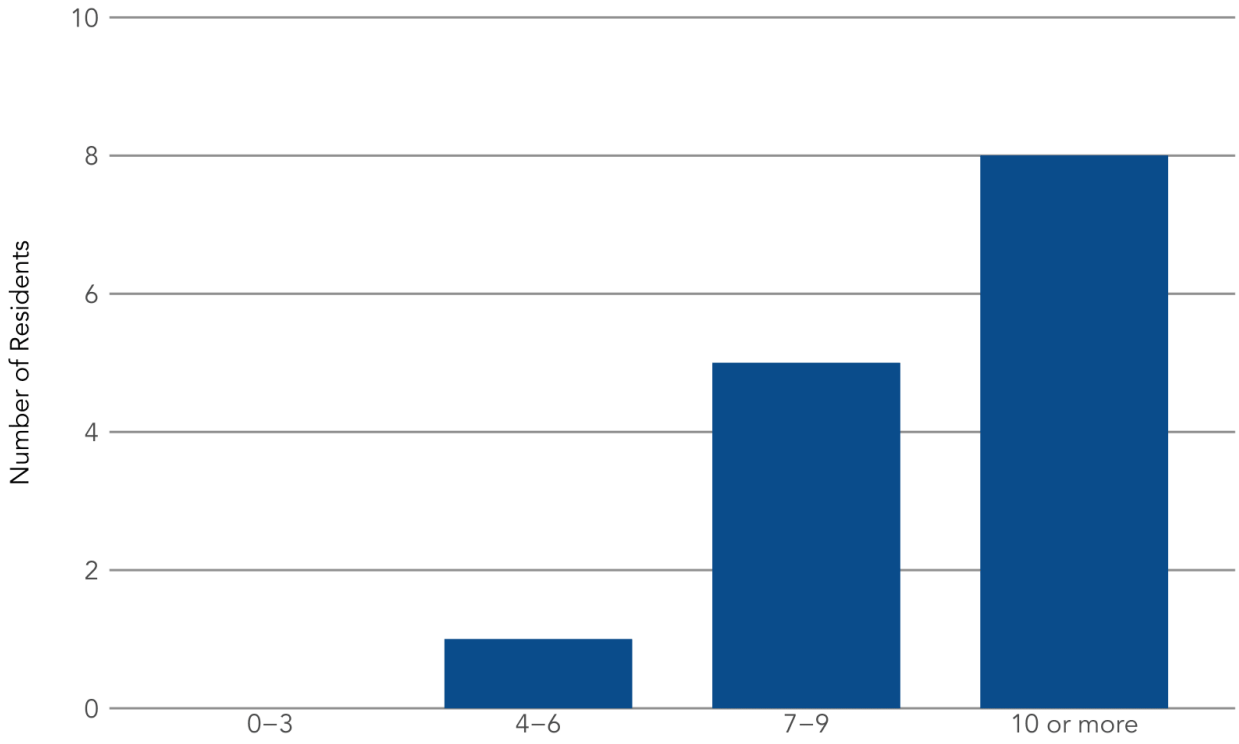

FIGURE A3. ATTENDANCE AT TOGETHER TIME MEETINGS FOR COHORT 3 CREATE RESIDENTS IN YEAR 1

Note. CREATE's expectation is that residents should attend between 8–10 meetings in Year 1.

N = 14

Source: Quarterly surveys and CREATE program rosters

COMPARISON OF EDTPA OUTCOMES

In this section, we present descriptive findings for how study participants performed on the edTPA assessment. edTPA, as described in Chapter 2 of this report, is a performance-based assessment for teacher candidates. During the 2017-18 school year, the state of Georgia requires that teacher candidates earn a passing score on the edTPA assessment in order to become certified teachers (GaPSC, n.d.).

The edTPA assessment includes 15 rubrics that are each scored on a scale of 1–5 points. The assignment of each score considers all required commentary responses and materials—which may include lesson plans, assessments, or video recordings—for each task. A score of 1 or 2 indicates the participant did not adequately complete some or all of the key requirements in a rubric. At a score of 1 or 2, a teacher candidate may have presented only vague connections to skills and conventions essential to student learning, and provided supports that are misaligned with the needs of the specific classroom or the focus of learning material. A score of 3 indicates the participant met the key requirements of the rubric, which include evidence of consideration of the needs of individual students and making explicit connections between

skills and academic content. A score of 4 or 5 means the participant exceeded the key requirements and met most or all of the goals listed in the rubric. A teacher candidate who earns a score of 4 or 5 will have demonstrated a clear central focus in their lessons, an ability to provide scaffolded instructional supports, and create a positive learning environment that challenges student thinking and encourages respect. At this level, a teacher candidate may show that they considered educational research and theory when completing tasks. The 15 rubrics are organized within three larger categories, called tasks. Rubrics 1–5 are categorized under Task 1: Planning. Rubrics 6–10 constitute Task 2: Instruction. Rubrics 11–15 fall under Task 3: Assessment. The titles of each individual rubric are listed in Table A7. During the 2017–18 school year, teacher candidates in the state of Georgia were required to earn a total score of 38 or higher on the 15 rubrics (edTPA, 2020).[2]

The following tables show descriptive statistics that compare the results of CREATE residents and comparison study participants' performance on the edTPA. One hundred percent of study participants in both the CREATE group and the comparison group received a passing score. Furthermore, as you can see from these findings, there is little difference between the mean scores of the CREATE and comparison groups (45.27 and 46.10, respectively), so we concluded there was no need to conduct an inferential test.

TABLE A6. AVERAGE EDTPA TOTAL SCORE AND PASS RATE

Rubric	CREATE (*N* = 13)	Comparison (*N* = 59)
Total Score	45.27 (3.78)	46.10 (4.86)
Pass Rate	100%	100%

Note. Standard deviation is in parentheses.

TABLE A7. MEANS OF EDTPA RUBRIC SCORES

Rubric	CREATE (*N* = 13)	Comparison (*N* = 59)
Rubric 1: Planning for Mathematical Understandings	3.12 (0.51)	3.15 (0.56)
Rubric 2: Planning to Support Varied Student Learning Needs	3.00 (0.91)	3.03 (0.67)
Rubric 3: Using Knowledge of Students to Inform Teaching and Learning	3.31 (0.48)	3.21 (0.64)

[2] In the 2015–16 and 2016–17 school years, teacher candidates in the state of Georgia were required to earn a total score of 35 or higher on the 15 rubrics. In September 2017, Georgia increased the passing score for the 15 rubrics to 38. The total possible score is 75. Cohort 1 study participants took the edTPA in spring 2016, Cohort 2 study participants took the edTPA in spring 2017, and Cohort 3 study participants took the edTPA in spring 2018.

TABLE A7. MEANS OF EDTPA RUBRIC SCORES

Rubric	CREATE (*N* = 13)	Comparison (*N* = 59)
Rubric 4: Identifying and Supporting Language Demands	3.31 (0.63)	2.88 (0.69)
Rubric 5: Planning Assessments to Monitor and Support Student Learning	3.23 (0.60)	3.16 (0.64)
Rubric 6: Learning Environment	3.00 (0)	3.09 (0.29)
Rubric 7: Engaging Students in Learning	3.04 (0.32)	2.93 (0.49)
Rubric 8: Deepening Student Learning	2.77 (0.60)	2.81 (0.63)
Rubric 9: Subject-Specific Pedagogy: Using Representations	2.85 (0.69)	2.68 (0.64)
Rubric 10: Analyzing Teaching Effectiveness	2.77 (0.60)	2.66 (0.64)
Rubric 11: Analysis of Student Learning	2.88 (0.58)	3.32 (0.59)
Rubric 12: Providing Feedback to Guide Learning	3.23 (0.60)	3.38 (0.55)
Rubric 13: Student Use of Feedback	2.85 (0.69)	2.82 (0.63)
Rubric 14: Analyzing Students' Language Use and Mathematics Learning	2.92 (0.49)	2.85 (0.78)
Rubric 15: Using Assessment to Inform Instruction	3.00 (0.41)	3.03 (0.67)
Average Rubric Score	3.02 (0.25)	3.00 (0.33)

Note. Standard deviation is in parentheses.

Figure A4 below shows the average score each participant in the CREATE and comparison group received on the rubrics.

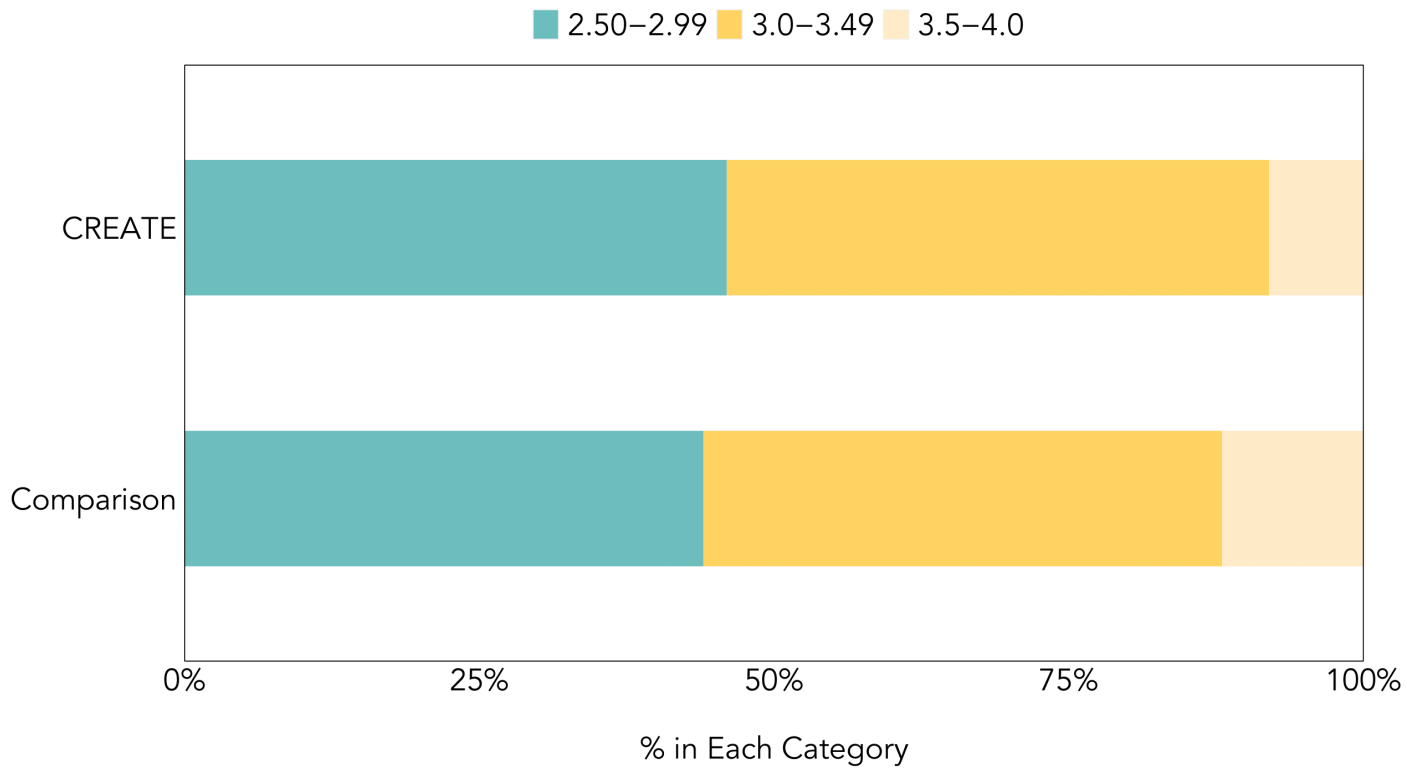

FIGURE A4. MEAN OF ALL RUBRIC SCORES, CREATE AND COMPARISON

Note. *N* (CREATE) = 13 and *N* (Comparison) = 59

No teacher received an average rating in the range of 0 – 2.49.

The following figures show the distribution of scores on each of the rubrics by group assignment (i.e., CREATE vs comparison group). The first five rubrics fall under Task 1—Planning for Instruction and Assessment—for which teacher candidates must plan a series of lessons in which the candidate aligns standards, learning objectives, learning tasks, and assessments based on knowledge of their students. As you can see in the following two figures, most participants in both groups scored a 3 or higher on all five rubrics, with a slightly higher percentage of CREATE residents scoring 3 or higher than in the comparison group. There is no 1.5 or 4.5 in the legend of the following figures because no participant received those scores on any of the rubrics.

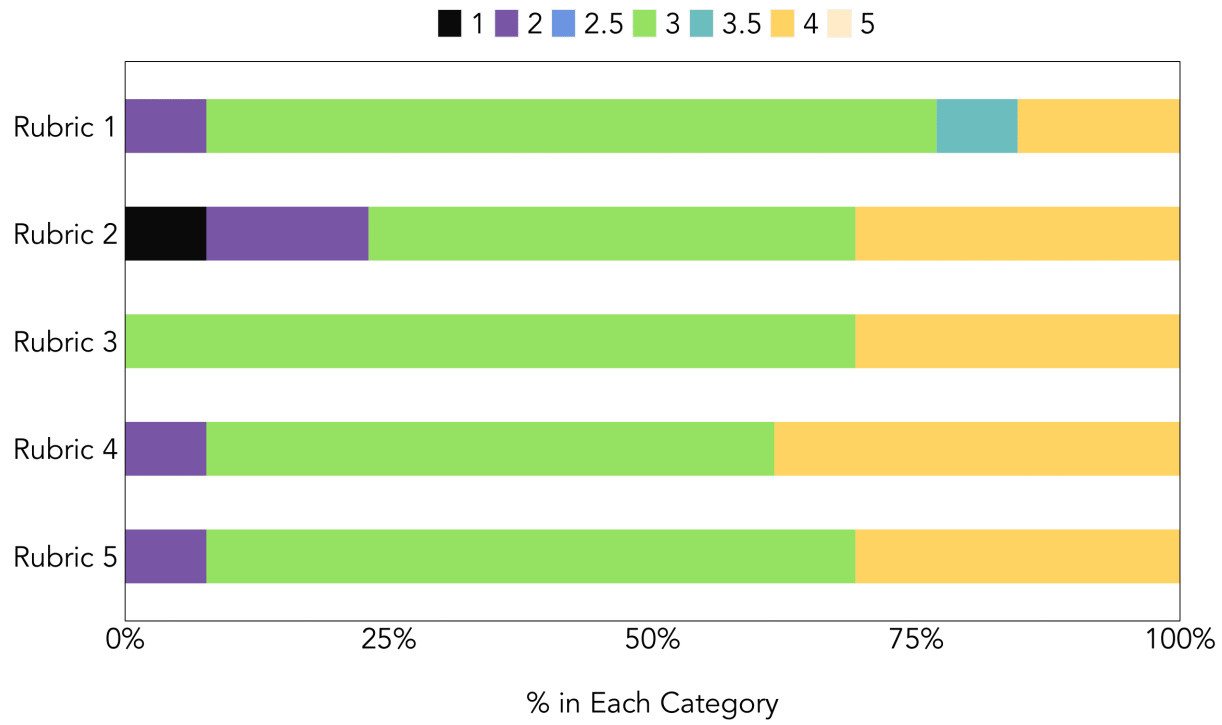

FIGURE A5. TASK 1, RUBRICS 1-5: CREATE RESIDENTS

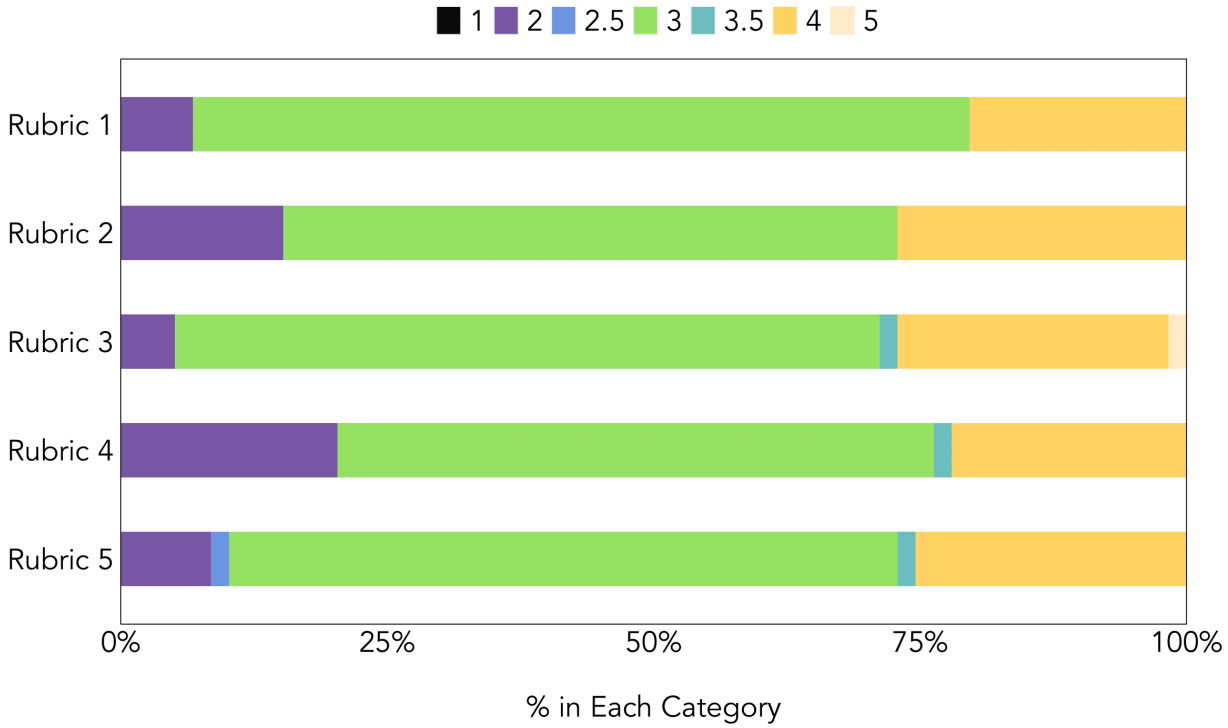

FIGURE A6. TASK 1, RUBRICS 1-5: COMPARISON GROUP

The next two figures show the distribution of scores on Rubrics 6–10, which fall under Task 2: Instructing and Engaging Students in Learning. To complete this task, teacher candidates must record themselves implementing their lessons. There is an evaluation on their ability to utilize specific pedagogical strategies, facilitate student responses, and deepen students' understanding of the main topic. As the figures show, most participants received a score of 3 or higher, and there is not a significant difference between CREATE and comparison participants' scores.

