Appendix

Evaluation of Georgia State University's CREATE Teacher Residency Program – Cohorts 3 through 5

A FINAL REPORT OF A QUASI-EXPERIMENT IN GEORGIA

March 31, 2023

Andrew P. Jaciw

Sze-Shun Lau

Jenna Zacamy

Li Lin

Empirical Education Inc.

Table of Contents

Reference this report: Jaciw, A. P., Lau, S., Zacamy, J., & Lin, L. (2023). Evaluation of Georgia State University's CREATE Teacher Residency Program – Cohorts 3 through 5: A Final Report of a Quasi-Experiment in Georgia. (Empirical Education Rep. No. Empirical_GSU-7035-FR-2023-O.2). Empirical Education Inc. https://www.empiricaleducation.com/create/

Appendix A. Description of CREATE Programming

CREATE's three-year teacher residency program is designed to strengthen novice teachers' professional knowledge and develop compassionate, skilled, and anti-racist educators. During Year 1 of the residency, while in their final year at Georgia State University's College of Education and Human Development (GSU CEHD), residents complete their preservice teaching practicum by serving as student teachers in K–8 classrooms within schools in Atlanta Public Schools under the supervision of a Cooperating Teacher (CT). As residents move through the three-year residency model, their role within the classroom changes. In Year 2 of the program, once residents have completed their program at GSU CEHD, CREATE residents complete their first year as a teacher-of-record while paired with another CREATE teacher, who serves as their co-teacher, in a single classroom. While working as a co-teacher in Year 2, each resident receives a full-time teaching salary. In Year 3, residents become the *sole* "teacher of record" in their own classroom. In addition to these "progressive classroom roles," throughout the three-year residency, CREATE residents receive support from their resident cohort and the CREATE program team each year in the form of regular Together Time meetings, access to mentors, and intensive summer training. Together Time meetings, which occur approximately monthly throughout the school year, serve as opportunities for residents to meet with those in their resident cohort. They share and collaborate, discuss their classroom experiences and dilemmas of practice, build classroom management skills, develop tactics for stress reduction, and engage about racism and oppression as it relates to their work. Residents also have access to multiple forms of mentorship, including CREATE-trained school-based mentors, instructional mentors, and the CREATE program team for support. Furthermore, residents participate in a Summer Resident Academy (SRA) for a five-week intensive training during the summer after their student teaching year (that is, between Year 1 and Year 2 of the residency), and for a two-week intensive training during the summer after their first year as full-time teachers (that is, between Year 2 and Year 3 of the residency). Through SRA, CREATE guides residents in developing social emotional competencies, pedagogical skills, content knowledge, and the confidence they will need for success in their beginning years as full-time teachers. The CREATE teacher residency program has evolved since the beginning of the research study, in response to observed need based on social context. CREATE has shifted its work towards an expansion of equity-centered professional learning (PL) opportunities, and the integration of mindfulness and critically-conscious practices. These programmatic shifts aim to effectively develop anti-racist and compassion-centered mindsets and practices in new teachers.

CREATE invites all students enrolled in GSU CEHD's teacher credentialing programs, in both the Early Childhood and Elementary Education (ECEE) and Middle and Secondary Education (MSE) departments, to apply to participate in the CREATE teacher residency program. Staff members at CREATE conduct presentations at GSU CEHD, providing students with an overview of the three-year residency program and inviting them to submit an application to become a resident. These presentations usually take place in the spring and summer, setting students up to begin their residency at the beginning of their final year of study at GSU CEHD. CREATE admits students into the program based on a variety of information in the applications, including the students' degree of interest in teaching in historically-underserved communities in Atlanta and how their goals and experiences relate to CREATE's mission. Through these recruitment efforts, the CREATE program team is dedicated to contributing to the diversification of the teacher workforce.

In addition to the residency program for new teachers, CREATE also offers a series of PL opportunities for experienced educators. The majority of participants in these PL opportunities work in CREATE schools, but those who work in schools that are not partners of CREATE are eligible to register to attend some of the PL opportunities. CREATE's PL opportunities—focused on developing teacher leaders within CREATE schools—include Cooperating Teacher trainings

and School Based Mentor trainings, which equip experienced educators to work one-on-one with CREATE residents as mentors in professionalism, instructional skills, classroom practice, collaboration, reflection skills, and other essential skills for teaching.