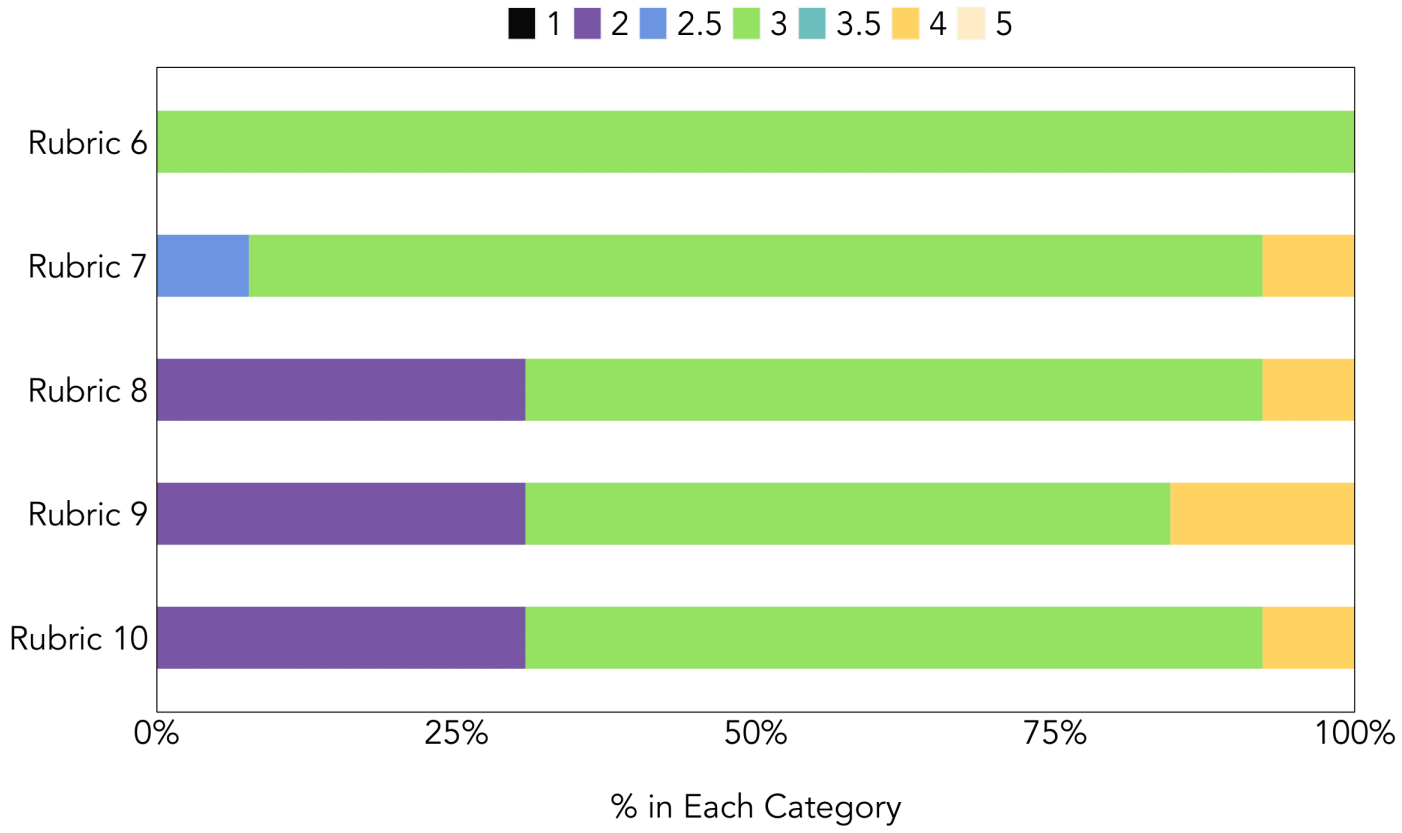

FIGURE A7. TASK 2, RUBRICS 6-10: CREATE RESIDENTS

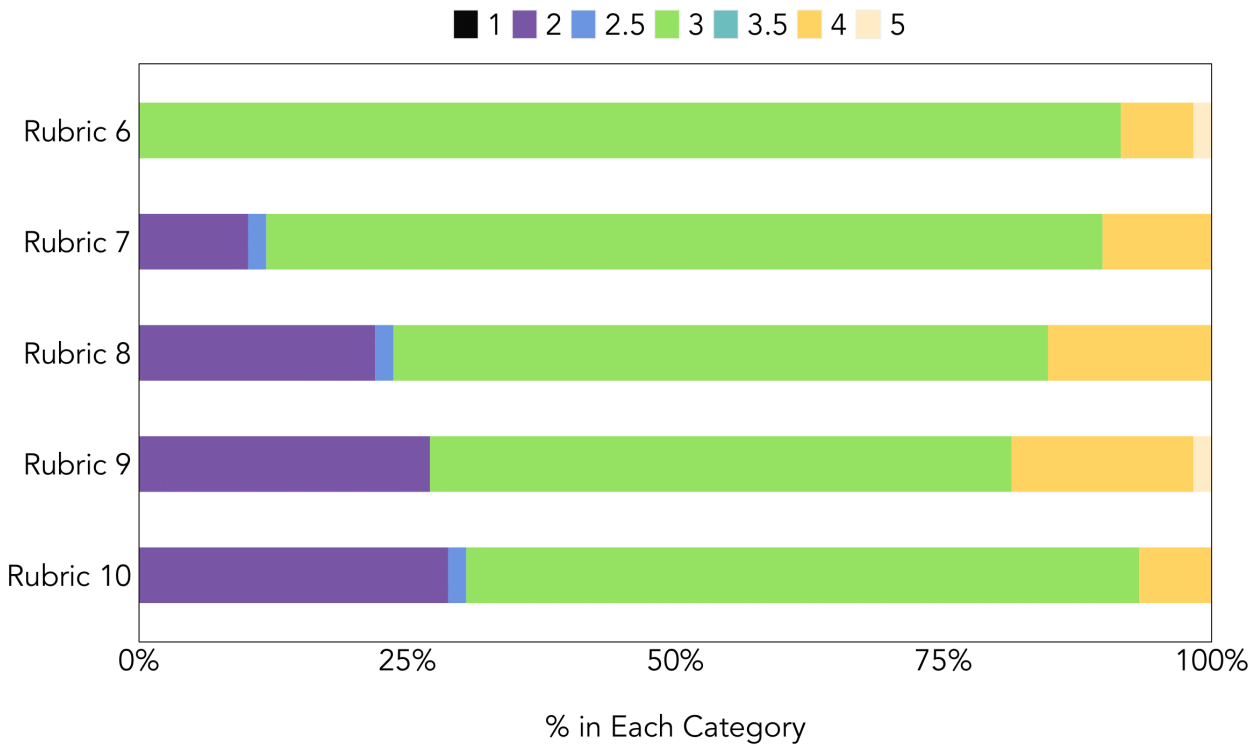

FIGURE A8. TASK 2, RUBRICS 6-10: COMPARISON GROUP

The final two figures show the distribution of scores on Rubrics 11–15, which are categorized within Task 3: Assessing Student Learning. Teacher candidates must compile and analyze assessment data from their students to demonstrate student progress toward the learning objectives. As with Tasks 1 and 2, most participants received a score of 3 or higher, and there is not a significant difference between CREATE residents' and comparison participants' scores.

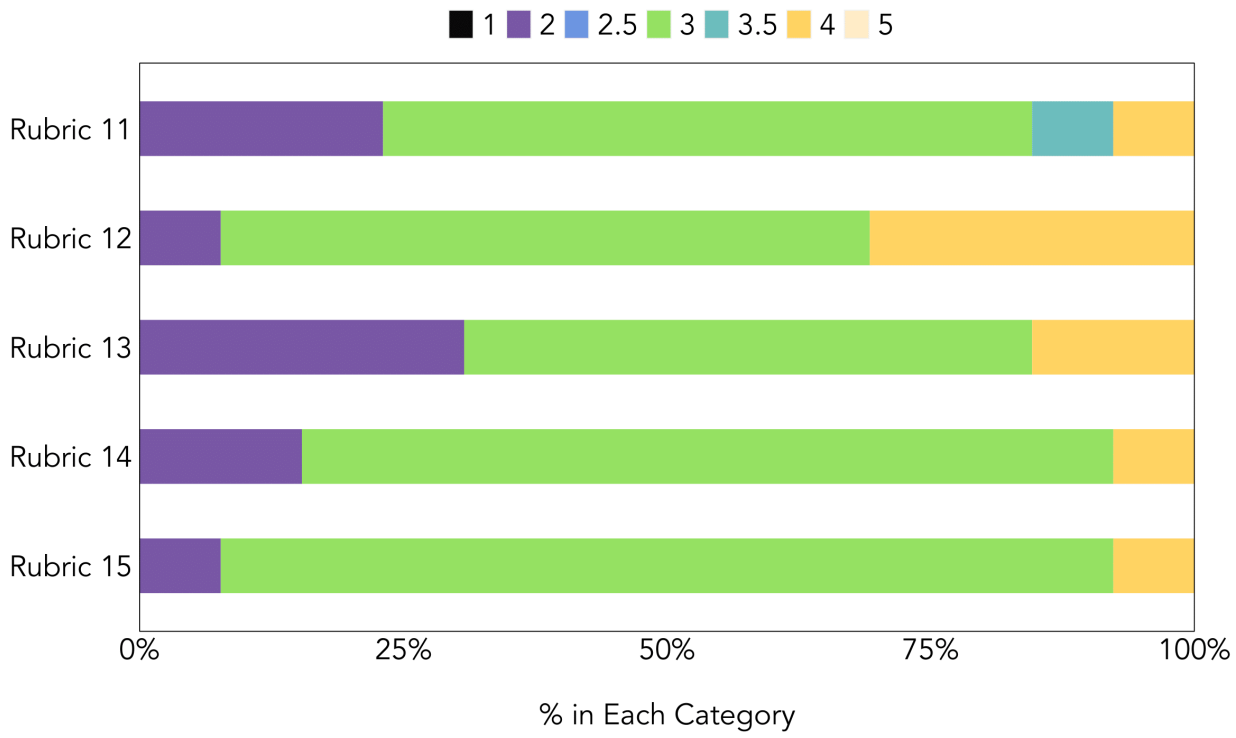

FIGURE A9. TASK 3, RUBRICS 11-15: CREATE RESIDENTS

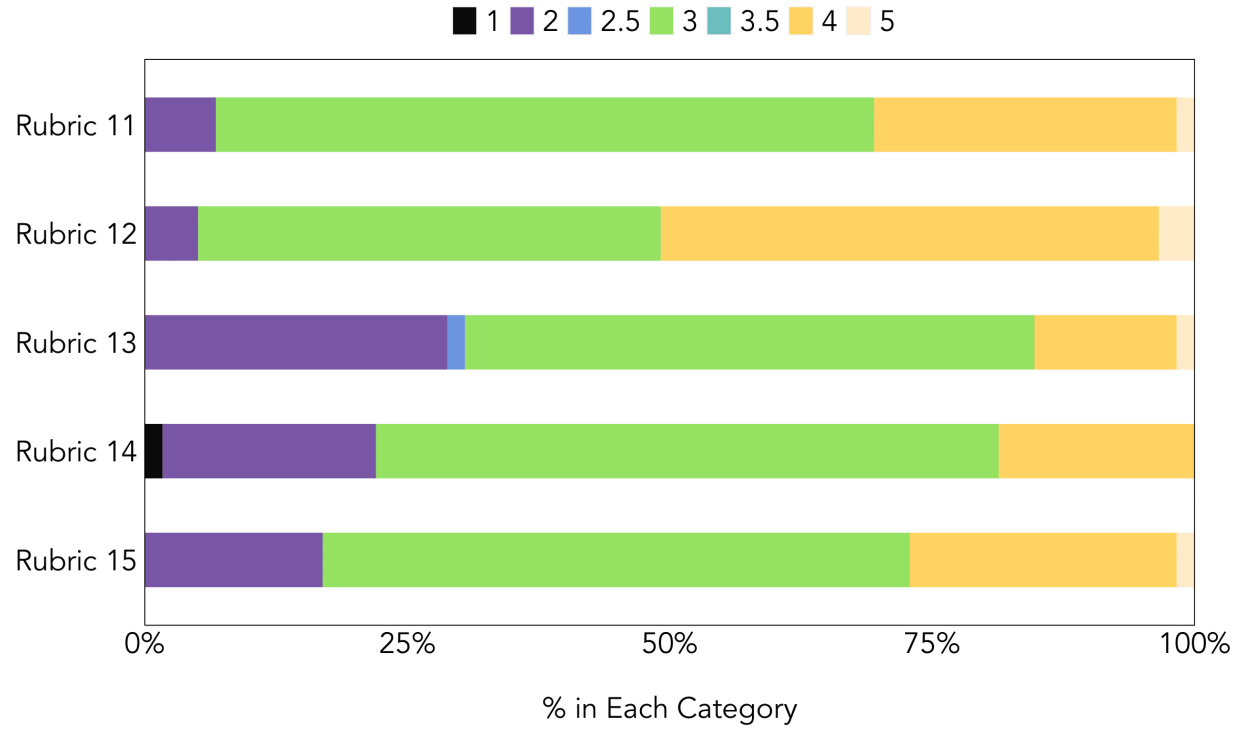

FIGURE A10. TASK 3, RUBRICS 11-15: COMPARISON GROUP

Appendix B. Details about Study Participation

After agreeing to participate in the study, study participants may leave the study for a number of reasons. The following section details how many people dropped out of each cohort during the study and for what reasons. Note that the sample for any given analysis varies vary on availability of data. We may still have collected outcomes from GaDOE or publicly available records after teachers left the study or stopped responding to surveys.

COHORT 1

Year 1

During recruitment, 20 CREATE and 59 comparison group study participants agreed to be part of the study. However, 1 participant in each group was deemed ineligible and not sent a baseline survey. The baseline sample, therefore, includes 19 CREATE and 58 comparison group study participants. A total of 15 study participants left the study in Year 1. We provide details below and in Table B1.

CREATE group

No CREATE group residents left the study in Year 1.

Comparison group

Seven of the study participants in the comparison group dropped out of the program at GSU CEHD, which made them ineligible to continue in the study. Three comparison group study participants no longer wanted to participate in the research. Three comparison group study participants did not meet the expected criteria for participating in the study. Two study participants did not have the required permissions from their districts to participate in the study.

TABLE B1. ATTRITION FROM STUDY FOR COHORT 1, YEAR 1

Event/Reason for attrition	CREATE group	Comparison group
Consented to participate in study and were eligible to participate at the time of consent	19	58
(Dropped out of GSU CEHD's program)	(0)	(7)
(No longer wanted to participate in research)	(0)	(3)
(Ineligible to continue based on eligibility criteria for the study)	(0)	(3)
(Unable to obtain district/school permission to survey participant)	(0)	(2)
Remaining sample at end of Year 1	19	43

Year 2

By the end of Year 2, an additional 27 study participants had left the study. They were all part of the comparison group. We provide details below and in Table B2.

CREATE group

No CREATE group residents left the study in Year 2.

Comparison group

Of the 27 comparison study participants who left the study in Year 2, we categorized 12 as inactive because we were unable to obtain permission to conduct research from their schools and/or districts. Seven study participants left the study because they were no longer working in school settings. Seven study participants were non-responsive to data collection attempts and one indicated that they did not want to continue participation in the study.

TABLE B2. ATTRITION FROM STUDY FOR COHORT 1, YEAR 2

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 2	19	43
(Unable to obtain district/school permission to survey participant)	(0)	(12)
(No longer working in a school/classroom setting)	(0)	(7)
(No longer wanted to participate in research/Non-responsive to follow-up)	(0)	(8)
Remaining sample at end of Year 2	19	16

Year 3

Nine study participants were made inactive in Year 3, and three study participants returned to the study. We provide details below and in Table B3.

CREATE group

Seven CREATE residents became inactive in Year 3. Six residents dropped from the CREATE program in Year 3: one moved and was substitute teaching in a new state, one started attending graduate school, and four started teaching in different districts. The final participant decided they were no longer interested in participating in the study.

Comparison group

Two comparison group study participants became inactive in Year 3. One participant stopped working in a classroom setting, which made them ineligible to continue in the study, and one participant was nonresponsive to data-collection attempts. Three study participants who became inactive in Year 2 due to research permissions re-entered the study in Year 3 when we were able to obtain permission from their school/district.

TABLE B3. ATTRITION FROM STUDY FOR COHORT 1, YEAR 3

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 3	19	16
(Unable to obtain district/school permission to survey participant)	(0)	(0)
Obtained district/school permission to survey participant	0	3
(Left CREATE school to teach in another district, move out of state, or return to graduate school)	(5)	N/A
(No longer working in a school/classroom setting)	(1)	(1)
(No longer wanted to participate in research/Non-responsive to follow-up)	(1)	(1)
Remaining sample at end of Year 3	12	17

Summary of Cohort 1 Attrition

A total of 48 study participants (41 comparison and 7 CREATE) became (and remained) inactive over the course of the study. See Table B4 for a summary table of attrition from the study sample in Cohort 1 during the three years of the study.

TABLE B4. SUMMARY OF COHORT 1 ATTRITION

Event/Reason for attrition	CREATE group	Comparison group
Consented to participate in study and were eligible to participate at the time of consent	19	58
(Unable to obtain district/school permission to survey participant)	(0)	(11)
(Ineligible to continue based on eligibility criteria for the study)	(6)	(18)
(No longer wanted to participate in research/Non-responsive to follow-up)	(1)	(12)
Remaining sample at end of Year 3	12	17

COHORT 2

Year 1

During recruitment, 23 CREATE and 40 comparison group study participants agreed to be part of the study. Eight study participants (3 CREATE and 5 comparison) left the study during Year 1. We provide details below and in Table B5.

CREATE group

Three CREATE residents dropped from the study in Year 1. One participant was not going to graduate in time, which made them ineligible for the study. The second participant dropped out of the CREATE program, with plans to teach in another state. The final participant, along with the program coordinators, took a year off and then re-joined the CREATE program the following year (year 3).

Comparison group

Of the five comparison group study participants who left the study in year 1, four became ineligible for the study for various reasons: one would not graduate on time, two never joined a practicum site, and one dropped out of GSU CEHD. The fifth participant who left in year 1 decided not to participate in the study.

TABLE B5. ATTRITION FROM STUDY FOR COHORT 2, YEAR 1

Event/Reason for attrition	CREATE group	Comparison group
Consented to participate in study and were eligible to participate at the time of consent	23	40
(Left CREATE program/school)	(2)	N/A
(Dropped out of GSU CEHD)	(0)	(1)
(No longer wanted to participate in research)	(0)	(1)
(Ineligible to continue based on eligibility criteria for the study: would not graduate on time/not placed in practicum site)	(1)	(3)
Remaining sample at end of Year 1	20	35

Year 2

By the end of Year 2, an additional 25 study participants (5 CREATE and 20 comparison) became inactive in the study. We provide details below and in Table B6.

CREATE group

Five CREATE residents became ineligible when one began teaching in another state, and the other four began teaching in non-CREATE schools in Georgia.

Comparison group

Researchers were unable to obtain research permissions from the schools and/or districts of seven comparison group study participants. Five comparison group study participants left the study because they were no longer working in classroom or school settings. Five comparison study participants did not want to participate in the study anymore. Three comparison study participants were nonresponsive to data-collection attempts.

TABLE B6. ATTRITION FROM STUDY FOR COHORT 2, YEAR 2

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 2	20	35
(Unable to obtain district/school permission to survey participant)	(0)	(7)
(Left CREATE program)	(5)	N/A
(No longer working in a school/classroom setting)	(0)	(5)
(No longer wanted to participate in research/Non-responsive to follow-up)	(0)	(8)
Remaining sample at end of Year 2	15	15

Year 3

Four Cohort 2 study participants left the study in Year 3. We provide details below and in Table B7.

CREATE group

Three CREATE residents left the study in Year 3. All three left the CREATE program for different reasons, one of which related to a school site undergoing the process of leveling—when the number of teachers at the school is reduced as a result of having lower student enrollment numbers than expected.

Comparison group

One comparison group study participant left the study in Year 3. This participant became inactive in the study due to their lack of response to data collection attempts. One participant who became inactive in Year 2 due to research permissions became active again in Year 3, when we were able to obtain permission from their school/district.

TABLE B7. ATTRITION FROM STUDY FOR COHORT 2, YEAR 3

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 3	15	15
Obtained district/school permission to survey participant	0	1
(Unable to obtain district/school permission to survey participant)	(0)	(0)
(Left CREATE program)	(3)	N/A
(No longer wanted to participate in research/Non-responsive to follow-up)	(0)	(1)
Remaining sample at end of Year 3	12	15

Summary of Cohort 2 Attrition

Over the course of the study, a total of 36 Cohort 2 study participants (25 comparison and 11 treatment) became inactive (and remained inactive) during the study. See Table B8 for a summary table of attrition from the study sample in Cohort 2 during the three years of the study.

TABLE B8. SUMMARY OF COHORT 2 ATTRITION

Event/Reason for attrition	CREATE group	Comparison group
Consented to participate in study and were eligible to participate at the time of consent	23	40
(Unable to obtain district/school permission to survey participant)	(0)	(6)
(Ineligible to continue based on eligibility criteria for the study)	(11)	(9)
(No longer wanted to participate in research/Non-responsive to follow-up)	(0)	(10)
Remaining sample at end of Year 3	12	15

COHORT 3

Year 1: Summary of Cohort 3 Attrition

During recruitment, 14 CREATE and 48 comparison group study participants agreed to be part of the study. Nine Cohort 3 participants, all of whom were part of the comparison group, became inactive during Year 1 of the study. Of these nine participants, four had dropped from the program at GSU CEHD, another four were ineligible for the study (including one participant who had taken a semester off from their program due to athletics, one who started teaching a grade not included in the study, and two who never joined practicum sites), and one participant was not responsive to data collection attempts and follow-up communication. Table B9 includes a summary of attrition for Cohort 3.

TABLE B9. ATTRITION FROM STUDY FOR COHORT 3, YEAR 1

Event/Reason for attrition	CREATE group	Comparison group
Consented to participate in study and were eligible to participate at the time of consent	14	48
(Dropped out of GSU CEHD)	(0)	(4)
(No longer wanted to participate in research/ Non-responsive to follow-up)	(0)	(1)
(Ineligible to continue based on eligibility criteria for the study)	(0)	(4)
Remaining sample at end of Year 1	14	41

Appendix C. Survey Response Rates

Tables C1 through C7 provide response rates to the quarterly surveys of participants who were active at the end of Year 3 of participation.

TABLE C1. COHORT 1, YEAR 1 SURVEY RESPONSE RATES (2015–16)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Baseline survey	May 2015	72/78 (92%)	19/19 (100%)	53/59 (90%)
Quarterly survey 1	November 2015	57/62 (92%)	19/19 100%	38/43 (88%)
Quarterly survey 2	January 2016	52/62 (84%)	18/19 (95%)	34/43 (79%)
Quarterly survey 3	March 2016	51/62 (82%)	19/19 (100%)	32/43 (74%)
Quarterly survey 4	April 2016	54/62 (87%)	19/19 (100%)	35/43 (81%)

TABLE C2. COHORT 1, YEAR 2 SURVEY RESPONSE RATES (2016–17)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	December 2016	31/33 (94%)	19/19 (100%)	12/14 (86%)
Quarterly survey 2	January 2017	27/33 (82%)	19/19 (100%)	8/14 (57%)
Quarterly survey 3	March 2017	34/35 (97%)	19/19 (100%)	15/16 (94%)
Quarterly survey 4	April 2017	33/35 (94%)	19/19 (100%)	14/16 (88%)

TABLE C3. COHORT 1, YEAR 3 SURVEY RESPONSE RATES (2017–18)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	November 2017	26/26 (100%)	12/12 (100%)	14/14 (100%)
Quarterly survey 2	January 2018	26/26 (100%)	12/12 (100%)	14/14 (100%)
Quarterly survey 3	March 2018	26/26 (100%)	12/12 (100%)	14/14 (100%)
Quarterly survey 4	April 2018	26/26 (100%)	12/12 (100%)	13/14 (93%)

TABLE C4. COHORT 2, YEAR 1 SURVEY RESPONSE RATES (2016–17)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Baseline survey	March 2016	52/55 (95%)	19/20 (95%)	33/35 (94%)
Quarterly survey 1	November 2016	51/55 (93%)	20/20 (100%)	31/35 (89%)
Quarterly survey 2	January 2017	49/55 (89%)	19/20 (95%)	30/35 (86%)
Quarterly survey 3	March 2017	50/55 (91%)	20/20 (100%)	30/35 (86%)
Quarterly survey 4	April 2017	49/55 (89%)	20/20 (100%)	29/35 (83%)

TABLE C5. COHORT 2, YEAR 2 SURVEY RESPONSE RATES (2017–18)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	November 2017	29/30 (97%)	15/15 (100%)	14/15 (93%)
Quarterly survey 2	January 2018	29/30 (97%)	15/15 (100%)	14/15 (93%)
Quarterly survey 3	March 2018	29/30 (97%)	15/15 (100%)	14/15 (93%)
Quarterly survey 4	April 2018	29/30 (97%)	15/15 (100%)	14/15 (93%)

TABLE C6. COHORT 2, YEAR 3 SURVEY RESPONSE RATES (2018–19)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	November 2018	27/27 (100%)	12/12 (100%)	15/15 (100%)
Quarterly survey 2	January 2019	26/27 (96%)	11/12 (92%)	15/15 (100%)
Quarterly survey 3	March 2019	26/27 (96%)	11/12 (92%)	15/15 (100%)
Quarterly survey 4	April 2019	27/27 (100%)	12/12 (100%)	15/15 (100%)

TABLE C7. COHORT 3, YEAR 1 SURVEY RESPONSE RATES (2017–18)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Baseline survey	July 2017	55/55 (100%)	14/14 (100%)	41/41 (100%)
Quarterly survey 1	November 2017	55/55 (100%)	14/14 (100%)	41/41 (100%)
Quarterly survey 2	January 2018	53/55 (96%)	13/14 (93%)	40/41 (98%)
Quarterly survey 3	March 2018	53/55 (96%)	13/14 (93%)	40/41 (98%)
Quarterly survey 4	April 2018	54/55 (98%)	13/14 (93%)	41/41 (100%)

Appendix D. Fidelity of Implementation Matrix

TABLE D1. KEY COMPONENT 1: PROGRESSIVE CORE CLASSROOM ROLES

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 1: Y1	(Y1) Resident is paired with another Y1 resident for fall semester (C1 = within same classroom; C2 = in same school building)	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not paired with another Y1 resident; 1 = Resident is paired with another Y1 resident for part of the semester 2 = Resident is paired with another Y1 resident for whole semester	Adequate implementation at resident level = score of "1"		20 residents (Cohort 1) in SY 2015-2016 26 residents (Cohort 2) in SY 2016-2017	Cohort 1: SY 2015-2016 Cohort 2: SY 2016-2017
Indicator 2: Y1	(Y1) Resident is placed in classroom of veteran educator trained in mentoring	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not placed with a veteran educator trained in mentoring; 1 = Resident is placed with a veteran educator trained in mentoring	Adequate implementation at resident level = score of "1"		20 residents (Cohort 1) in SY 2015-2016 26 residents (Cohort 2) in SY 2016-2017	Cohort 1: SY 2015-2016 Cohort 2: SY 2016-2017
Indicator 3: Y1	(Y1) Resident is placed in a CREATE school	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is placed in a non-CREATE school; 1 = Resident is placed in a CREATE school	Adequate implementation at resident level = score of "1"		20 residents (Cohort 1) in SY 2015-2016 26 residents (Cohort 2) in SY 2016-2017	Cohort 1: SY 2015-2016 Cohort 2: SY 2016-2017
Indicator 1: Y2	(Y2) Resident is *co-teaching with another Y2 resident* as a teacher of record	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not paired with co-resident; 1 = Resident is paired with co-resident	Adequate implementation at resident level = score of "1"		20 residents (Cohort 1) in SY 2016-2017 26 residents (Cohort 2) in SY 2017-2018	Cohort 1: SY 2016-2017 Cohort 2: SY 2017-2018

TABLE D1. KEY COMPONENT 1: PROGRESSIVE CORE CLASSROOM ROLES

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 2: Y2	(Y2) Resident is teaching in *subject and grade level for which resident is certified*	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not teaching in subject and grade level for which resident is certified; 1 = Resident is teaching in subject and grade level for which resident is certified	Adequate implementation at resident level = score of "1"		20 residents (Cohort 1) in SY 2016-2017 26 residents (Cohort 2) in SY 2017-2018	Cohort 1: SY 2016-2017 Cohort 2: SY 2017-2018
Indicator 3: Y2	(Y2) Resident is teaching *at a CREATE school*	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not teaching at a CREATE school; 1 = Resident is teaching at a CREATE school	Adequate implementation at resident level = score of "1"		20 residents (Cohort 1) in SY 2016-2017 26 residents (Cohort 2) in SY 2017-2018	Cohort 1: SY 2016-2017 Cohort 2: SY 2017-2018
Indicator 1: Y3	(Y3) Resident is teaching solo as teacher of record	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not teaching solo as teacher of record; 1 = Resident is teaching solo as teacher of record	Adequate implementation at resident level = score of "1"		20 residents (Cohort 1) in SY 2017-2018 26 residents (Cohort 2) in SY 2018-2019	Cohort 1: SY 2017-2018 Cohort 2: SY 2018-2019
Indicator 2: Y3	(Y3) Resident is teaching in subject and grade level for which resident is certified	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not teaching in subject area and grade level for which resident is certified; 1 = Resident is teaching in subject and grade level for which resident is certified	Adequate implementation at resident level = score of "1"		20 residents (Cohort 1) in SY 2017-2018 26 residents (Cohort 2) in SY 2018-2019	Cohort 1: SY 2017-2018 Cohort 2: SY 2018-2019

TABLE D1. KEY COMPONENT 1: PROGRESSIVE CORE CLASSROOM ROLES

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 3: Y3	(Y3) Resident is teaching at a CREATE school	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not teaching at a CREATE school; 1= Resident is teaching at a CREATE school	Adequate implementation at resident level = score of "1		20 residents (Cohort 1) in SY 2017-2018 26 residents (Cohort 2) in SY 2018-2019	Cohort 1: SY 2017-2018 Cohort 2: SY 2018-2019
All indicators					Resident-level component score: Number of indicators implemented with fidelity per year. Possible range (per year) = 0 - 3		Component score ranges from 0-1: Fidelity not met: Y1: Less than 95% of residents meet fidelity on 2+ indicators Y2: Less than 75% of residents meet fidelity on 2+ indicators Y3: Less than 85% of residents meet fidelity on 2+ indicators Fidelity met = Y1: 95% or more of residents meet fidelity on 2+ indicators; Y2: 75% or more of residents meet fidelity on 2+ indicators; Y3: 85% or more of residents meet fidelity on 2+ indicators	20 (Cohort 1) + 26 (Cohort 2) = 46 total residents	Cohort 1: SY 2015-2016 (Y1), SY 2016-2017 (Y2), SY 2017-2018 (Y3); Cohort 2: SY 2016-2017(Y1), SY 2017-2018 (Y2), SY 2018-2019 (Y3)

TABLE D2. KEY COMPONENT 2: CRITICAL FRIENDSHIP

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 1	CREATE administrators host 2 or more options for veteran teachers to attend 4-day CFG Institute each year. (Institute will not be held if fewer than 16 educators sign up).	Program	CREATE attendance rosters	Rosters emailed to evaluators from CREATE	0 = CREATE administrators host 0-1 institute; 1= CREATE administrators host 2+ institutes	Adequate implementation at program level = score of 1"	0 = Fidelity not met at unit level; 1 = Fidelity met at unit level		SY 2015-2016, SY 2016-2017, SY 2017-2018, SY 2018-2019
Indicator 2	Veteran educators attend 4-day CFG Institute	Non-resident educators from CREATE schools	CREATE attendance rosters	Rosters emailed to evaluators from CREATE	0 = Educator attends 1-2 days; 1 = Educator attends 3-4 days	Adequate implementation = score of "1"	0 = Of the veteran educators who attend, less than 85% meet fidelity at the unit level; 1 = Of the veteran educators who attend, 85% or more meet fidelity at the unit level	Approx. 40 non-resident educators	SY 2015-2016, SY 2016-2017, SY 2017-2018, SY 2018-2019
Indicator 3	Residents attend 4-day CF Institute by the end of Year 3 of their program. Residents may attend the CF Institute in Year 2 or Year 3.	CREATE residents	CREATE attendance rosters Resident surveys	Rosters emailed to evaluators from CREATE Surveys administered quarterly by evaluators	0 = Resident attends 0-2 days 1 = Resident attends 3-4 days	Adequate implementation = score of "1"	0 = Less than 25% of residents meet fidelity at the unit level 1 = 25% or more of residents meet fidelity at the unit level	*n* = 20 CREATE residents in Cohort 1 *n* = 26 CREATE residents in Cohort 2	Cohort 1: Summer 2016, Summer 2017 Cohort 2: Summer 2017, Summer 2018

TABLE D2. KEY COMPONENT 2: CRITICAL FRIENDSHIP

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 4	CREATE residents attend monthly CFG meetings.	Residents	CREATE attendance rosters Resident surveys	Rosters emailed to evaluators from CREATE Surveys administered quarterly by evaluators	Y1 residents: 0 = Attend 0-7 meetings; 1 = Attend 8-10 meetings Y2 residents: 0 = Attend 0-6 meetings, 1 = Attend 7+ meetings Y3 residents: 0 = Attend 0-2 meetings; 1 = Attend 3 meetings	Adequate implementation = score of "1"	0 = Less than 95% of residents meet fidelity 1 = 95% or more of residents meet fidelity at the unit level	*n* = 20 CREATE residents in Cohort 1 *n* = 26 CREATE residents in Cohort 2	Cohort 1: SY 2015-2016, SY 2016-2017, SY 2017-2018; Cohort 2: SY 2016-2017, SY 2017-2018; SY 2018-2019
All indicators							Years 1 and 2 (indicator 3 not measured) 0 = Fidelity not met= Fidelity not met for any indicator, or fidelity met for Indicator 1 but no other indicator 1 = Fidelity met= Fidelity was met for Indicator 1 and at least one other indicator Year3 0 = Fidelity not met= fidelity not met for indicators 1 or 3, OR fidelity met for indicators 1 and 3 but not for Indicators 2 and 4 1 = Fidelity met= Fidelity was met for Indicator 1, Indicator 3 and at least one other indicator	*n* = 20 CREATE residents in Cohort 1 *n* = 26 CREATE residents in Cohort 2 *n* = 40 veteran teachers per year	Cohort 1: SY 2015-2016, SY 2016-2017, SY 2017-2018 Cohort 2: SY 2016-2017, SY 2017-2018; SY 2018-2019