CREATE also offers equity-centered and compassion-centered PL opportunities, which allow experienced educators to work and learn alongside other educators to collaborate and develop their own teaching practices. CREATE's primary equity-centered PL is the Equity-Centered Critical Friendship (ECCF) Institute, at which educators focus on developing understandings of individual and collective identity and liberatory practices as they contribute to educational equity. CREATE offered Critical Friendship (CF) Institutes through fall 2018, where facilitators guided educator participants through examining problems of practice and assumptions about students and families at their school sites using structured protocols. iGroup was an additional equity-centered PL offered by CREATE through Fall 2019. iGroup provided an opportunity for participants to explore their socio-political identities, develop an understanding of institutionalized racism, and strengthen their ability to identify and interrupt inequities in education (CREATE, personal communication, 2019).

CREATE's compassion-based PL includes Presence, Power, Impact (PPI), which supports participants in strengthening self-awareness, self-compassion, and their contemplative practices that help cultivate personal resilience during challenging teaching moments (CREATE personal communication, 2021). PPI draws upon practices taught at Cognitively-Based Compassion Training© (CBCT), at which CREATE engaged research-based practices born out of Emory University's Center for Contemplative Science and Compassion-Based Ethics to foster educator resilience through increased mindfulness (CREATE, personal communication, 2019). CREATE offered CBCT PL through summer 2019, when PPI was introduced. CREATE's partner schools are given a limited number of complimentary seats for these PL opportunities, increasing accessibility for educators in the CREATE community. CREATE integrates practices from both its equity-centered PLs and compassion-based PLs as central components of residents' Together Time meetings.

Appendix B. Participant Recruitment

Recruitment began in summer 2017 for Cohort 3, in spring 2018 for Cohort 4, and in spring 2019 for Cohort 5. Each year, we presented the research study to students who were both in their final year of GSU CEHD's teacher credentialing program and also identified as eligible for participation in the research. We recruited treatment cases from the pool of students who had applied to participate in CREATE and who chose to join the program. We recruited comparison cases for the study from the pool of students who were eligible for CREATE, but who chose to not apply for the program (for a variety of reasons described in the General Design section below).

In order for student teachers to be eligible for inclusion in the research study, they needed to:

- be enrolled in GSU CEHD,
- be completing their student teaching during their first year of participation in the research,
- plan to teach in a public school in Georgia,
- plan to teach in an elementary or middle school, and
- expect to complete the teacher certification requirements and graduate from GSU CEHD in the spring of the first year of participation in the research.

To recruit participants for all cohorts, researchers held recruitment events at which eligible GSU CEHD students learned about the research study and participation in it. The recruitment presentation occurred in-person for Cohort 3 and virtually for Cohorts 4 and 5. We conducted presentations during courses or orientation events that GSU CEHD students attended, during which they provided potential study participants with information about the research study, data collection activities, and related stipends. We then provided attendees with an opportunity to ask questions. The end of each recruitment presentation gave attendees instructions on how to complete a consent form if they were interested in participating in the study. Interested students received consent forms as hard copies or via an online link, depending on the arrangement that was made with the point-of-contact at GSU. The GSU department contact mailed the completed hard copy consent forms to Empirical Education. Researchers also emailed a link to the CREATE residents who had not yet consented. The link opened a recorded version of the recruitment presentation and an invitation to complete an online consent form.

Appendix C. Details about Study Participation

After agreeing to participate in the 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 based on availability of data. We may still have collected outcomes from the Georgia Department of Education (GaDOE) or publicly available records after teachers left the study or stopped responding to surveys.

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. Seven Cohort 3 participants, all of whom were part of the comparison group, became inactive during Year 1 of the study. Of these seven participants, four had dropped from the program at GSU CEHD, one participant was not responsive to data collection attempts and follow-up communication, and another two were ineligible for the study (including one participant who had taken a semester off from their program due to athletics, and one who did not continue their practicum until the following fall semester). Table C1 includes a summary of attrition for Cohort 3.

TABLE C1. 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
(Left 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)	(2)
Remaining sample at end of Year 1	14	41

Year 2

By the end of Year 2, an additional 23 study participants had left the study. Eighteen were part of the comparison group, and five were part of the CREATE group. We provide details below and in Table C2.

CREATE Group

Of the five CREATE participants that left the study in Year 2, we made four inactive because they left the CREATE program Researchers made one additional participant inactive because they did not consent to extend their participation to Year 2. Thus, the researchers chose to not include them in the study for year 2. We should note that Cohort 3 was initially recruited into CREATE with funds from the i3 grant. CREATE was unsure if they would receive additional funding and therefore recruited Cohort 3 to participate for just one year (with the hope that they could extend

participation). Due to this, CREATE (and Empirical Education) needed to extend participation agreements in Year 2 when the 2017 SEED grant was awarded. Some participants chose not to continue participation beyond the first year.

Comparison Group

Of the 18 comparison study participants who left the study in Year 2, two participants indicated that they did not want to continue participation in the study. Eight participants became inactive due to not meeting eligibility criteria for the following reasons.

- one participant started working in a non-teaching education-related position
- one participant started working outside of Georgia
- two participants did not graduate in year one
- four participants did not start working in a classroom setting

An additional eight participants were made inactive because we were unable to obtain permission to conduct research from their schools and/or districts.

TABLE C2. ATTRITION FROM STUDY FOR COHORT 3, YEAR 2

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 2	14	41
(Left CREATE program)	(2)	(0)
(Left CREATE program, Moved outside of Georgia)	(2)	(0)
(No longer wanted to participate in research/ Non-responsive to follow-up)	(0)	(2)
(Ineligible to continue based on eligibility criteria for the study)	(1)	(8)
(Unable to obtain district/school permission to survey participant)	(0)	(8)
Remaining sample at end of Year 2	9	23

Year 3

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

CREATE Group

One CREATE resident became inactive in Year 3. The participant was non-responsive to our outreach attempts prompting completion of data collection activities.

Comparison Group

Three comparison group study participants became inactive in Year 3. Two participants were nonresponsive to data-collection attempts. One participant was no longer working in a school setting and thus in ineligible to participate in the study. Two 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 C3. ATTRITION FROM STUDY FOR COHORT 3, YEAR 3

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 3	9	23
(No longer wanted to participate in research/Non-responsive to follow-up)	(1)	(2)
(Ineligible to continue based on eligibility criteria for the study)	(0)	(1)
(Unable to obtain district/school permission to survey participant)	(0)	(0)
Obtained district/school permission to survey participant	0	2
Remaining sample at end of Year 3	8	22

Summary of Cohort 3 Attrition

A total of 32 study participants (26 comparison and 6 CREATE) became (and remained) inactive over the course of the study. See Table C4 for a summary table of attrition from the study sample in Cohort 3 during the three years of the study.

TABLE C4. SUMMARY OF COHORT 3 ATTRITION

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
(No longer wanted to participate in research/Non-responsive to follow-up)	(1)	(9)
(Ineligible to continue based on eligibility criteria for the study)	(5)	(15)
(Unable to obtain district/school permission to survey participant)	(0)	(2)

TABLE C4. SUMMARY OF COHORT 3 ATTRITION

Event/Reason for attrition	CREATE group	Comparison group
Remaining sample at end of Year 3	8	22

COHORT 4

Year 1: Summary of Cohort 4 Attrition

During recruitment, 16 CREATE and 51 comparison group study participants agreed to be part of the study. Twenty-one Cohort 4 participants—one CREATE participant and 20 comparison participants—became inactive during Year 1 of the study. The one CREATE participant converted to inactive because they were no longer at GSU CEHD. Of these 20 comparison participants, two participants left GSU CEHD (one for medical reasons), 6 were not responsive to data collection attempts, and 12 participants' placement settings did not grant us research permissions. Table C5 includes a summary of attrition for Cohort 4.

TABLE C5. ATTRITION FROM STUDY FOR COHORT 4, 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	16	51
(Left GSU CEHD)	(1)	(2)
(No longer wanted to participate in research/ Non-responsive to follow-up)	(0)	(6)
(Unable to obtain district/school permission to survey participant)	(0)	(12)
Remaining sample at end of Year 1	15	31

Year 2

By the end of Year 2, an additional 11 study participants had left the study and two study participants returned to the study. Of the 11 participants made inactive, nine were part of the comparison group, and two were part of the CREATE group. We provide details below and in Table C6.

CREATE Group

Of the two CREATE participants that left the study in Year 2, we made one inactive because they left the CREATE program. The other participant became inactive because they were not working in a classroom setting and thus were no longer eligible to participate in the study.

Comparison Group

Of the nine comparison study participants who left the study in Year 2, four participants did not respond to outreach. Five study participants left the study because they were no longer working in school settings and were no longer eligible to participate. Two study participants who became inactive in Year 1, due to research permissions, re-entered the study in Year 2 when we were able to obtain permission from their school/district.