TABLE D3. KEY COMPONENT 3: COGNITIVELY-BASED COMPASSION TRAINING

Indicators	Definition	Unit of implementation	Data source (s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 1	Program administrators offer at least one Cognitively-Based Compassion Training (CBCT®) per year for general population of teachers at CREATE schools	Program	Email communication with CREATE administrator	Annual email communication with CREATE administrators	0 = CREATE administrators offer 0 courses 1 = CREATE administrators offer 1 or more courses	Adequate implementation = score of "1"	0 = Fidelity not met at unit level 1 = Fidelity met at unit level		SY 2015-2016, SY 2016-2017, SY 2017-2018, SY 2018-2019
Indicator 2	Residents attend CBCT classes	Resident	CREATE attendance rosters Resident surveys	Rosters emailed to evaluators from CREATE Surveys administered quarterly by evaluators	Y1 residents: 0 = Attend 0-7 meetings; 1 = Attend 8-10 meetings Y2 residents: 0 = Attend 0-6 meetings, 1 = Attend 7+ meetings Y3 residents: 0 = Attend 0-2 meetings; 1 = Attend 3 meetings	Adequate implementation = score of "1"	0 = Less than 95% of residents meet fidelity at the unit level 1= 95% or more of residents meet fidelity at the unit level		Cohort 1: SY 2015-2016, SY 2016-2017, SY 2017-2018 Cohort 2: SY 2016-2017, SY 2017-2018; SY 2018-2019
All indicators							Component score is sum of indicator scores. Range is 0-2 Fidelity not met = score of 0-1 Fidelity met = score of 2	20 (Cohort 1)+26 (Cohort 2)=46 total residents Y1: Cohort 1 SY 2015-2016; Cohort 2 SY 2016-2017	Cohort 1: SY 2015-2016, SY 2016-2017, SY 2017-2018; Cohort 2: SY 2016-2017, SY 2017-2018, SY 2018-2019

TABLE D4. KEY COMPONENT 4: MULTIPLE FORMS OF MENTORING

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 1	Residents have mentors who attended training prior to mentoring	Mentor	Rosters from CREATE	Rosters emailed to evaluators from CREATE	0 = Resident has a mentor who did not attend training 1 = Resident has a mentor who attended training	Adequate implementation = score of "1"	0 = Less than 100% of residents have mentors that meet fidelity at the unit level 1 = 100% of residents have mentors that meet fidelity at the unit level		
Indicator 2	Residents have mentors who attended training during their mentoring year (2 sessions)	Mentor	Rosters from CREATE	Rosters emailed to evaluators from CREATE	0 = Resident has a mentor who attends 0-1 sessions 1 = Resident has a mentor who attends 2 or more sessions	Adequate implementation = score of "1"	0 = Less than 90% of residents have mentors that meet fidelity at the unit level 1 = 90% or more of mentors have residents that meet fidelity at the unit level		

TABLE D4. KEY COMPONENT 4: MULTIPLE FORMS OF MENTORING

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 3	Residents attend semi-monthly meetings with their mentor (SBM and IM)	Resident	Resident surveys, CREATE mentor logs (and spreadsheet)	Mentor logs emailed to evaluators from CREATE, Surveys administered quarterly by evaluators	C1Y2: 0 = Resident attends 0-8 meetings; 1 = Resident attends 9-11 meetings; 2 = Resident attends 12 or more meetings C2Y2: *0 = Resident attends 0-22 meetings;* *1 = Resident attends 23-27 meetings;* *2 = Resident attends 28 or more meetings* C1Y3 and C2Y3: 0 = Resident attends 0-8 meetings; 1 = Resident attends 9-11 meetings; 2 = Resident attends 12 or more meetings	Adequate implementation = score of "1" in Y2 and Y3	0 = Less than 95% of residents earn a score of 2 at the unit level *or* anyone earns a score of 0; 1 = 95% of residents earn a score of 2 at the unit level *and* no residents earn a score of 0		

TABLE D4. KEY COMPONENT 4: MULTIPLE FORMS OF MENTORING

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 4	Resident participates in mentor-resident observation cycles	Resident	Resident surveys CREATE mentor logs (and spreadsheet)		0 = Resident participates in 0-1 cycles 1 = Resident participates in 2-3 cycles	Adequate implementation = score of 1	Y1: N/A Y2: 0 = Less than 90% of residents meet fidelity at the unit level 1 = 90% or more of residents meet fidelity at the unit level Y3: 0 = Less than 80% of residents meet fidelity at the unit level 1 = 80% or more of residents meet fidelity at the unit level		
All indicators							Sum of indicator scores Component score ranges from 0-4: Fidelity not met = score of 0-3 Fidelity met = score of 4	*n* = 20 CREATE residents and mentors in Cohort 1 *n* = 26 CREATE residents and mentors in Cohort 2	Cohort 1: SY 2015-2016, SY 2016-2017, SY 2017-2018 Cohort 2: SY 2016-2017, SY 2017-2018; SY 2018-2019

TABLE D5. KEY COMPONENT 5: SUMMER RESIDENT ACADEMY

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Indicator 1	Residents attend summer Internship/ academy	Residents	CREATE attendance rosters Resident surveys	Rosters emailed to evaluators from CREATE Surveys administered quarterly by evaluators	For Cohort 1: 0 = Resident does not enroll in internship; 1 = Resident attends more than 2 weeks and less than 3 weeks of internship; 2 = Resident attends internship for 3 weeks For Cohort 2: 0 = Resident attends less than 15 days; 1 = Resident attends 15-20 days; 2 = resident attends more than 20 days	Adequate implementation = score of "1"			
All indicators							Component score ranges from 0-1: 0 = Fidelity not met: Less than 95% of residents earn a score of at least 1 *or* less than 85% of residents earn a score of 2 at the unit level 1= Fidelity Met: 95% or more of residents earn a score of at least 1 *and* 85% or more of residents earn a score of 2 at the unit level	19 (Cohort 1) + 15 (Cohort 2) = 34 total residents	Cohort 1: Summer 2016 Cohort 2: Summer 2017

Appendix E. Fidelity of Implementation Results

In each section below, we provide a detailed description of each of CREATE's five key components. Then, we provide the Fidelity of Implementation (FOI) results for Cohorts 1 and 2 for each component, during all three years of the study.

KEY COMPONENT 1: PROGRESSIVE CORE CLASSROOM ROLES

As CREATE residents move through the three-year residency model, their role within the classroom changes. The progressive core classroom roles provide supports for residents while also providing space for increased autonomy, agency, and independence each year.

Residents enter the CREATE residency during the final year of their GSU CEHD teacher certification program. In this first year, two Year 1 residents are paired with one highly-skilled veteran teacher—known as a cooperating teacher—in a classroom at a CREATE school, for at least the fall semester.

Upon graduating from the GSU CEHD teacher certification program, residents in year 2 should continue on at a CREATE school, paired with each other as co-lead teachers of record. The lighter load and flexibility of having two teachers in one classroom allows more time for mentor-resident reflection and observations of other teachers. This arrangement is also intended to address the sense of being overwhelmed with new responsibilities that new teachers often cite as the reason they leave teaching. In addition, each resident should be teaching in the subject and grade level for which they are certified. As the program developed, CREATE recognized that some teachers worked more successfully in their first year of teaching if allowed to work as the sole teacher of record. Because CREATE has a close working relationship with each of their residents by year 2 of the residency, CREATE administrators determined who would be best suited to teach alone versus with a co-teacher. We have taken this into account in our calculation of FOI for the associated indicator.

In year 3, all residents become lead teachers in their own classrooms after having one to two years of supported co-teaching experiences. The CREATE program expects residents to teach at a CREATE school in the subject and grade for which they are certified.

We present data on the following indicators related to progressive core classroom roles.

- Indicator 1
    - Year 1: Resident is paired with another Y1 resident for fall semester (same classroom for Cohort 1, same building for Cohort 2)
    - Year 2: Resident is co-teaching with another Y2 resident, both as full-time teachers of record
    - Year 3: Resident is teaching as the solo teacher of record
- Indicator 2
    - Year 1: Resident is placed in a classroom of a veteran educator trained in mentoring
    - Year 2: Resident is teaching in a subject and grade for which the resident is certified
    - Year 3: Resident is teaching in a subject and grade for which the resident is certified
- Indicator 3
    - Year 1: Resident is placed in a CREATE school

- Year 2: Resident is teaching at a CREATE school
- Year 3: Resident is teaching at a CREATE school

The following tables of results display data from left to right. *Column 1* describes the indicators (or details) of each CREATE program component that should be implemented with fidelity. *Column 2* indicates the threshold that CREATE has set in order to meet fidelity on the respective indicator. (The threshold CREATE needs to meet is bolded in column 2.) *Column 3* indicates how many residents in Cohort 1 met the threshold for fidelity. *Column 4* indicates how many residents in Cohort 2 met the threshold for fidelity. The bottom row of each table is the final result showing whether or not CREATE met FOI for each component at the program level.

Result for Key Component 1 in Year 1

CREATE administrators expressed that they intended to have 95% or more of residents meet fidelity on at least two of three indicators listed above in Year 1. The records from CREATE attendance rosters and survey data show that 100% of residents in Cohort 1 and Cohort 2 met fidelity in year 1. More details about the number of residents who met each specific indicator are in the table below.

TABLE E1. DETAILS OF COMPONENT 1 IN YEAR 1

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Indicator 1: **Resident is paired with another Y1 resident for fall semester (C1 = within same classroom; C2 = in same school building)**	0 = Resident is not paired with a co-resident **1 = Resident is paired with a co-resident**	15/19 (79%)	20/20 (100%)
Indicator 2: **Resident is placed in classroom of veteran educator trained in mentoring**	0 = Resident is not placed with a veteran educator trained in mentoring **1 = Resident is placed with a veteran educator trained in mentoring**	15/19 (79%)	9/20 (45%)
Indicator 3: **Resident is placed in a CREATE school**	0 = Resident is not teaching at a CREATE school **1 = Resident is teaching at a CREATE school**	19/19 (100%)	20/20 (100%)
Program level	95% or more of residents meet fidelity on 2+ indicators	19/19 (100%) Fidelity Met	20/20 (100%) Fidelity Met

Note. While 23 people joined Cohort 2, 3 of them dropped out of the study during Year 1. More information about the sample attrition can be found in Appendix B.

Result for Key Component 1 in Year 2

CREATE administrators expressed that they intended to have 75% or more of residents meet fidelity on at least two of the three indicators of Component 1 in Year 2. The records from CREATE attendance rosters and survey data show that 100% of active residents in Cohort 1 and 87% of active residents in Cohort 2 met fidelity in Year 2; both cohorts met fidelity. More details about the number of residents who met each specific indicator are in the table below.

TABLE E2. DETAILS OF COMPONENT 1 IN YEAR 2

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Indicator 1: **Resident is co-teaching with another Y2 resident as a teacher of record**	0 = Resident is not paired with co-resident 1 = Resident is paired with co-resident	16/19 (80%)	10/15 (67%)
Indicator 2: **Resident is teaching in a subject and grade level for which the resident is certified**	0 = Resident is not teaching in a subject and grade level for which resident is certified 1 = Resident is teaching in a subject and grade level for which resident is certified	19/19 (100%)	13/15 (87%)
Indicator 3: **Resident is teaching at a CREATE school**	0 = Resident is not teaching at a CREATE school 1 = Resident is teaching at a CREATE school	19/19 (100%)	15/15 (100%)
Program level	75% or more of residents meet fidelity on 2+ indicators	19/19 (100%) Fidelity Met	13/15 (87%) Fidelity Met

Note. While 23 people joined Cohort 2, 8 of them dropped out of the study for various reasons by the end of Year 2. More information about the sample attrition can be found in Appendix B.

Result for Key Component 1 in Year 3

CREATE administrators expressed that they intended to have 85% or more of residents meet fidelity on at least two of the three indicators in component 1 in year 3. The records from CREATE attendance rosters and survey data show that all active residents in Cohort 1 and Cohort 2 met fidelity in year 3. More details about the number of residents who met each specific indicator are in the table below.

TABLE E3. DETAILS OF COMPONENT 1 IN YEAR 3

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Indicator 1: Resident is teaching as the solo teacher of record	0 = Resident is not teaching as the solo teacher of record 1 = Resident is teaching as the solo teacher of record	11/12 (92%)	11/12 (92%)
Indicator 2: Resident is teaching in a subject and grade level for which resident is certified	0 = Resident is not teaching in a subject and grade level for which resident is certified 1 = Resident is teaching in a subject and grade level for which resident is certified	12/12 (100%)	12/12 (100%)
Indicator 3: Resident is teaching at a CREATE school	0 = Resident is not teaching at a CREATE school 1 = Resident is teaching at a CREATE school	12/12 (100%)	9/12 (75%)
Program level	85% or more of residents meet fidelity on 2+ indicators	12/12 (100%) Fidelity Met	12/12 (100%) Fidelity Met

Note. While 19 people joined Cohort 1, 7 of them had left the study by the end of Year 3. Similarly, while 23 people joined Cohort 2, 11 of them dropped out of the study for various reasons by the end of Year 3. More information about the sample attrition can be found in Appendix B.

Overall Results for Component 1

The CREATE program met fidelity on component 1 for both Cohorts 1 and 2 in all three years of their participation. We can, therefore, conclude that CREATE successfully implemented the core progressive classroom roles from the 2015–16 to 2018–19 school years for Cohort 1 and Cohort 2 residents.

KEY COMPONENT 2: CRITICAL FRIENDSHIP

CF work is designed to enhance pedagogical skills and build and sustain support networks for teachers. CREATE residents are placed at a CREATE school that has prepared itself for the arrival of residents by building a culture of collaboration through engagement in the work of CF. In CF meetings, educators gather together to discuss student work, educator work (such as unit plans, instructional strategies, and use of assessments or rubrics), and dilemmas of practice.

CREATE schools send teams of teachers to learn together at a four-day summer CF institute. CREATE administrators provide two or more options for a summer and/or a fall CF institute for veteran educators to choose from.

While CF meetings are ultimately offered as a form of professional development for all teachers at all CREATE schools, the year 1 residents themselves are placed in a CF group consisting solely of year 1 residents.

All CREATE residents attend monthly meetings with other residents, called Together Time. These meetings include both CF and CBCT work. There were ten meetings per year in year 1 and seven meetings per year in years 2 and 3 of the program. Each resident should also attend a four-day CF Institute in either year 2 or year 3 of their program.

We present data on the following indicators related to CF.

- Indicator 1 (Years 1, 2, and 3): CREATE administrators host 2 or more options for veteran teachers to attend a 4-day CF Institute each year. (Institutes were not held if fewer than 16 educators signed up.)
- Indicator 2 (Years 1, 2, and 3): Veteran educators attend a 4-day CF Institute.
- Indicator 3 (Years 2 and 3): Residents in year 2 or year 3 of their program attend a 4-day CF Institute.
- Indicator 4 (Years 1, 2, and 3): CREATE residents attend monthly CF meetings.

Results of Key Component 2 for Year 1

CREATE administrators expect that at least two of the three indicators listed above (i.e., from among Indicators 1, 2, and 4) will reach fidelity in Year 1. Cohort 1 and Cohort 2 met or exceeded the thresholds for fidelity on all Year 1 indicators. More details about the number of residents who met each specific indicator are in the table below.

The first two indicators in the tables below refer to veteran educators, who are not part of a particular cohort of residents. For this reason, *Columns 3* and *4* indicate the particular school year.

TABLE E4. DETAILS OF COMPONENT 2 IN YEAR 1

Indicator	Fidelity threshold	2015–16 school year	2016–17 school year
Indicator 1: **CREATE administrators host 2 or more options for veteran educators to attend a 4-day CF Institute each year**	0 = CREATE administrators host 0-1 institute 1= CREATE administrators host 2+ institutes	CREATE administrators host 3 institutes Fidelity met	CREATE administrators host 4 institutes Fidelity met
Indicator 2: **Veteran educators attend a 4-day CF Institute**	0 = Educator attends 1-2 day 1 = Educator attends 3-4 days Of the veteran educators who attend, 85% or more meet fidelity at the unit level	79/85 (93%) Fidelity met	99/105 (94%) Fidelity met

TABLE E4. DETAILS OF COMPONENT 2 IN YEAR 1

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Indicator 4: CREATE residents attend monthly CF meetings	0 = Attend 0-7 meetings 1 = Attend 8-10 meetings 95% or more of residents meet fidelity at the unit level	18/19 (95%) Fidelity met	20/20 (100%) Fidelity met
Program level	0 = Fidelity not met for any indicator, or fidelity met for Indicator 1 but no other indicator 1 = Fidelity was met for Indicator 1 and at least one other indicator	Fidelity met	Fidelity met

Note. Indicator 3 (Residents in year 2 or year 3 of their program attend a 4-day CF Institute) is not included in this table because it does not apply to Year 1 residents.

Results of Key Component 2 for Year 2

CREATE administrators expect that indicator 1, indicator 3,[3] and at least one other indicator are met in year 2 in order to reach FOI. The records from CREATE attendance rosters and survey data show that both Cohort 1 and Cohort 2 met fidelity on all but one Year 2 indicator which, again, exceeds expectations. There were three residents who did not attend the required seven (or more) CF meetings in Year 2. However, two of those residents attended six meetings, and one of them attended five meetings, so they were just below the threshold needed to meet fidelity on indicator 4. Overall, component 2 meets fidelity for both cohorts in Year 2.

More details about the number of residents who met each specific indicator are in the table below.

[3] See Table E6. Details of Component 2 in Year 3 for details regarding Indicator 3 in Year 2.

TABLE E5. DETAILS OF COMPONENT 2 IN YEAR 2

Indicator	Fidelity threshold	2016–17 school year	2017–18 school year
Indicator 1: CREATE administrators host 2 or more options for veteran teachers to attend a 4-day CF Institute each year	0 = CREATE administrators host 0-1 institute 1= CREATE administrators host 2+ institutes	CREATE administrators host 4 institutes Fidelity met	CREATE administrators host 6 institutes Fidelity met
Indicator 2: Veteran teachers attend a 4-day CF Institute	0 = Educator attends 1-2 day 1 = Educator attends 3-4 days 85% or more meet fidelity at the unit level	99/105 (94%) Fidelity met	62/68 (91%) Fidelity met
Indicator	**Fidelity threshold**	**Cohort 1 met fidelity?**	**Cohort 2 met fidelity?**
Indicator 4: CREATE residents attend monthly CF meetings	0 = Attend 0-6 meetings 1 = Attend 7+ meetings 95% or more of residents meet fidelity at the unit level	19/19 (100%) Fidelity met	12/15 (80%) Fidelity not met
Program level	0 = Fidelity not met for indicators 1 or 3, OR fidelity met for indicators 1 and 3 but not for Indicator 2 and 4. 1 = Fidelity was met for Indicator 1, Indicator 3 and at least one other indicator	Fidelity met	Fidelity met

Note. Indicator 3 (Residents in year 2 or year 3 of their program attend a 4-day CF Institute) is included in the next (Y3) table.