TABLE C6. ATTRITION FROM STUDY FOR COHORT 4, YEAR 2

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 2	15	31
(Left CREATE program)	(1)	(0)
(No longer wanted to participate in research/ Non-responsive to follow-up)	(0)	(4)
(Ineligible to continue based on eligibility criteria for the study)	(1)	(5)
Obtained district/school permission to survey participant	0	2
Remaining sample at end of Year 2	13	24

Year 3

Eight study participants, all in the comparison group, became inactive in Year 3. Two study participants returned to the study. We provide details below and in Table C7.

CREATE Group

No CREATE participants became inactive in Year 3.

Comparison Group

Eight comparison group study participants became inactive in Year 3. One participant was nonresponsive to data-collection attempts. Four participants were teaching at a high school and one was no longer teaching, so five participants were made inactive due to ineligibility. Two participants were inactive because we were unable to obtain permission from their schools and/or districts to conduct research. Two 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 C7. ATTRITION FROM STUDY FOR COHORT 4, YEAR 3

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 3	13	24

TABLE C7. ATTRITION FROM STUDY FOR COHORT 4, YEAR 3

Event/Reason for attrition	CREATE group	Comparison group
(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)	(5)
(Unable to obtain district/school permission to survey participant)	(0)	(2)
Obtained district/school permission to survey participant	0	2
Remaining sample at end of Year 3	13	18

Summary of Cohort 4 Attrition

A total of 36 study participants (33 comparison and 3 CREATE) became (and remained) inactive over the course of the study. See Table C8 for a summary table of attrition from the study sample in Cohort 4 during the three years of the study.

TABLE C8. SUMMARY OF COHORT 4 ATTRITION

Event/Reason for attrition	CREATE group	Comparison group
Consented to participate in study and were eligible to participate at the time of consent	16	51
(No longer wanted to participate in research/Non-responsive to follow-up)	(0)	(11)
(Ineligible to continue based on eligibility criteria for the study)	(3)	(10)
(Unable to obtain district/school permission to survey participant)	(0)	(12)
Remaining sample at end of Year 3	13	18

COHORT 5

Year 1: Summary of Cohort 5 Attrition

During recruitment, 17 CREATE and 25 comparison group study participants agreed to be part of the study. Six Cohort 5 participants—one CREATE participant and five comparison participants—became inactive during Year 1 of the study. The one CREATE participant who became inactive dropped out of GSU CEHD. Of the five comparison participants that became inactive, one dropped from the program at GSU CEHD, and four participants' placement settings did not grant us research permissions. Table C9 includes a summary of attrition for Cohort 3.

TABLE C9. ATTRITION FROM STUDY FOR COHORT 5, 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	17	25
(Left GSU CEHD)	(1)	(1)
(Unable to obtain district/school permission to survey participant)	(0)	(4)
Remaining sample at end of Year 1	16	20

Year 2

By the end of Year 2, an additional 17 study participants had left the study. Eleven were part of the comparison group, and six were part of the CREATE group. We provide details below and in Table C10.

CREATE Group

Of the six CREATE participants that left the study in Year 2, all were inactive because they left the CREATE program. One of these participants also dropped out of GSU CEHD.

Comparison Group

Of the 11 comparison study participants who left the study in Year 2, five did not respond to outreach. Six participants were not part of the expecting sample (two were no longer working in a school setting, one was not teaching at a public school, one was not graduating on time, and two were not working in a K–8 setting).

TABLE C10. ATTRITION FROM STUDY FOR COHORT 5, YEAR 2

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 2	16	20
(Left CREATE program)	(5)	(0)

TABLE C10. ATTRITION FROM STUDY FOR COHORT 5, YEAR 2

Event/Reason for attrition	CREATE group	Comparison group
(Left GSU CEHD)	(1)	(0)
(No longer wanted to participate in research/ Non-responsive to follow-up)	(0)	(5)
(Ineligible to continue based on eligibility criteria for the study)	(0)	(6)
Remaining sample at end of Year 2	10	9

Year 3

One comparison participant became inactive in Year 3, and one returned to the study. We provide details below and in Table C11.

CREATE Group

No CREATE participants became inactive in Year 3.

Comparison Group

The one comparison participant that became inactive in Y3 was no longer working in a classroom setting and was ineligible to participate in the study. One study participant that was inactive in Year 2 returned to the study in Year 3 because we were able to receive the necessary research permissions.

TABLE C11. ATTRITION FROM STUDY FOR COHORT 5, YEAR 3

Event/Reason for attrition	CREATE group	Comparison group
Sample at beginning of Year 3	10	9
(Ineligible to continue based on eligibility criteria for the study)	(0)	(1)
Obtained district/school permission to survey participant	0	1
Remaining sample at end of Year 3	10	9

Summary of Cohort 5 Attrition

A total of 23 study participants (16 comparison and 7 CREATE) became (and remained) inactive over the course of the study. See Table C12 for a summary of attrition from the study sample in Cohort 5, during the three years of the study.

TABLE C12. SUMMARY OF COHORT 5 ATTRITION

Event/Reason for attrition	CREATE group	Comparison group
Consented to participate in study and were eligible to participate at the time of consent	17	25
(Left CREATE program)	(5)	(0)
(Left GSU CEHD)	(2)	(1)
(No longer wanted to participate in research/Non-responsive to follow-up)	(0)	(5)
(Ineligible to continue based on eligibility criteria for the study)	(0)	(7)
(Unable to obtain district/school permission to survey participant)	(0)	(3)
Remaining sample at end of Year 3	10	9

Appendix D. Data Collection Sources

This study utilized multiple sources of data to measure desired outcomes.

Participant Surveys

Baseline Survey

We invited study participants in both the CREATE teacher residency program and the traditional credentialing program to complete the initial baseline survey, after they agreed to participate in the research study. This baseline survey asked study participants questions about their teaching-related background, motivation, perspective, and interests. Responses to this survey allowed us to confirm participants' eligibility for the research study, as well as inform the selection of the comparison group. We also used data from this survey in analysis: as variables for matching, as covariates in ANCOVA analysis, and as moderators in assessments of differential impacts. We administered this survey to study participants one time only, when they joined the research study.

Quarterly Surveys

Participants in both conditions completed quarterly online surveys for the duration of the three years of the study for their respective cohort. These surveys took no more than 20 minutes each to complete, on average. Surveys included questions related to support during their student teaching year and first two years of teaching, classroom experiences, and plans for continued teaching. They also included established scales, such as the Maslach Teacher Burnout Inventory, the Self-Compassion scale, and questions from the PRIDE Teaching Environment survey. See Appendix F. Survey Response Rates for study participant response rates for each individual survey.

PRIDE Teaching Environment Survey

Included in the final quarterly survey of each school year were items from the PRIDE Teaching Environment Survey. The survey assessed factors shown to be related to the likelihood that a teacher will remain in the education profession. These factors include levels of teacher satisfaction, motivation, self-efficacy, support, career goals and intentions, school climate, and the teaching experience (Elfers et al., 2006). In the first year of participation, study participants were not yet full-time teachers and were placed at their practicum schools for varying amounts of time. Therefore, some items were adjusted to more accurately reflect the participant context.

Maslach Burnout Inventory

Teachers, among others who frequently work with people, are at risk for burnout. The Maslach Burnout Inventory for Educators assesses the three components of burnout using a 22-item scale: 1) emotional exhaustion (depletion of emotional resources), 2) depersonalization (negative feelings about one's students), and 3) reduced personal accomplishment (tendency to evaluate oneself negatively with regard to work). Respondents read a series of statements and respond with how frequently they experience the stated feelings or attitudes on a 6-point scale ranging from *Never* to *Everyday*. Cronbach's alpha is 0.90 for emotional exhaustion, 0.79 for depersonalization, and 0.71 for personal accomplishment (Maslach et al., 2018).

Self-Compassion Scale

The Self-Compassion Scale (SCS) is a 26-item scale that assesses an individual's levels of self-compassion according to three main components of self-compassion: 1) self-kindness (versus self-judgment), 2) a sense of common humanity

(versus isolation), and 3) mindfulness (versus over-identification). We asked respondents a series of statements about their thoughts and feelings with regard to the above three components on a 5-point scale of *Almost Never* to *Almost Always*. Total SCS scores have good internal reliability (Cronbach's alpha = 0.92). The six scales have Cronbach's alphas ranging from 0.75–0.81. A large body of research also demonstrates strong construct validity; thus, SCS scores correlate with wellbeing. The SCS administration takes place in the spring of each study school year, for all study participants (Neff, 2003). [1]

Georgia Department of Education Data

Teacher Level Data

We collected teacher-level data from GaDOE, which included Teacher Assessment on Performance Standards (TAPS) ratings, gender, race, ethnicity, and termination information, if applicable. TAPS is a rubrics-based evaluation method used by GaDOE to measure Georgia public school teachers' performance on a set of designated performance standards. TAPS allows teacher effectiveness to be measured consistently throughout the state. There are ten performance standards that TAPS uses to rate teachers on a scale of 0 to 3: Level 0 is Emerging, Level I is Developing, Level II is Proficient, and Level III is Advanced. Through the programming and support it offers, CREATE aims primarily to improve teacher efficacy in two of the ten performance standards measured by TAPS, both of which we planned to analyze for this project.