Results of Key Component 2 for Year 3

CREATE administrators expect that indicator 1, indicator 3, and at least one other indicator in Year 3 will reach FOI. The records from CREATE attendance rosters and survey data show that both Cohort 1 and Cohort 2 met fidelity on Component 2 in Year 3.

More details about the number of residents who met each specific indicator are in the table below.

TABLE E6. DETAILS OF COMPONENT 2 IN YEAR 3

Indicator	Fidelity threshold	2017–18 school year	2018–19 school year
Indicator 1: **CREATE administrators host 2 or more options for veteran teachers to attend a 4-day CF Institute each year**	0 = CREATE administrators host 0-1 institute 1= CREATE administrators host 2+ institutes	CREATE administrators host 4 institutes Fidelity met	CREATE administrators host 5 institutes Fidelity met
Indicator 2: **Veteran teachers attend a 4-day CF Institute**	0 = Educator attends 1-2 day 1 = Educator attends 3-4 days 85% or more meet fidelity at the unit level	62/68 (91%) Fidelity met	69/78 (88%) Fidelity met

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Indicator 3: **Residents attend a 4-day CF Institute by the end of Year 3 of their program** **Residents may attend the CF Institute in Year 2 or Year 3**	0 = Resident attends 0-2 days 1 = Resident attends 3-4 days 25% or more of residents meet fidelity at the unit level	6/19 (32%) Fidelity met	4/15 (27%) Fidelity met
Indicator 4: **CREATE residents attend monthly CF meetings**	0 = Attend 0-2 meetings 1 = Attend 3 meetings 95% or more of residents meet fidelity at the unit level	11/12 (92%) Fidelity not met	12/12 (100%) Fidelity met
Program level	0 = Fidelity not met for indicators 1 or 3, OR fidelity met for indicators 1 and 3 but not for Indicator 2 and 4. 1 = Fidelity was met for Indicator 1, Indicator 3 and at least one other indicator	Fidelity met	Fidelity met

Note. Indicator 3 applies to residents in Year 2 or Year 3 of the residency. The result included here is the total number of residents who attended a 4-day CF institute in either Year 2 or Year 3. The total sample size provided is the active sample from Year 2 to ensure that residents who attended in Year 2 but dropped out of the study by Year 3 are included in the calculation.

Because Component 2 and Component 3 have similar requirements and results, a summary of Component 2 is at the end of the next section describing Component 3. Furthermore, monthly Together Time meetings include both CF and CBCT

components. Therefore, the results for monthly CF meetings (Key Component 2, Indicator 4) and monthly CBCT meetings (Key Component 3, Indicator 2) are identical.

Key Component 3: CBCT

Acknowledging that negative collegial and student relationships can diminish a teacher's energy for teaching, all CREATE residents will also engage in regularly scheduled CBCT classes throughout their residency. CBCT is a research-based course developed at Emory University. The course "teaches mental exercises to promote a healthier response to stress and to strengthen empathic concern for others." These techniques are designed to help residents develop a greater awareness of their own attitudes and behaviors and how these attitudes and behaviors impact their relationship with others. Through progressive mental exercises and meditation, residents learn to focus their attention and become more aware of their thoughts and feelings as they occur in each moment. CBCT teaches educators how to respond to stressful situations in a healthy way, connect with others who are facing similar challenges, and strengthen their ability and willingness to express empathy and compassion for others (CREATE Teacher Residency Program, n.d.b.). The veteran educators at all CREATE schools will have the opportunity to engage in these trainings as well.

We present data on the following indicators related to CBCT.

- Indicator 1 (Years 1, 2, and 3): Program administrators offer at least one CBCT class per year to the general population of teachers at CREATE schools
- Indicator 2 (Years 1, 2, and 3): Residents attend CBCT classes

Results of Key Component 3 for Year 1

In Year 1, CREATE expected to hold at least one CBCT course for the general population of teachers at CREATE schools. CREATE exceeded this goal by offering three CBCT institutes during the 2015–16 school year and two CBCT institutes during the 2016–17 school year. Additionally, they expected 95% or more of the residents to attend at least 8 CBCT classes. They met this goal for both Cohort 1 and Cohort 2. Overall, both cohorts met fidelity for Component 3 in Year 1.

TABLE E7. DETAILS OF COMPONENT 3 IN YEAR 1

Indicator	Fidelity threshold	2015–16 school year	2016–17 school year
Indicator 1: Program administrators offer at least one CBCT course per year for general population of teachers at CREATE schools	0 = CREATE administrators offer 0 courses 1 = CREATE administrators offer 1 or more courses	CREATE administrators host 3 institutes Fidelity met	CREATE administrators host 2 institutes Fidelity met
Indicator	**Fidelity threshold**	**Cohort 1 met fidelity?**	**Cohort 2 met fidelity?**
Indicator 2: Residents attend CBCT classes	0 = Attend 0-7 meetings 1 = Attend 8-10 meetings 95% or more of residents meet fidelity at the unit level	18/19 (95%) Fidelity met	20/20 (100%) Fidelity met

TABLE E7. DETAILS OF COMPONENT 3 IN YEAR 1

Program level	Component score is sum of indicator scores. Fidelity not met = score of 0-1 Fidelity met = score of 2	Fidelity met	Fidelity met

Note. CBCT stands for Cognitively-Based Compassion Training

Results of Key Component 3 for Year 2

In Year 2, CREATE expected to hold at least one CBCT course for the general population of teachers at CREATE schools. CREATE exceeded this goal by offering two CBCT institutes during the 2017–18 school year. Additionally, they expected 95% or more of the residents to attend at least seven CBCT classes in Year 2. CREATE met this goal for Cohort 1, but did not meet this goal for Cohort 2. At the program level, Cohort 1 met fidelity, but Cohort 2 did not meet fidelity, for Component 3 in Year 2.

TABLE E8. DETAILS OF COMPONENT 3 IN YEAR 2

Indicator	Fidelity threshold	2016–17 school year	2017–18 school year
Indicator 1: Program administrators offer at least one CBCT class per year for general population of teachers at CREATE schools	0 = CREATE administrators offer 0 courses 1 = CREATE administrators offer 1 or more courses	CREATE administrators host 2 institutes Fidelity met	CREATE administrators host 2 institutes Fidelity met
Indicator	**Fidelity threshold**	**Cohort 1 met fidelity?**	**Cohort 2 met fidelity?**
Indicator 2: Residents attend CBCT classes	0 = Attend 0-6 meetings 1 = Attend 7+ meetings 95% or more of residents meet fidelity at the unit level	19/19 (100%) Fidelity met	12/15 (80%) Fidelity not met
Program level	Component score is sum of indicator scores. Fidelity not met = score of 0-1 Fidelity met = score of 2	Fidelity met	Fidelity not met

Note. CBCT stands for Cognitively-Based Compassion Training

Results of Key Component 3 for Year 3

In Year 3, CREATE expected to hold at least one CBCT course for the general population of teachers at CREATE schools. CREATE met this goal by offering one CBCT institute during the 2018–19 school year. Additionally, they expected 95% or more of the residents to attend at least seven CBCT classes in Year 3. CREATE met this goal for Cohorts 1 and 2. At the program level, Cohort 1 did not meet fidelity, but Cohort 2 did, for Component 3 in Year 3.

TABLE E9. DETAILS OF COMPONENT 3 IN YEAR 3

Indicator	Fidelity threshold	2017–18 school year	2018–19 school year
Indicator 1: **Program administrators offer at least one CBCT course per year for general population of teachers at CREATE schools**	0 = CREATE administrators offer 0 courses 1 = CREATE administrators offer 1 or more courses CREATE administrators offer 1 or more courses	CREATE administrators host 3 institutes Fidelity met	CREATE administrators host 1 institute Fidelity met
Indicator	**Fidelity threshold**	**Cohort 1 met fidelity?**	**Cohort 2 met fidelity?**
Indicator 2: **Residents attend CBCT classes**	0 = Attend 0-2 meetings 1 = Attend 3 meetings 95% or more of residents meet fidelity at the unit level	11/12 (92%) Fidelity not met	12/12 (100%) Fidelity met
Program level	Component score is sum of indicator scores. Fidelity not met = score of 0-1 Fidelity met = score of 2	Fidelity not met	Fidelity met

Note. CBCT stands for Cognitively-Based Compassion Training

Overall Results of Key Component 2 and 3

CREATE regularly offered more than one CF and CBCT institute a year, in which they achieved a high level of consistent attendance by veteran educators. They also meet fidelity for CF and CBCT meetings for both cohorts in Year 1. However, Cohort 1 did not meet fidelity for Together Time meetings in year 3, and Cohort 2 did not meet fidelity for Together Time meetings in year 2. During Year 1, residents are in their student teaching year and taking classes at GSU CEHD, which may make it easier for CREATE to maintain high attendance at these meetings. However, once the residents become full-time teachers, they are likely to face additional stressors and competing priorities, making it more difficult for CREATE to maintain the high level of attendance at these meetings that occurred in Year 1.

The CREATE program met fidelity on Component 2 for Cohort 1 in all three years and for Cohort 2 in years 1 and 3. FOI is met for Component 3 in all three years for both Cohorts 1 and 2, except for Cohort 2 in Year 2 and Cohort 1 in Year 3.

KEY COMPONENT 4: MULTIPLE FORMS OF MENTORING

CREATE equips experienced educators with the skills needed to guide CREATE residents through their first two years as full-time teachers. School-based mentors, who are veteran educators working the same school as their CREATE resident mentee, receive training from CREATE in how to best coach their residents in professional practice techniques, the development of a growth mindset, and skills for maintaining their well-being during the high stress situations that occur in the early days of a teacher's career. The training is delivered in several sessions beginning the summer prior to the start of the school and continuing throughout the year. Residents also receive instructional mentors (IMs) who are members of the CREATE staff trained in supporting residents in their development of compassion-based, equitable, effective teaching practices. Component 4 only applies to years 2 and 3. CREATE residents do not receive school-based mentors in Year 1, as they are still student teaching through GSU CEHD.

CREATE Key Component 4 covers the multiple forms of mentoring offered to CREATE residents. We present data on the following indicators related to mentorship.

- Indicator 1 (Years 2 and 3): Mentors attend training prior to mentoring
- Indicator 2 (Years 2 and 3): Mentors attend training during their mentoring year (at least 1 session)
- Indicator 3 (Years 2 and 3): Residents attend semi-monthly meetings with their mentor (SBM and IM)
- Indicator 4 (Years 2 and 3): Residents participate in mentor-resident observation cycles

Results of Key Component 4 for Year 2

CREATE expects that all mentors will attend a summer training prior to beginning their work as a mentor and that at least 90% of them will attend one or more trainings during the school year while they serve as mentors. Additionally, CREATE expects that 95% of residents meet with their mentors at least 28 times during the year and that 90% of them complete at least 2–3 observation cycles with their mentors during the year. CREATE met these goals for Cohort 1. Cohort 2 did not meet fidelity on indicators 2 and 3. Because all indicators must be met for Component 4, Cohort 2 does not meet fidelity on this component for Year 2.

TABLE E10. DETAILS OF COMPONENT 4 IN YEAR 2

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Indicator 1: **Residents have mentors who attend training prior to mentoring**	0 = Resident has a mentor who did not attend training 1 = Resident has a mentor who attended training 100% of residents have mentors that meets fidelity at the unit level	19/19 (100%) Fidelity met	15/15 (100%) Fidelity met

TABLE E10. DETAILS OF COMPONENT 4 IN YEAR 2

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Indicator 2: **Residents' mentors attend training during their mentoring year (1 sessions)**	0 = Resident has a mentor who attends 0 sessions 1 = Resident has a mentor attends 1 or more sessions 90% or more of residents have mentors that meets fidelity at the unit level	19/19 (100%) Fidelity met	15/15 (100%) Fidelity met
Indicator 3: **Residents attend semi-monthly meetings with their mentor (School-based mentor or Instructional Mentor)**	Cohort 1: 0 = Resident attends 0-8 meetings; 1 = Resident attends 9-11 meetings; 2 = Resident attends 12 or more meetings	18/19 (95%) receive score of 2 1/19 (5%) receive score of 1 None receive score of 0 Fidelity met	9/15 (60%) receive score of 2 None receive score of 1 6/15 (40%) receive a score of 0 Fidelity not met
	Cohort 2: 0 = Resident attends 0-22 meetings 1 = Resident attends 23-27 meetings 2 = Resident attends 28 or more meetings 95% of residents earn a score of 2 at the unit level and no residents earn a score of 0		
Indicator 4: **Residents participate in mentor-resident observation cycles**	0 = Resident participates in 0-1 cycles, 1 = Resident participates in 2-3 cycles 90% or more of residents meet fidelity at the unit level	19/19 (100%) Fidelity met	13/15 (87%) Fidelity not met
Program level	All indicators meet fidelity	Fidelity met	Fidelity not met

Note. There was one mentor who served two residents in Cohort 1 and one mentor who served three residents in Cohort 2. All other mentors served one resident each.

Results of Key Component 4 for Year 3

As mentioned above, CREATE expects that all mentors attend a summer training prior to beginning their work as a mentor and that at least 90% of them attend one or more trainings during the school year in which they serve as a mentor. In year 3, CREATE expects that 95% of residents will meet with their mentors at least 12 times during the year and that 80% of them complete at least 2–3 observation cycles with their mentors during the year. CREATE met these goals for Cohort 1. Cohort 2 did not meet fidelity on indicators 1 and 2 and, as a result, did not meet fidelity on this component in Year 3.

TABLE E11. DETAILS OF COMPONENT 4 IN YEAR 3

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Indicator 1: Residents have a mentor that attended training prior to mentoring	0 = Resident has a mentor that did not attend training 1 = Resident has a mentor that attended training 100% of residents have mentors that meets fidelity at the unit level	12/12 (100%) Fidelity met	6/12 (50%) Fidelity not met
Indicator 2: Residents have a mentor that attended training during their mentoring year (1 sessions)	0 = Resident has a mentor that attends 0 sessions 1 = Resident has a mentor that attends 1 or more sessions 90% or more of residents have mentors that meet fidelity at the unit level	11/12 (92%) Fidelity met	10/12 (83%) Fidelity not met
Indicator 3: Residents attend semi-monthly meetings with their mentor (School-based mentor or Instructional Mentor)	0 = Resident attends 0-8 meetings; 1 = Resident attends 9-11 meetings; 2 = Resident attends 12 or more meetings 95% of residents earn a score of 2 at the unit level and no residents earn a score of 0	12/12 (100%) receive a score of 2 Fidelity met	12/12 (100%) receive a score of 2 Fidelity met
Indicator 4: Resident participates in mentor-resident observation cycles	0 = Resident participates in 0-1 cycles, 1 = Resident participates in 2-3 cycles 80% or more of residents meet fidelity at the unit level	10/12 (83%) Fidelity met	12/12 (100%) Fidelity met
Program level	All indicators meet fidelity	Fidelity met	Fidelity not met

Note. Each mentor served one resident in Year 3.

KEY COMPONENT 5: SUMMER RESIDENT ACADEMY (SRA)

There were programmatic changes to the summer participation between Cohort 1 and Cohort 2 so the thresholds vary for this component by cohort. SRA is a 5-week intensive training for new teachers that occurs between the summer after residents graduate from GSU CEHD and before they begin their first year as full-time teachers. Activities that residents participate in during the academy include preparing lesson plans that will contribute to a safe and culturally responsive classroom, learning research-based instructional skills, practicing strategies to maintain personal and physical heath, among other skills and competencies they will need as a full-time teacher. We present data on the following indicator related to the summer internships and academy.

- Indicator 1: Residents attend SRA

CREATE expects that among Cohort 1 residents, 95% or more attend SRA for at least two weeks and 85% or more attend 3 weeks. CREATE expects that among Cohort 2 residents, 95% or more attend SRA for at least 15 days and 85% or more attend at least 20 days.

TABLE E12. DETAILS OF COMPONENT 5 IN YEAR 2

Indicator	Fidelity threshold	Cohort 1 met fidelity?	Cohort 2 met fidelity?
Residents attend Summer Resident Academy	Cohort 1 0=Resident does not enroll 1=Resident attends more than two week and less than 3 weeks 2=Resident attends for 3 weeks 95% or more of residents earn a score of at least 1 and 85% or more of residents earn a score of 2 at the unit level	18/19 (95%) earned a score of at least 1 15/19 (79%) earned a score of 2	15/15 (100%) of active residents receive a score of 2
	Cohort 2 0 = Resident attends less than 15 days 1 = Resident attends 15-20 days 2 = Resident attends more than 20 days 95% or more of residents earn a score of at least 1 and 85% or more of residents earn a score of 2 at the unit level		
Program level	95% or more of residents earn a score of at least 1 and 85% or more of residents earn a score of 2 at the unit level	Fidelity not met	Fidelity met

Note. While 20 Cohort 2 residents were active at the end of Year 1, 4 of them indicated they did not intend to return to CREATE. As a result, they were not expected to attend the Summer Resident Academy.

Overall Results of Key Component 5

Cohort 1 did not meet fidelity, but Cohort 2 did, for this component.

Appendix F. The Five Main Survey Outcome Scales and Moderator Scales

SCALE DETAILS

Self-efficacy in Teaching

Cronbach's Alpha = .81

We assessed self-efficacy in teaching using seven items from the PRIDE Teaching Environment Survey. Each item in the scale allowed four response options ranging from 'Not True at All' to 'Very True', which we coded as integers ranging from 1 to 4. We estimated the outcome for an individual by averaging that person's responses across the following seven items.

1. I know how to deliver instruction so that all my students can learn.
2. I have the ability to assess student learning problems.
3. When students in my class struggle, I have the expertise to use alternate teaching strategies.
4. I use many effective strategies to restore order in a classroom.
5. I use a number of effective strategies for motivating students to engage in their classwork.
6. Some students just cannot be motivated to do the work. [reverse coded.]
7. I give students an opportunity to make decisions about class activities.

A higher score on the scale means a participant self-reports that he or she knows better how to deliver instruction so that all his/her students can learn. For this and the other scales on which we assessed impacts, we recoded reverse-coded items before averaging responses across items.