1. Instructional strategies (the teacher promotes student learning by using research-based instructional strategies relevant to the content area to engage students in active learning and to facilitate the students' acquisition of key knowledge and skills)
2. Positive learning environment (the teacher provides a well-managed, safe, and orderly environment that is conducive to learning and encourages respect for all) (GaDOE, 2020)

The ordinal alpha, a measure similar to Cronbach's alpha, for the ten items in TAPS is 0.95, which indicates high internal consistency. Due to COVID-19 interruptions, TAPS scores were not available for spring 2020 and spring 2021.

Student Level Data

Student level data collected from GaDOE include gender, age, grade level, race, ethnicity, special education status, limited English proficiency status, and Georgia Milestones ELA and mathematics scores. The Georgia Milestones assesses ELA and mathematics student achievement for students in grades 3–8, according to state-adopted content standards. The Georgia Milestones is a valid and reliable measure for student achievement in Georgia. Cronbach's alpha reliability coefficient for the Georgia Milestones ranges from 0.89 to 0.94 across all subjects, which is an adequate level of reliability for the stated goals of the assessment (GaDOE, 2019).

Georgia State University College of Education and Human Development Data

We collected teacher level data from GSU CEHD, which included study participants' practicum placements and Intern Keys ratings.

The teacher Intern Keys assessment (Elder et al., n.d.) is a rubrics-based evaluation that aligns directly with TAPS. University supervisors and cooperating teachers use this rubric to measure student teachers' performance on 10 state

[1] We added the Maslach Burnout Inventory and the Self-Compassion Scale to the quarterly surveys starting in 2018–19.

performance standards during their practicum on a rating scale of 1 to 4: Level I is Ineffective, Level II is Needs Development, Level III is Proficient, and Level IV is Exemplary. The Cronbach's alpha reliability coefficient for the teacher Intern Keys assessment is 0.90, indicating a high degree of reliability. We use the Intern Keys ratings as a baseline measure for TAPS.

Publicly Available Data on Teacher Certification and Teaching Status

Certification Data

The Educator Certification Division of the Georgia Professional Standards Commission provides a publicly available database to confirm certification status for Georgia educators (Georgia Professional Standards Commission, 2023). The database includes certification type, level, field, and issue and validity dates. This database was used to triangulate self-reported data (if needed) or fill in missing values for teacher preparation program completion.

Teaching Status

The State of Georgia provides a publicly available database to provide information on state expenditures (Open Georgia, 2008). The database includes annual salaries and travel expenses for employees of Local Boards of Education, including teachers. The research team used this information, along with a variety of other data sources, to determine teachers' teaching status.

CREATE Program Data

Researchers collected various program data from the CREATE program team in order to corroborate resident self-report survey data on FOI measures (and to report on FOI measures not addressed by resident survey data). Program data for residents include classroom placement rosters, Together Time attendance, logs for mentor meetings and observation cycles, and summer internship/academy attendance. We also collect program data for experienced educators at CREATE schools participating in CREATE activities such as attendance rosters for ECCF, PPI, and mentor trainings.

Appendix E. Schedule of Major Milestones

Table E1 lists the study's major milestones for Cohorts 3, 4, and 5.

TABLE E1. RESEARCH MILESTONES FOR COHORTS 3, 4, AND 5

Date	Milestone
May – September 2017	Recruited CREATE and comparison participants for Cohort 3 of the research study Collected signed study consent forms and administered the baseline survey to Cohort 3 participants
December 2017	Submitted application to external IRB and received exemption from full review for the SEED study
November 2017	Deployed first 2017–18 quarterly survey to Cohort 3 (subsequent surveys deployed in January, March, and April 2018)
March – July 2018	Recruited CREATE and comparison study participants for Cohort 4 Collected signed consent forms from and administered the baseline survey to Cohort 4 participants
November 2018	Deployed first 2018–19 quarterly survey to Cohorts 3 & 4 (subsequent surveys deployed in January, March, and April 2019) Delivered first interim report to GSU Annual data collection from GaDOE
March – July 2019	Recruited CREATE and comparison study participants for Cohort 5 Collected signed consent forms from and administered the baseline survey to Cohort 5 participants
November 2019	Delivered second interim report to GSU
December 2019	Deployed first 2019–20 quarterly survey to Cohorts 3, 4, & 5 (subsequent surveys deployed in January, March, and April 2020) Annual data collection from GaDOE
December 2020	Deployed first 2020–21 quarterly survey to Cohorts 4 & 5 (subsequent surveys deployed in January, March, and May 2021) Delivered third interim report to GSU Annual data collection from GaDOE
November 2021	Deployed first 2021–22 quarterly survey to Cohort 5 (subsequent surveys deployed in January, March, and May 2022) Delivered fourth interim report to GSU Annual data collection from GaDOE
Summer 2022 – Winter 2023	Warehousing final data, triangulated retention data, conducted final data analysis, and drafted report

Note. IRB = Institutional Review Board

Appendix F. Survey Response Rates

Tables F1 through F9 provide response rates to the quarterly surveys of participants who were active at the end of Year 3 of participation.

TABLE F1. 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 – September 2017	55/55 (100%)	14/14 (100%)	41/41 (100%)
Quarterly survey 1	November 2017	54/55 (98%)	14/14 100%	40/41 (98%)
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%)

TABLE F2. COHORT 3, YEAR 2 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	29/32 (91%)	8/9 (89%)	21/23 (91%)
Quarterly survey 2	January 2019	29/32 (91%)	8/9 (89%)	21/23 (91%)
Quarterly survey 3	March 2019	29/32 (91%)	8/9 (89%)	21/23 (91%)

TABLE F2. COHORT 3, YEAR 2 SURVEY RESPONSE RATES (2018–19)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 4	April 2019	29/32 (91%)	8/9 (89%)	21/23 (91%)

TABLE F3. COHORT 3, YEAR 3 SURVEY RESPONSE RATES (2019–20)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	December 2019	30/30 (100%)	8/8 (100%)	22/22 (100%)
Quarterly survey 2	January 2020	30/30 (100%)	8/8 (100%)	22/22 (100%)
Quarterly survey 3	March 2020	30/30 (100%)	8/8 (100%)	22/22 (100%)
Quarterly survey 4	April 2020	29/30 (97%)	8/8 (100%)	21/22 (95%)

TABLE F4. COHORT 4, YEAR 1 SURVEY RESPONSE RATES (2018–19)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Baseline survey	April – August 2018	61/61 (100%)	16/16 (100%)	45/45 (100%)
Quarterly survey 1	November 2018	45/46 (98%)	15/15 (100%)	30/31 (97%)
Quarterly survey 2	January 2019	46/46 (100%)	15/15 (100%)	31/31 (100%)
Quarterly survey 3	March 2019	46/46 (100%)	15/15 (100%)	31/31 (100%)

TABLE F4. COHORT 4, YEAR 1 SURVEY RESPONSE RATES (2018–19)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 4	April 2019	45/46 (98%)	15/15 (100%)	30/31 (97%)

TABLE F5. COHORT 4, YEAR 2 SURVEY RESPONSE RATES (2019–20)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	November 2019	36/37 (97%)	12/13 (92%)	24/24 (100%)
Quarterly survey 2	January 2020	36/37 (97%)	12/13 (92%)	24/24 (100%)
Quarterly survey 3	March 2020	36/37 (97%)	12/13 (92%)	24/24 (100%)
Quarterly survey 4	April 2020	36/37 (97%)	12/13 (92%)	24/24 (100%)

TABLE F6. COHORT 4, YEAR 3 SURVEY RESPONSE RATES (2020–21)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	December 2020	26/27 (96%)	13/13 (100%)	13/14 (93%)
Quarterly survey 2	January 2021	25/27 (93%)	12/13 (92%)	13/14 (93%)
Quarterly survey 3	March 2021	30/32 (94%)	12/13 (92%)	18/19 (95%)

TABLE F6. COHORT 4, YEAR 3 SURVEY RESPONSE RATES (2020–21)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 4	April 2021	30/32 (94%)	12/13 (92%)	18/19 (95%)

Note. Due to varied district research permission requirements, five Cohort 4 comparison participants were invited only to the third and fourth quarterly survey during the 2020–21 school year.