Commitment to Teaching

Cronbach's Alpha = .82

We assessed commitment to teaching using four adapted items from the PRIDE Teaching Environment Survey. Each item in the scale allowed four response options ranging from 'Not true at all' to 'Very true', which we coded as integers ranging from 1 to 4. We estimated the outcome for an individual by averaging that person's responses across the following four items.

1. I have the same motivation now that I did when I started my program at GSU COE.
2. I question if teaching is right for me.
3. If I had to do it over, I would still want to become a teacher.
4. I still want to teach because I truly enjoy the work.

A higher score on the scale means a participant self-reports that he or she has a greater commitment to teaching.

Stress Management and Empathy Related to Teaching

Cronbach's Alpha = .92

We assessed stress management and empathy related to teaching using six items from a researcher-developed scale. Each item in the scale allowed five response options ranging from 'Strongly Disagree' to 'Strongly Agree', which we coded as integers ranging from 1 to 5. We estimated the outcome for an individual by averaging that person's responses across the following six items.

1. Since starting this school year, I have learned techniques that enable me to manage my reactions in a healthy, constructive way when faced with stressful situations.
2. Since starting this school year, I feel more confident that I will be able to handle the stress of being a teacher.
3. Since starting this school year, I feel more confident identifying and advocating for my own professional needs.
4. Since starting this school year, I have increased my commitment to helping my students learn and thrive.
5. Since starting this school year, I feel more able to understand the perspective of my students.
6. Since starting this school year, I feel more able to understand the perspective of my fellow teachers/colleagues.

A higher score on the scale means a participant self-reports that he or she is more empathetic and effective at managing stress.

Resilience (CD-RISC)

Cronbach's Alpha = .91

We assessed teacher resilience using the CD-RISC 10. Each item in the scale allowed response options from "Not true at all" to "True nearly all the time" which we coded as integers ranging from 0 to 4. We estimated the outcome for an individual by averaging that person's responses across 10 items. The scale is proprietary, and therefore, the specific items composing the scale are not included here. A higher score on the scale means a participant self-reports that he or she has greater resilience generally (i.e., the items are not specific to resilience as related to teaching).

Mindfulness

Cronbach's Alpha = .68

We assessed levels of emotional regulation using 12 items from the Five Facets Mindfulness Questionnaire. Each item in the scale allowed five response options ranging from 'Never of Rarely True' to 'Very Often to Always True', which we coded as integers ranging from 1 to 5. We reverse-coded negatively phrased items. We estimated the outcome for an individual by averaging that person's responses across the following 12 items.

1. I criticize myself for having irrational or inappropriate emotions.
2. I perceive my feelings and emotions without having to react to them.
3. I don't pay attention to what I'm doing because I'm daydreaming, worrying, or otherwise distracted.
4. I pay attention to sensations, such as the wind in my hair or sun on my face.
5. I make judgments about whether my thoughts are good or bad.
6. I find it difficult to stay focused on what's happening in the present.

7. I pay attention to sounds, such as clocks ticking, birds chirping, or cars passing.
8. In difficult situations, I can pause without immediately reacting.
9. It seems I am "running on automatic" without much awareness of what I'm doing.
10. I tell myself that I shouldn't be thinking the way I'm thinking.
11. When I have distressing thoughts or images, I just notice them and let them go.
12. I pay attention to how my emotions affect my thoughts and behavior.

A higher score on the scale means a participant regulates his or her own feelings more effectively.

Scales Used in Moderator Analysis

We assessed differential impacts using 5 moderators collected via teacher surveys: whether the teacher is Black, current GPA, and scales on confidence in general teaching skills, motivation for entering teaching, and math anxiety. We provide more information on the three scales below.

Confidence in General Teaching Skills

Cronbach's Alpha = .81

We assessed *Confidence in General Teaching Skills* using an evaluator-developed scale. Each item in the scale allowed five response options from "Not at all confident" to "Extremely confident" which we coded as integers ranging from 1 to 5. We estimated the score for an individual by averaging that person's responses across 8 items. A higher score on the scale means a participant self-reports that he or she has greater confidence in general teaching skills.

1. How confident are you in your abilities within each of the following areas?
    a. Collaborating with others
    b. Reflecting on my actions
    c. Planning and organization
    d. Oral communication
    e. Written communication
    f. Working with children
    g. Working with parents
    h. Classroom management

Motivation for Entering Teaching Profession

Cronbach's Alpha: Alpha=.51

We assessed *Motivation for Entering Teaching* using an evaluator-developed scale, which consisted of 5 items. Each item allowed five response options from "Strongly disagree" to "Strongly agree" which we coded as integers ranging from 1 to 5. We estimated the score for an individual by averaging that person's responses across the 5 items. A higher score on the scale means a participant self-reports that he or she has greater motivation in entering teaching.

2. To what extent do you agree with each of the following statements about your motivation for entering the teaching profession?
    a. I am interested in making a difference in the lives of children in Atlanta.
    b. I believe that as a teacher I will contribute to social justice.
    c. I am interested in how students learn.
    d. I look at teaching as a process of self-discovery.
    e. I am entering teaching because I want to eventually mentor teachers.

Math Anxiety

Cronbach's Alpha = .96

We assessed *Math Anxiety* using an evaluator-developed scale. Each item in the scale allowed five response options from "Not at all" to "Very much" which we coded as integers ranging from 1 to 5. We estimated the score for an individual by averaging that person's responses across 10 items about their anxiety related to math coursework. A higher score on the scale means a participant self-reports that he or she has greater math anxiety.

Current GPA

What is your current GPA?

Appendix G. Confirmatory Analysis of Impact on Teacher TAPS Ratings

ATTRITION IN TEACHER SAMPLE FOR TAPS ANALYSIS

The final sample of study participants included in the analysis was dependent on a series of criteria being met with regard to consent for data collection and available data from the GaDOE. Tables G1 and G2 outline reasons for attrition in the analytic sample for TAPS.

TABLE G1. ATTRITION FROM TAPS SAMPLE, COHORT 1

Event/reason for attrition	CREATE group	Comparison group
Agreed to participate in study and were eligible to participate at the time of consent	19	58
Active at time of TAPS consent request	19	43
Agreed to have researchers collect TAPS data from GaDOE [a]	14	24
Requested TAPS data	14	24
Received TAPS data	12	15
Active at time of Intern Keys consent request (January 2018)[b]	13	15
Agreed to have researchers collect Intern Keys data from GSU	9	13
Requested Intern Keys data	9	13
Received Intern Keys data for the semester we needed	9	12
Had score in Instructional Practices TAPS dimension	12	15
Had score in Positive Learning Environment TAPS dimension	12	15
Received both Intern Keys data + TAPS data	6	10
Remained in sample after matching (Instructional Strategies)	6	9
Remained in sample after matching (Quality of Learning Environment)	6	10

[a] During negotiations for the data sharing agreement with GaDOE, we were informed that we needed more specific language in our study participant consent forms in order for GaDOE to release the teacher and student data. Therefore, we had to obtain additional consent from participants to collect this data. This resulted in a reduced sample with necessary data for analysis. In CREATE, five teachers did not agree to release TAPS data, reducing the sample from 19 to 14. In the comparison group of 43 individuals, two did not agree to release TAPS, and 17 were either never sent the consent form due to having previously left the study, or we lacked school or district permissions, or they never responded to our request in spite of repeated follow-up, reducing the sample from 43 to 24.

[b] Intern Keys were added as a baseline measure after participants signed original study consent forms. During negotiations for the data sharing agreement with GSU, we were informed that we needed specific language in our study participant consent forms in order for GSU to release Intern Keys data. Therefore, we had to obtain additional consent from participants to collect these data. This further reduced the sample with data needed for analysis.

TABLE G2. ATTRITION FROM TAPS SAMPLE, COHORT 2

Event/Reason for attrition	CREATE group	Comparison group
Agreed to participate in study and were eligible to participate at the time of consent	23	40
Requested TAPS data	23	40
Received TAPS data	14	27
Active at time of Intern Keys consent request[a]	15	18
Agreed to have researchers collect Intern Keys data from GSU	13	13
Requested Intern Keys data (January 2018)	13	13
Received Intern Keys data for the semester we needed	13	7
Had score in Instructional Practices TAPS dimension	14	27
Had score in Positive Learning Environment TAPS dimension	14	27
Received both Intern Keys and TAPS data	9	6
Remained in sample after matching (Instructional Strategies)	8	4
Remained in sample after matching (Quality of Learning Environment)	7	6

[a] Intern Keys were added as a baseline measure after participants signed original study consent forms. During negotiations for the data sharing agreement with GSU, we were informed that we needed specific language in our study participant consent forms in order for GSU to release Intern Keys data. Therefore, we had to obtain additional consent from participants to collect these data. This further reduced the sample with data needed for analysis.

DESCRIPTION OF IMPACT ANALYSIS

Through the matching strategies, we arrived at the analytic samples for assessing the confirmatory impacts of CREATE on the two dimensions of TAPS. We used a teacher-level linear regression model to estimate impacts. We regressed the TAPS outcome against the Intern Keys (baseline measure) and other covariates. The relatively small sample size of teachers limited the number of covariates that can be used in analysis. That is, with only 27 or 29 teachers (depending on the performance standard of TAPS), we were concerned with including too many covariates in analysis and possibly leading

to problems with multicollinearity that could bias the impact estimate. At the same time, realizing that this is a quasi-experimental analysis, it was important to account for as many confounders of treatment as possible to limit selection bias.

To address this problem of selecting an optimal number of covariates that limits potential for bias from both model overspecification, as well as selection, we used three methods of covariate selection: (1) an adaptation of a method by Hosmer and Lemeshow (2000) (Method 1), (2) like the first method but with more liberal settings of the parameters used to decide covariate inclusion (Method 2, which supports the benchmark impact model), and (3) a standard algorithmic forward selection approach (Method 3). We describe each briefly below.

Method 1

1. Assume we have outcome y_j for teacher j and covariates $treatment_j$, X_{1j},….X_{nj} (where treatment indicates condition $treatment_j = 1$ for CREATE, and $treatment_j = 0$ for comparison).
2. Regress y_j against each covariate individually: $y_j = \beta_0 + \beta_1 X_{kj} + e_j$. If the p value for the effect of X_{kj} is less than .25, then retain X_{kj} for inclusion in the multivariate model.
3. Include the treatment variable and all covariates in (2) that satisfy the p value criterion in the multivariate model:

   $y_j = \beta_0 + \beta_1 treatment_j + \sum_{k=1}^{N} \beta_k X_{kj} + e_j.$

4. Take note all of the covariates in (3) that are significant at the $p < .20$ level.
5. Check for confounding by running models removing one covariate at a time:

   $y_j = \beta_0 + \beta_1 treatment_j + \sum_{k \neq 1} \beta_k X_{kj} + e_j.$
   $y_j = \beta_0 + \beta_1 treatment_j + \sum_{k \neq 2} \beta_k X_{kj} + e_j.$
   …
   $y_j = \beta_0 + \beta_1 treatment_j + \sum_{k \neq N} \beta_k X_{kj} + e_j.$

6. If removing a covariate causes the coefficient for $treatment_j$ to change more than 20% of its value, then designate it as confounded with treatment.
7. Remove from the multivariate model in (3) covariates that are both non-significant in (4) and not confounders in (6).
8. Add to the model specified in (7) each of the covariates excluded in (2). If any are significant at level $p < .10$, then add them to the final model.

Method 2 (Benchmark Confirmatory)

This is like Model 1 but with $treatment_j$ and the baseline measure forced into the model, and with the setting in Step 8 changed to $p < .20$.

Method 3

We used a standard forward selection method that algorithmically includes covariates one at a time and selects a combination that maximizes the variability in outcomes accounted for. The baseline and the variable indicating treatment assignment status were forced into the model.

ESTABLISHING BASELINE EQUIVALENCE

Before conducting the impact analysis, we examined baseline equivalence for the available sample. We tested the difference between treatment and comparison on the baseline (separately for the final analytic sample associated with analysis of impact on each performance standard). With each test, we regressed the baseline against a dummy variable indicating condition (CREATE or comparison) and a dummy variable for cohort, but without covariates. The form of the model has the same structure for fixed and random error terms that we apply in the benchmark impact model. We also tested equivalence separately by cohort.

For Quality of Instructional Strategies, after limiting the sample to cases with baseline and outcome ratings and with non-missing values of covariates used in the impact analysis (N = 27), we observed a difference of = -0.073 in baseline equivalence, expressed in effect size units across the two cohorts. Given that this baseline effect size is within the adjustment range for the WWC (difference between 0.05 and 0.25 standard deviations), we included the baseline as a covariate in the impact analysis model to establish baseline equivalence. Full descriptives with adjusted and unadjusted means and corresponding effect sizes are in Table G3.

TABLE G3. BASELINE EQUIVALENCE FOR THE INTERN KEYS (BASELINE) FOR THE SAMPLE USED TO EVALUATE IMPACT ON QUALITY OF INSTRUCTIONAL STRATEGIES

	CREATE			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	SD	Sample size	Unadjusted mean	SD	CREATE – comparison	ES
Intern Keys (model-based approach)	14	2.64	0.74	13	2.69	0.63	-0.052	-0.073
Unadjusted approach	14	2.64	0.74	13	2.69	0.63	-0.05	-0.073

For Positive Learning Environment, after limiting the sample to cases with baseline and outcome ratings and with non-missing values of covariates (N = 31), we observed an effect size difference of 0.259. The difference for this baseline performance standard is large enough that the results would not satisfy the baseline equivalence requirement. Therefore, we used a rudimentary matching strategy, described below, to improve baseline equivalence.

Matching on the Positive Learning Environment Baseline Measure

Given that our analytic sample of teachers in each condition is small to start with, we focused on a basic matching strategy that would preserve the sample size and directly address baseline equivalence on the baseline measure. We noticed that for Cohort 2, no comparison cases had a baseline value of "2" on the baseline measure (observed responses included values 3 and 4 only); whereas in treatment, there were some teachers who achieved ratings for each of three baseline values (2, 3, and 4). By removing two of the treatment teachers who achieved a baseline rating of "2", we were able to even out the distribution of the baseline ratings resulting in a baseline difference in the Positive Learning Environment baseline with a value for the adjusted effect size of -0.192 (with final N = 29).

The size of the baseline difference between the treatment and comparison groups for the Positive Learning Environment baseline ratings requires that we adjust for the baseline ratings in the impact analysis to offset possible biasing effects. Full descriptives with adjusted and unadjusted means and corresponding effect sizes are in Table G4.

TABLE G4. BASELINE EQUIVALENCE FOR THE INTERN KEYS (BASELINE) FOR THE SAMPLE USED TO EVALUATE IMPACT ON POSITIVE LEARNING ENVIRONMENT STRATEGIES (OUTCOME)

	CREATE			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	SD	Sample size	Unadjusted mean	SD	CREATE – comparison	ES
Intern Keys (model-based approach)	13	2.78	0.555	16	2.88	0.500	-0.10	-0.192
Unadjusted approach	13	2.85	0.555	16	2.88	0.500	-0.03	-0.057

SAMPLE SIZES AND DETAILED IMPACT RESULTS

Impacts on Instructional Strategies

Table G5 provides case counts for the benchmark analysis using Model 2 and sensitivity analysis using Model 3. Table G6 provides case counts for the sensitivity analysis using Model 1. Table G7 provides detailed results for the benchmark and sensitivity analyses.

TABLE G5. COUNTS FOR EACH RATING ON THE INTERN KEYS AND TAPS INSTRUCTIONAL STRATEGIES, MODEL 2 (BENCHMARK) AND MODEL 3 (*N* = 27)

	CREATE	Comparison
Counts per condition and cohort		
Cohort 1	6	9
Cohort 2	8	4
Total	14	13
Counts for each response option (cohorts 1 and 2 combined) on the Intern Keys (baseline)		
Level 1	0	0
Level 2	7	5
Level 3	5	7
Level 4	2	1
Counts for each response option (cohorts 1 and 2 combined) on the TAPS Instructional Strategies (outcome)		
Level 0	0	0
Level 1	2	1

TABLE G5. COUNTS FOR EACH RATING ON THE INTERN KEYS AND TAPS INSTRUCTIONAL STRATEGIES, MODEL 2 (BENCHMARK) AND MODEL 3 (*N* = 27)

	CREATE	Comparison
Level 2	12	12
Level 3	0	0

Note.
Intern Key ratings: Level I = Ineffective; Level 2 = Needs Development; Level 3 = Proficient; Level 4 = Exemplary

TAPS ratings: Level 0 = Emerging; Level I = Developing; Level 2 = Proficient; Level 3 = Advanced

TABLE G6. COUNTS FOR EACH RATING ON THE INTERN KEYS AND TAPS INSTRUCTIONAL STRATEGIES, MODEL 1 (*N* = 28)

	CREATE	Comparison
Counts per condition and cohort		
Cohort 1	6	10
Cohort 2	8	4
Total	14	14
Counts for each response option (cohorts 1 and 2 combined) on the Intern Keys (baseline)		
Level 1	0	0
Level 2	7	6
Level 3	5	7
Level 4	2	1
Counts for each response option (cohorts 1 and 2 combined) on the TAPS Instructional Strategies (outcome)		
Level 0	0	0
Level 1	2	1
Level 2	12	12
Level 3	0	1

Note.
Intern Key ratings: Level I = Ineffective; Level 2 = Needs Development; Level 3 = Proficient; Level 4 = Exemplary

TAPS ratings: Level 0 = Emerging; Level I = Developing; Level 2 = Proficient; Level 3 = Advanced

TABLE G7. IMPACTS ON INSTRUCTIONAL STRATEGIES

Covariates	Method 1 for covariate selection (*N* = 28) *n* (CREATE) = 14 *n* (comparison) = 14	Method 2 for covariate selection (*N* = 27) (Benchmark Model) *n* (CREATE) = 14 *n* (comparison) = 13	Method 3 for covariate selection (*N* = 27) (Forward selection) *n* (CREATE) = 14 *n* (comparison) = 13
Intercept	-0.362 (0.745) *df* = 23 *t* = -.49 *p* = .632	-0.948 (0.523) *df* = 20 *t* = -1.81 *p* = .085	-1.095 (0.512) *df* = 7 F = 4.57 *p* = .047
Baseline	-0.053 (0.094) *df* = 23 *t* = .579 *p* = .579	-0.045 (0.075) *df* = 20 *t* = -.59 *p* = .559	-0.062, (0.074) *df* = 7 F = .72 *p* = .409
Condition (CREATE minus comparison)	-0.106 (0.123) *df* = 23 *t* = .396 *p* = .396	-0.110 (0.086) *df* = 20 *t* = -1.26 *p* = .221	-0.151 (0.091) *df* = 7 F = 2.77 *p* = .114
High school GPA	0.298 (0.133) *df* = 23 *t* = 2.24 *p* = .035	0.421 (0.104) *df* = 20 *t* = 4.06 *p* = .001	0.408 (0.100) *df* = 7 F = 16.77 *p* < .001
Agrees with teaching as a process of self-discovery (survey item)	0.325 (0.130) *df* = 23 *t* = 2.49 *p* = .030	0.277 (0.089) *df* = 20 *t* = 3.11 *p* = .006	0.297 (0.089) *df* = 7 F = 11.23 *p* = .004
Belongs to Cohort 2		0.211 (0.087) *df* = 20 *t* = 2.42 *p* = .025	0.266 (0.092) *df* = 7 F = 8.38 *p* = .010
Gender is male		0.231 (0.144) *df* = 20 *t* = 1.60 *p* = .125	0.266 (0.139) *df* = 7 F = 3.65 *p* = .072
Teacher ethnicity: Black			0.172 (0.087) *df* = 7 F = 3.90 *p* = .064
Effect size (Impact)[a]	ES = -0.106 / 0.377 = -0.281	ES = -0.110 / 0.324 = -0.339	ES = -0.151/ 0.324 = -0.466
Baseline equivalence 2 (with cohort effect)	ES = 0.019	ES = -0.073	ES = -0.073

[a] All effect sizes are the regression adjusted estimate for condition (CREATE or comparison) from the impact model (baseline equivalence or impact) divided by the pooled standard deviation of the outcome variable.

Impacts on Positive Learning Environment

Table G8 provides case counts for all analyses. Table G9 provides detailed results for the benchmark and sensitivity analyses.

TABLE G8. COUNTS FOR EACH RATING ON THE INTERN KEYS AND TAPS POSITIVE LEARNING ENVIRONMENT (N= 29)

	CREATE	Comparison
Counts per condition and cohort		
Cohort 1	6	10
Cohort 2	7	6
Total	13	16
Counts for each response option (cohorts 1 and 2 combined) on the Intern Keys (baseline)		
Level 1	0	0
Level 2	3	3
Level 3	9	12
Level 4	1	1
Counts for each response option (cohorts 1 and 2 combined) on the TAPS Instructional Strategies (outcome)		
Level 0	0	0
Level 1	1	0
Level 2	12	14
Level 3	0	2

Note.
Intern Key ratings: Level I = Ineffective; Level II = Needs Development; Level III = Proficient; Level IV = Exemplary

TAPS ratings: Level 0 = Emerging; Level I = Developing; Level II = Proficient; Level III = Advanced

TABLE G9. IMPACTS ON POSITIVE LEARNING ENVIRONMENT

Covariates	Method 1 for covariate selection (*N* = 29) *n* (CREATE) = 13 *n* (comparison) = 16	Method 2 for covariate selection (*N* = 29) (Benchmark Model) *n* (CREATE) = 13 *n* (comparison) = 16	Method 3 for covariate selection (*N* = 29) (Forward selection) *n* (CREATE) = 13 *n* (comparison) = 16
Intercept	1.008 (0.715) *df* = 25 *t* = 1.41 *p* = .171	0.636 (0.807) *df* = 24 *t* = 0.79 *p* = .438	0.120 (0.695) *df* = 4 F = .03 *p* = .865
Baseline	-0.026 (0.116) *df* = 25 *t* = -.220 *p* = .826	-0.026 (0.116) *df* = 24 *t* = -.23 *p* = .821	0.215 (0.118) *df* = 4 F = 3.30 *p* = .085
Condition	-0.192 (0.116) *df* = 25 *t* = -1.660 *p* = .110	-0.175 (0.117) *df* = 24 *t* = -1.50 *p* = .192	-0.077 (0.108) *df* = 4 F = .50 *p* = .488
Has dual certification			0.289 (0.126) *df* = 4 F = 5.24 *p* = .034
Scale on motivation to contribute to social justice	0.247 (0.145) *df* = 25 *t* = 1.71 *p* = .100	0.203 (0.151) *df* = 24 *t* = 1.34 *p* = .192	0.247 (0.143) *df* = 4 F = 2.94 *p* = .103
Effect size (Impact)[a]	- 0.192 / 0.314 = -0.611	-0.175 / 0.314 = -0.557	-0.077 / 0.314 = -0.245
Baseline equivalence 2 (with cohort effect)	ES = -0.192	ES = -0.192	ES = -0.192

[a] All effect sizes are the regression adjusted estimate for condition (CREATE or comparison) from the impact model (baseline equivalence or impact) divided by the pooled standard deviation of the outcome variable.

Appendix H. Confirmatory Analysis of Impact on Student Achievement

RATES OF ATTRITION IN SAMPLE FOR STUDENT ACHIEVEMENT ANALYSIS

The final sample of study participants included in this analysis depended on a series of factors being met with regard to consent for data collection and available data from the GaDOE. Tables H1 and H2 outline reasons for attrition in the analytic sample for the analysis of impact on student achievement.

TABLE H1. ATTRITION FROM SAMPLE FOR ANALYSIS OF IMPACT ON STUDENT ACHIEVEMENT, COHORT 1

Event/reason for attrition	Number of teachers in CREATE group	Number of teachers in comparison group
Agreed to participate in study and were eligible to participate at the time of consent	19	58
Active at time researchers requested teachers' consent to collect Milestones data from GaDOE (Fall 2016)	19	43[b]
Agreed to have researchers collect Milestones data from GaDOE	14[a]	24
Had any Milestones data in any subject	8	16
Had Milestones data in math	3	10
Had Milestones data in ELA	3	10
Had students with pretest and posttest in math	2	7
Had students with pretest and posttest in ELA	2	6
Remained in sample after matching (final sample for math and ELA)	1	2

[a] During negotiations for the data sharing agreement with GaDOE, we were informed that we needed more specific language in our study participant consent forms in order for GaDOE to release the teacher and student data. Therefore, we had to obtain additional consent from participants to collect this data. This resulted in a reduced sample with necessary data for analysis. 5 of the 19 CREATE teachers did not agree to release Milestones scores from GaDOE, resulting in 14 for whom we requested outcomes.

[b] Of the 58 comparison teachers who agreed to the study, 7 dropped out of GSU, 3 changed their minds about participating in the study or were unresponsive, 3 were ineligible to continue, and for 2, we were unable to get district/school permissions to obtain outcomes. This resulted in 43 comparison teachers who were active at the time we requested consent to obtain student Milestones scores.

TABLE H2. ATTRITION FROM SAMPLE FOR ANALYSIS OF IMPACT ON STUDENT ACHIEVEMENT, COHORT 2

Event/Reason for attrition	Number of teachers in CREATE group	Number of teachers in comparison group
Agreed to participate in study and were eligible to participate at the time of consent	23	40
Requested Milestones data	23	40
Received Milestones data	12	30
Had Milestones data in math	9	12
Had Milestones data in ELA	9	23
Had students with pretest and posttest in math	5	3
Had students with pretest and posttest in ELA	5	12
Remained in sample after matching (final sample for math)	0	0
Remained in sample after matching (final sample for ELA)	4	7

DETAILS OF HOW WE ARRIVED AT THE FINAL ANALYTIC SAMPLES FOR ASSESSING CONFIRMATORY IMPACTS

Student ELA Achievement

Tables H3 and H4 detail student and teacher sample sizes for assessing confirmatory impacts on ELA achievement. We start with all students linked to teachers and for whom we obtained outcomes. We then limit cases to those with pre- and posttests. Using the criterion that we must match cases within-cohort and within-grade, available cases are limited to 7th grade in Cohort 1, and 4th and 6th grades in Cohort 2. In Cohort 1 seventh-grade sample, 52 students are retained in each condition, following matching. In Cohort 2, in the 4th grade sample, 49 students are retained in each condition, following matching. In 6th grade, 121 students are retained in each condition, following matching, for a sum of 170 students in each condition. Summing across cohorts, we arrive at the total of 52 + 170 = 222 students in each condition for the confirmatory analysis of impact on ELA achievement.

TABLE H3. COUNTS OF STUDENTS AVAILABLE FOR MATCHING TO ASSESS IMPACTS ON ELA ACHIEVEMENT (COHORT 1)

	Number of students			Number of teachers		
	All	Comparison	CREATE	All	Comparison	CREATE
Cases for whom we obtained outcomes linkable to teachers	929	807	122	13	10	3
With pretest and posttest	620	522	98	8	6	2
Grade 4[a]	25	25	0	2	2	0
Grade 5[a]	20	0	20	1	0	1
Grade 6[a]	232	232	0	1	1	0
Grade 7	195	117 (52[b])	78 (52[b])	3	2	1
Grade 8[a]	148	148	0	2	2	0

[a] Samples at grade level include CREATE and comparison cases. We conduct within-grade matching, and therefore, this grade level is removed from analysis.
[b] Available for impact analysis after matching of students

TABLE H4. COUNTS OF STUDENTS AVAILABLE FOR MATCHING TO ASSESS IMPACTS ON ELA ACHIEVEMENT (COHORT 2)

	Number of students			Number of teachers		
	All	Comparison	CREATE	All	Comparison	CREATE
Cases for whom we obtained outcomes linkable to teachers	1728	1401	327	31	22	9
With pretest and posttest	1043	872	171	17	12	5
Grade 3[a]	1	0	1	1	0	1
Grade 4	106	57 (49[b])	49 (49[b])	5	2	3
Grade 5[a]	23	23	0	1	1	0
Grade 6	471	350 (121[b])	121 (121[b])	6	5	1
Grade 7[a]	250	250	0	3	3	0
Grade 8[a]	192	192	0	2	2	0

[a] Samples at grade level include CREATE and comparison cases. We conduct within-grade matching, and therefore, this grade level is removed from analysis.
[b] Available for impact analysis after matching of students

Student Mathematics Achievement

Tables H5 and H6 detail student and teacher sample sizes for assessing confirmatory impacts on math achievement. We start with all students linked to teachers and for whom we've obtained outcomes. We then limit cases to those with pre- and posttests. Using the criterion of within-cohort and within-grade matching, the available cases include only 7th grade in Cohort 1, and no cases from Cohort 2. Following matching, we retain 52 cases in each condition, for a total student sample size of 104.

TABLE H5. COUNTS OF STUDENTS AVAILABLE FOR MATCHING TO ASSESS IMPACTS ON MATHEMATICS ACHIEVEMENT (COHORT 1)

	Number of students			Number of teachers		
	All	Comparison	CREATE	All	Comparison	CREATE
Cases for whom we obtained outcomes linkable to teachers	648	518	130	13	10	3
With pretest and posttest	516	410	106	9	7	2
Grade 4[a]	25	25	0	2	2	0
Grade 5[a]	21	1	20	2	1	1
Grade 6[a]	268	268	0	4	4	0
Grade 7	202	116 (52[b])	86 (52[b])	3	2	1

[a] Samples at grade level include CREATE and comparison cases. We conduct within-grade matching, and therefore, this grade level is removed from analysis.

[b] Available for impact analysis after matching of students

TABLE H6. COUNTS OF STUDENTS AVAILABLE FOR MATCHING TO ASSESS IMPACTS ON MATHEMATICS ACHIEVEMENT (COHORT 2)

	Number of students			Number of teachers		
	All	Comparison	CREATE	All	Comparison	CREATE
Cases for whom we obtained outcomes linkable to teachers	715	521	194	21	12	9
With pretest and posttest	407	271	136	8	3	5
Grade 3[a]	1	0	1	1	0	1
Grade 4[a]	48	0	48	3	0	3
Grade 5[a]	24	24	0	1	1	0
Grade 6[a]	87	0	87	1	0	1
Grade 7[a]	247	247	0	2	2	0

[a] Samples at grade level include CREATE and comparison cases. We conduct within-grade matching, and therefore, this grade level is removed from analysis. (No students could be matched within-grade for Cohort 2.)

ESTABLISHING BASELINE EQUIVALENCE

To determine baseline equivalence, we regressed the pretest against the indicator of treatment status, a dummy variable indicating cohort (where possible), and the same random effects as in the impact model. For confirmatory analyses, students of teachers in the CREATE and comparison groups were equivalent at baseline for the analysis of impact on ELA (ES = -0.06 standard deviations; see Table H7), on math (ES = .05 standard deviations; see Table H8), and on math and ELA combined (ES = -0.08 standard deviations; see Table H9).

For exploratory analyses, in which we allowed students of CREATE and comparison group teachers to be matched within grades and across cohorts, baseline equivalence is achieved for samples associated with analysis of impact on ELA (ES = 0.10 standard deviations; see Table H10) on math (ES = 0.11 standard deviations; see Table H11), and both subjects combined (ES = 0.12 standard deviations; see Table H12).

TABLE H7. TEST OF BASELINE EQUIVALENCE ON THE ELA PRETEST FOR THE ANALYTIC SAMPLE USED TO EVALUATE CONFIRMATORY IMPACT OF CREATE ON STUDENT ELA ACHIEVEMENT

	Comparison		CREATE				
	Mean	Count	Mean	Count	Effect size	*p* value	Imbalance
Pretest	-0.16	222	-0.21	222	-0.06	.684	No

Note. Means are the raw means reported in scale score units. The effect size is the regression adjusted difference between conditions on the pretest, figuring in the fixed effect for cohort and teacher random effects, divided by the pooled standard deviation (sd = .99). CREATE stands for the group of students in classes of CREATE teachers.

TABLE H8. TEST OF BASELINE EQUIVALENCE ON THE MATH PRETEST FOR THE ANALYTIC SAMPLE USED TO EVALUATE CONFIRMATORY IMPACT OF CREATE ON STUDENT MATH ACHIEVEMENT

	Comparison		CREATE				
	Mean	Count	Mean	Count	Effect size	*p* value	Imbalance
Pretest	-0.48	52	-0.67	52	0.05	.939	No

Note. Means are the raw means reported in scale score units. The effect size is the regression adjusted difference between conditions on the pretest, figuring in the fixed effect for cohort and teacher random effects, divided by the pooled standard deviation (sd = .77). CREATE stands for the group of students in classes of CREATE teachers.

TABLE H9. TEST OF BASELINE EQUIVALENCE ON THE PRETEST FOR THE ANALYTIC SAMPLE USED TO EVALUATE CONFIRMATORY IMPACT OF CREATE POOLED ACROSS ELA AND MATH ACHIEVEMENT OUTCOMES

	Comparison		CREATE				
	Mean	Count	Mean	Count	Effect size	*p* value	Imbalance
Pretest	-0.22	274	-0.29	274	-0.08	.651	No

Note. Means are the raw means reported in scale score units. The effect size is the regression adjusted difference between conditions on the pretest, figuring in the fixed effect for cohort and teacher random effects, divided by the pooled standard deviation (sd = .96). CREATE stands for the group of students in classes of CREATE teachers.

Thirty students in the treatment condition (7th grade Cohort 1) yield both math and ELA pretest scores (i.e., there are 518 unique student IDs). Repeated measures for these individuals were accounted for in the model used to assess baseline equivalence.

TABLE H10. TEST OF BASELINE EQUIVALENCE FOR THE ANALYTIC SAMPLE USED TO EVALUATE EXPLORATORY IMPACT OF CREATE ON STUDENT ELA ACHIEVEMENT

	Comparison		CREATE				
	Mean	Count	Mean	Count	Effect size	*p* value	Imbalance
Pretest	-0.26	252	-0.24	252	0.10	.611	No

Note. Means are the raw means reported in scale score units. The effect size is the regression adjusted difference between conditions on the pretest, figuring in the fixed effect for cohort and teacher random effects, divided by the pooled standard deviation (sd = 1.04). CREATE stands for the group of students in classes of CREATE teachers.

TABLE H11. TEST OF BASELINE EQUIVALENCE FOR THE ANALYTIC SAMPLE USED TO EVALUATE EXPLORATORY IMPACT OF CREATE ON STUDENT MATH ACHIEVEMENT

	Comparison		CREATE				
	Mean	Count	Mean	Count	Effect size	*p* value	Imbalance
Pretest	-0.24	158	-0.28	158	0.11	.629	No

Note. Means are the raw means reported in scale score units. The effect size is the regression adjusted difference between conditions on the pretest, figuring in the fixed effect for cohort and teacher random effects, divided by the pooled standard deviation (sd = 1.02). CREATE stands for the group of students in classes of CREATE teachers.

TABLE H12. TEST OF BASELINE EQUIVALENCE FOR THE ANALYTIC SAMPLE USED TO EVALUATE EXPLORATORY IMPACT OF CREATE POOLED ACROSS ELA AND MATH ACHIEVEMENT OUTCOMES

	Comparison		CREATE				
	Mean	Count	Mean	Count	Effect size	*p* value	Imbalance
Pretest	-0.25	410	-0.26	410	0.12	.497	No

Note. Means are the raw means reported in scale sore units. The effect size is the regression adjusted difference between conditions on the pretest, figuring in the fixed effect for cohort and teacher random effects, divided by the pooled standard deviation (sd = 1.03). CREATE stands for the group of students in classes of CREATE teachers.

95 students in the treatment condition and 13 students in the comparison group yield both math and ELA pretest scores (i.e., there are 712 unique student ID's). Repeated measures for these individuals were accounted for in the model used to assess baseline equivalence.

FULL IMPACT RESULTS FOR CONFIRMATORY ANALYSIS OF STUDENT ACHIEVEMENT

Results of the Benchmark Analysis of Impacts on ELA (*N* = 14 Teachers, *N* = 444 Students)

TABLE H13. ESTIMATES OF FIXED EFFECTS

Fixed effects	Coefficient	Standard error	*df*	t-ratio	*p* value
Intercept	0.9875	0.1787	11	5.53	.0002
Belongs to cohort 1	0.3103	0.1039	422	2.99	.0030
Treatment (belongs to CREATE)	-0.1181	0.1520	11	-0.78	.4535
Pretest	0.7336	0.03677	422	19.95	<.0001
Gender is male	-0.1035	0.06113	422	-1.69	.0913
Ethnicity is Black	-0.8373	0.07075	422	-11.83	<.0001
Ethnicity is Hispanic	-0.6374	0.1150	422	-5.54	<.0001
Ethnicity is Mixed	-0.7772	0.2911	422	-2.67	.0079
Ethnicity is White	-0.5280	0.1142	422	-4.62	<.0001
Ethnicity is Asian (reference category)	0	.	.	.	.
Belongs to a category of special education	-0.3076	0.1187	422	-2.59	.0099
Is English proficient	-0.2533	0.1475	422	-1.72	.0868

TABLE H14. ESTIMATES OF LEVEL-1 AND LEVEL-2 VARIANCE COMPONENTS (RANDOM EFFECTS)

Random effect	Variance component	Standard error	Z value	*p* value
Teacher	0.02876	0.01632	1.76	.0390
Residual (student)	0.3108	0.02114	14.70	<.0001

Results of the Benchmark Analysis of Impacts on Math (N = 3 Teachers, N = 104 Students)

TABLE H15. ESTIMATES OF FIXED EFFECTS

Fixed effects	Coefficient	Standard error	*df*	t-ratio	*p* value
Intercept	-0.7328	0.3748	1	-1.96	.3010
Treatment (belongs to CREATE)	-0.1479	0.1840	1	-0.80	.5690
Pretest	0.8303	0.02302	94	36.06	<.0001
Gender is male	0.04577	0.1191	94	0.38	.7017
Ethnicity is Black	0.7909	0.07210	94	10.97	<.0001
Ethnicity is Hispanic	0.8550	0.06424	94	13.31	<.0001
Ethnicity is White	0.6749	0.07584	94	8.90	<.0001
Ethnicity is Mixed (reference category)	0	.	.	.	.
Belongs to a category of special education	0.04539	0.07107	94	0.64	.5246
Is English proficient	-0.2424	0.3202	94	-0.76	.4509

TABLE H16. ESTIMATES OF LEVEL-1 AND LEVEL-2 VARIANCE COMPONENTS (RANDOM EFFECTS)

Random effect	Variance component	Standard error	Z value	*p* value
Teacher	0.03676	0.04009	0.92	.1796
Residual (student)	0.1823	0.02575	7.08	<.0001

Results of the Benchmark Analysis of Impacts on ELA and Math (N = 17 Teachers, N = 548 Students)

TABLE H17. ESTIMATES OF FIXED EFFECTS

Fixed effects	Coefficient	Standard error	*df*	t-ratio	*p* value
Intercept	0.5515	0.4064	13	1.36	.1979
Subject is ELA (reference category is math)	0.3909	0.1672	28	2.34	.0267
Belongs to cohort 1	0.3158	0.1384	28	2.28	.0303
Treatment (belongs to CREATE)	-0.1351	0.1083	13	-1.25	.2343
Pretest	0.7452	0.02730	28	27.30	<.0001
Gender is male	-0.06988	0.04791	28	-1.46	.1558
Ethnicity is Black	-0.8191	0.3246	28	-2.52	.0176
Ethnicity is Hispanic	-0.6389	0.3282	28	-1.95	.0616
Ethnicity is Mixed	-0.9113	0.4259	28	-2.14	.0413
Ethnicity is White	-0.5436	0.3318	28	-1.64	.1126
Ethnicity is Asian (reference category)	0	.	.	.	.
Belongs to a category of special education	-0.2860	0.08135	28	-3.52	.0015
Is English proficient	-0.2347	0.1811	28	-1.30	.2056

TABLE H18. ESTIMATES OF LEVEL-1 AND LEVEL-2 VARIANCE COMPONENTS (RANDOM EFFECTS)

Random effect	Variance component	Standard error	Z value	*p* value
Teacher	0.03051	0.01508	2.02	.0215
Student	0.04646	0.07454	0.62	.2665
Repeated observation (within student)	0.2434	0.07465	3.26	.0006

Appendix I. Retention Coding Details for the Three-Year Early Career Trajectory

As described in Chapter 7, we relied on a variety of sources to determine the graduation and teaching status for each study participant at three time points in their early career trajectory: 1) Graduated from GSU CEHD, 2) Teaching in Year 2 (first year of teaching), and 3) Teaching in Year 3 (second year of teaching). We assigned each participant a 0 or 1 to designate their retention status at each time point. Note that the ensuing description focuses on the process for triangulating across different data sources to assign retention status *without* taking into account the timing of graduation and of teaching. We provide additional information on the process of recoding these records to factor in *on-time* graduation and teaching in chapter 7.

For Graduation from GSU CEHD, we relied on data from our participant tracker, which provided detailed contextual information about each participant's status in teaching, participant surveys, data provided to the research team by GSU or the CREATE program team, and teacher certification records from the Georgia Professional Standards Commission (GaPSC, 2014).

- If a participant taught in the first year after graduating from GSU CEHD, we assigned them a 1 for graduated from GSU CEHD.
- If a participant received an induction certification the year they were expected to graduate, we assigned them a 1 for graduated from GSU CEHD.

For Teaching in Year 2 (first year of teaching) and Teaching in Year 3 (second year of teaching), we first triangulated data received from teacher placement and quarterly surveys, data received from the CREATE program team, communication with the CREATE program team, and the TAPS and Milestones data from GaDOE using the following rules.

- If a participant listed a teaching position in a GA public school on the teacher placement or quarterly survey, or if the participant has TAPS or Milestones data for that specific year, we designate the participant as teaching, and assign them a 1 for that year.
- If a participant indicated that they were not teaching, teaching outside of Georgia, teaching in a private school—or if we obtained this information at some point during follow ups with the teacher or with the CREATE program team—we assigned the participant a 0 for that year.
- If there were any discrepancies, we referred to the data CREATE provided to us for additional contextual information.

If the participant's teaching status for a particular year was inconclusive based on the various sources of information above, then the participant's retention status was considered to be unknown up to that point, and the research team investigated each case. Researchers used the Open Georgia: Transparency in Government travel and salary database (Open Georgia, 2008)—which included employment and salary information for all teachers teaching in public schools in GA—and GA's teacher certification database, in conjunction with all other available data. The research team also discussed and determined the retention designation on a case-by-case basis, and assigned each participant one of following codes.

- **1**: The participant had a successful match on the open records. *Successful match* was defined as having one unique entry in the open records with the same first and last name as in study records, and had a salary that is commensurate with a full-time teaching position.

- **0**: There was overwhelming evidence that the participant did not teach that year; for example, the participant had a record in the open records (for the same district as the previous year), but the record indicated that they did not have a full-time salary in that year.
- **2**: There was not enough evidence to determine whether the participant was likely to be teaching in that particular year or not, even after the search of the open records (e.g., based on a non-unique name match in the open records).
- **3**: The participant did not have a match in the open records but either graduated from GSU or taught in their second year.
- **5**: The participant did not have a match in the open records, and did not graduate from GSU or did not teach in the previous year (if applicable).

The team reasoned that not having a match in the open records did not necessarily mean not teaching; an unsuccessful match could have also been due to a name change, a non-unique name match, or an alternate spelling of the name. We documented decisions in a decision tree to ensure consistency across cases and conditions.

For the analyses and results presented in Chapter 7 of this report, we collapsed the three unknown categories (2, 3, and 5) into one single category for unknown (i.e., lost to follow-up).

Table I1 shows the frequency of counts in each category of the three-year trajectory with the different categories of loss to follow-up.

TABLE I1. RAW AND RECODED COUNTS IN EACH CATEGORY

Completed GSU CEHD	Teaching in a GA public school in Y1	Teaching in a GA public school in Y2	Frequency
0	0	1	1
0	0	2	1
0	0	5	9
0	2	2	1
0	3	3	1
0	5	5	2
1	0	0	5
1	0	1	3
1	0	5	7
1	1	0	6
1	1	1	94
1	2	2	2
1	2	3	2

0 = did not graduate / not retained

1 = graduated / retained

2 = Unknown (e.g., cannot verify through direct communication or GSU records, and ambiguous information in the open records, such as more than one person having the same name)

3 = Unknown AND not found in the open records based on the name we have in our database

5 = Unknown but probably not retained (NOT found in the open records based on the name we have in our database, and either did not graduate from GSU CEHD or was not teaching the year prior, or had internal notes to suggest person is not teaching)

Table I2 shows the frequency of counts in each category of the three-year trajectory without differentiating categories of loss to follow-up.

TABLE I2. RAW AND RECODED COUNTS IN EACH CATEGORY, WITH UNKNOWN CATEGORIES COLLAPSED INTO ONE

Completed GSU CEHD	Teaching in a GA public school in Y1	Teaching in a GA public school in Y2	Frequency
0	•	•	4
0	•	•	10
0	•	•	1
1	•	•	4
1	0	•	7
1	0	•	5
1	0	•	3
1	1	0	6
1	1	1	94

0 = did not graduate / not retained

1 = graduated / retained

• = Unknown

TABLE I3. NUMBERS RETAINED AND DETAILED DESCRIPTIVE RESULTS FOR THE FULL SAMPLE

	Graduated	Teaching Year 1	Teaching Year 2
CREATE (N = 40)			
Retained (1)	39	38	34
Not Retained (0)	1	2	5
Unknown (2)	0	0	0
Unknown (3)	0	0	0
Unknown (5)	0	0	1
Comparison (*N* = 94)			
Retained (1)	80	62	64
Not Retained (0)	14	24	6
Unknown (2)	0	5	4
Unknown (3)	0	1	3
Unknown (5)	0	2	17
Difference in proportion	39/40-80/94=12.4%	38/40-62/86=22.9%	34/39-64/70 = -4.2%
***p* value (difference in proportions)** [a]	0.037	0.003	0.480
Result of logistic regression	0.070	0.011	0.446
Add Fisher's exact test	0.039	0.001	0.518

Note. In the logistic regression the log odds of retention were regressed against treatment status (CREATE or comparison) and cohort.

We relied on multiple sources to determine the retention status of a teacher (direct communication, database from GaDOE, information about teaching status in open public records). Based on this, we developed three codes for unknown status, as follows.

2 = Unknown (e.g., cannot verify through direct communication or GSU records, and ambiguous information in open records, such as more than one person having the same name)

3 = Unknown AND not found in open records based on the name we have in our database

5 = Unknown but probably not retained (NOT found in open records based on the name we have in our database, and either did not graduate from GSU or was not teaching the year prior, or had internal notes to suggest person is not teaching)

[a] Software used for parametric analysis of results for "Graduated" and "Teaching Year 1" yielded a warning that small cell sizes in Chi-squared test could produce non-reliable results

TABLE I4. NUMBERS RETAINED AND DETAILED DESCRIPTIVE RESULTS FOR THE SUBSAMPLE OF BLACK EDUCATORS

	Graduated	Teaching Year 1	Teaching Year 2
CREATE (*N* = 22)			
Retained (1)	22	21	21
Not Retained (0)	0	1	1
Unknown (2)	0	0	0
Unknown (3)	0	0	0
Unknown (5)	0	0	0
Comparison (*N* = 31)			
Retained (1)	24	21	21
Not Retained (0)	7	8	1
Unknown (2)	0	1	1
Unknown (3)	0	1	1
Unknown (5)	0	0	7
Difference in proportion	22/22-24/31=22.6%	21/22-21/29=23.0%	21/22-21/22 = -0%
***p* value (difference in proportions)** [a]	.017	.033	1.000
Result of logistic regression	0.937	0.044	0.904
Add Fisher's exact test	0.033	0.060	1.000

Note. In the logistic regression, the log odds of retention were regressed against treatment status (CREATE or comparison) and cohort.

We relied on multiple sources to determine the retention status of a teacher (direct communication, database from GaDOE, information about teaching status in open public records). Based on this, we developed three codes for unknown status, as follows.

2 = Unknown (e.g., cannot verify through direct communication or GSU records, and ambiguous information in open records, such as more than one person having the same name)

3 = Unknown AND not found in open records based on the name we have in our database

5 = Unknown but probably not retained (NOT found in open records based on the name we have in our database, and either did not graduate from GSU or was not teaching the year prior, or had internal notes to suggest person is not teaching)

[a] Software used for parametric analysis of results for all three years yielded a warning that small cell sizes in Chi-squared test could produce non-reliable results

Appendix J. Tests of Baseline Equivalence on Covariates for Retention Analysis

In each of the tables below, the standardized effect size from the model-based approach is the estimated coefficient corresponding to the indicator of treatment status, from a model where the pretest is regressed against that indicator (i.e., belongs to CREATE or not) and an indicator of cohort (i.e., belong to Cohort 2 or not), divided by the pooled standard deviation of the outcome variable. The effect size from the unadjusted approach is the raw difference in means divided by the pooled standard deviation.

TABLE J1. TESTS OF BASELINE EQUIVALENCE IN SELF-REPORTED CONFIDENCE IN GENERAL TEACHING SKILL (FULL SAMPLE)

	CREATE group			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	Standard deviation	Sample size	Unadjusted mean	Standard deviation	CREATE-comparison difference	Standardized effect size
Model-based approach	38	4.001	0.547	83	4.069	0.581	-0.067	-0.118
Unadjusted approach	38	4.012	0.547	83	4.069	0.581	-0.056	-0.098

Note. Sample includes teachers from both cohorts 1 and 2

TABLE J2. TESTS OF BASELINE EQUIVALENCE FOR MOTIVATION IN TEACHING (FULL SAMPLE)

	CREATE group			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	Standard deviation	Sample size	Unadjusted mean	Standard deviation	CREATE-comparison difference	Standardized effect size
Confidence in Teaching (model-based approach)	38	4.542	0.325	83	4.395	0.764	0.147	0.223
Unadjusted approach	38	4.558	0.325	83	4.395	0.764	0.163	0.247

Note. Sample includes teachers from both cohorts 1 and 2

TABLE J3. TESTS OF BASELINE EQUIVALENCE IN MATH ANXIETY LEVEL (FULL SAMPLE)

	CREATE group			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	Standard deviation	Sample size	Unadjusted mean	Standard deviation	CREATE-comparison difference	Standardized effect size
Confidence in Teaching (model-based approach)	38	2.611	1.074	83	2.487	1.050	0.124	0.117
Unadjusted approach	38	2.647	1.074	83	2.487	1.050	0.160	0.151

Note. Sample includes teachers from both cohorts 1 and 2

TABLE J4. TESTS OF BASELINE EQUIVALENCE IN GPA (FULL SAMPLE)

	CREATE group			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	Standard deviation	Sample size	Unadjusted mean	Standard deviation	CREATE-comparison difference	Standardized effect size
Confidence in Teaching (model-based approach)	38	3.494	0.421	83	3.417	0.454	0.076	0.172
Unadjusted approach	38	3.500	0.421	83	3.417	0.454	0.083	0.187

Note. Sample includes teachers from both cohorts 1 and 2

TABLE J5. TESTS OF BASELINE EQUIVALENCE IN SELF-REPORTED CONFIDENCE IN GENERAL TEACHING SKILL (BLACK EDUCATORS SUBSAMPLE)

	CREATE group			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	Standard deviation	Sample size	Unadjusted mean	Standard deviation	CREATE-comparison difference	Standardized effect size
Confidence in Teaching (model-based approach)	19	4.163	0.535	19	4.184	0.555	-0.022	-0.040
Unadjusted approach	19	4.167	0.535	19	4.184	0.555	-0.018	-0.032

Note. Sample includes teachers from both cohorts 1 and 2

TABLE J6. TESTS OF BASELINE EQUIVALENCE IN MOTIVATION TO ENTER THE TEACHING PROFESSION (BLACK EDUCATORS SUBSAMPLE)

	CREATE group			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	Standard deviation	Sample size	Unadjusted mean	Standard deviation	CREATE-comparison difference	Standardized effect size
Confidence in Teaching (model-based approach)	19	4.611	0.343	19	4.674	0.284	-0.063	-0.200
Unadjusted approach	19	4.613	0.343	19	4.674	0.284	-0.061	-0.193

Note. Sample includes teachers from both cohorts 1 and 2

TABLE J7. TESTS OF BASELINE EQUIVALENCE IN MATH ANXIETY LEVEL (BLACK EDUCATORS SUBSAMPLE)

	CREATE group			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	Standard deviation	Sample size	Unadjusted mean	Standard deviation	CREATE-comparison difference	Standardized effect size
Confidence in Teaching (model-based approach)	19	2.477	1.090	19	2.374	1.119	0.103	0.094
Unadjusted approach	19	2.484	1.090	19	2.374	1.119	0.111	0.100

Note. Sample includes teachers from both cohorts 1 and 2

TABLE J8. TESTS OF BASELINE EQUIVALENCE IN GPA (BLACK EDUCATORS SUBSAMPLE)

	CREATE group			Comparison group			Baseline difference	
Baseline measure	Sample size	Model-adjusted mean	Standard deviation	Sample size	Unadjusted mean	Standard deviation	CREATE-comparison difference	Standardized effect size
Confidence in Teaching (model-based approach)	19	3.350	0.401	19	3.273	0.378	0.078	0.199
Unadjusted approach	19	3.362	0.401	19	3.273	0.378	0.089	0.228

Note. Sample includes teachers from both cohorts 1 and 2

References

edTPA. (2020). *edTPA for Georgia*. Pearson Education, Inc. https://www.edtpa.com/PageView.aspx?f=GEN_Georgia.html

Georgia Professional Standards Commission (GaPSC) (2014). *Check Certification Status*. https://www.gapsc.com/Certification/Lookup.aspx

Georgia Professional Standards Commission (GaPSC). (n.d.). *Induction*. https://www.gapsc.com/Certification/TieredCertification/induction.aspx

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley Online Library. https://onlinelibrary.wiley.com/doi/book/10.1002/0471722146

Open Georgia. (2008). *State of Georgia Financial Reports*. http://open.georgia.gov/