TABLE F7. COHORT 5, YEAR 1 SURVEY RESPONSE RATES (2019–20)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Baseline survey	March – July 2019	42/42 (100%)	17/17 (100%)	25/25 (100%)
Quarterly survey 1	December 2019	35/36 (97%)	16/16 (100%)	19/20 (95%)
Quarterly survey 2	January 2020	35/36 (97%)	16/16 (100%)	19/20 (95%)
Quarterly survey 3	March 2020	35/36 (97%)	16/16 (100%)	19/20 (95%)
Quarterly survey 4	April 2018	35/36 (97%)	16/16 (100%)	19/20 (95%)

TABLE F8. COHORT 5, YEAR 2 SURVEY RESPONSE RATES (2020–21)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	December 2020	16/16 (100%)	10/10 (100%)	6/6 (100%)
Quarterly survey 2	January 2021	16/16 (100%)	10/10 (100%)	6/6 (100%)
Quarterly survey 3	March 2021	19/19 (100%)	10/10 (100%)	9/9 (100%)

TABLE F8. COHORT 5, YEAR 2 SURVEY RESPONSE RATES (2020–21)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 4	May 2021	18/19 (95%)	10/10 (100%)	8/9 (95%)

Note. Due to varied district research permission requirements, three Cohort 5 comparison participants were invited only to the third and fourth quarterly survey during the 2020–21 school year.

TABLE F9. COHORT 5, YEAR 3 SURVEY RESPONSE RATES (2021–22)

Survey	Date	Total response rate (# who responded / # active)	CREATE group response rate	Comparison group response rate
Quarterly survey 1	December 2021	18/19 (95%)	9/10 (90%)	9/9 (100%)
Quarterly survey 2	January 2022	18/19 (95%)	9/10 (90%)	9/9 (100%)
Quarterly survey 3	March 2022	17/19 (89%)	9/10 (90%)	8/9 (89%)
Quarterly survey 4	May 2022	17/19 (89%)	9/10 (90%)	8/9 (89%)

Appendix G. Fidelity of Implementation Matrix

TABLE G1. KEY COMPONENT 1: PROGRESSIVE CORE CLASSROOM ROLES

Indicators	Definition	Unit of implementation	Data sources	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 placed in the same building as another Y1 resident for fall semester	Resident	CREATE program rosters, Resident surveys	Rosters emailed to evaluators from CREATE; Surveys administered quarterly by evaluators	0 = Resident is not placed in the same building as another Y1 resident; 1 = Resident is placed in the same building as another Y1 resident for the fall semester	Adequate implementation at resident level = score of 1		14 residents (Cohort 3) 18 residents (Cohort 4) 18 residents (Cohort 5)	Cohort 3: 2017–2018 Cohort 4: 2018–2019 Cohort 5: 2019–2020
Indicator 2: Y1	(Y1) Resident is placed in classroom of experienced 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 an experienced educator trained in mentoring; 1 = Resident is placed with an experienced educator trained in mentoring	Adequate implementation at resident level = score of 1		14 residents (Cohort 3) 18 residents (Cohort 4) 18 residents (Cohort 5)	Cohort 3: 2017–2018 Cohort 4: 2018–2019 Cohort 5: 2019–2020
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		14 residents (Cohort 3) 18 residents (Cohort 4) 18 residents (Cohort 5)	Cohort 3: 2017–2018 Cohort 4: 2018–2019 Cohort 5: 2019–2020

TABLE G1. KEY COMPONENT 1: PROGRESSIVE CORE CLASSROOM ROLES

Indicators	Definition	Unit of implementation	Data sources	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: 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		10 residents (Cohort 3) 15 residents (Cohort 4) 10 residents (Cohort 5)	Cohort 3: 2018–19 Cohort 4: 2019–20 Cohort: 2020–21
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		10 residents (Cohort 3) 15 residents (Cohort 4) 10 residents (Cohort 5)	Cohort 3: 2018–19 Cohort 4: 2019–20 Cohort: 2020–21
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		10 residents (Cohort 3) 15 residents (Cohort 4) 10 residents (Cohort 5)	Cohort 3: 2018–19 Cohort 4: 2019–20 Cohort: 2020–21

TABLE G1. KEY COMPONENT 1: PROGRESSIVE CORE CLASSROOM ROLES

Indicators	Definition	Unit of implementation	Data sources	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: 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		10 residents (Cohort 3) 15 residents (Cohort 4) 10 residents (Cohort 5)	Cohort 3: 2019–20 Cohort 4: 2020–21 Cohort: 2021–22
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		10 residents (Cohort 3) 15 residents (Cohort 4) 10 residents (Cohort 5)	Cohort 3: 2019–20 Cohort 4: 2020–21 Cohort: 2021–22
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		10 residents (Cohort 3) 15 residents (Cohort 4) 10 residents (Cohort 5)	Cohort 3: 2019–20 Cohort 4: 2020–21 Cohort: 2021–22

TABLE G1. KEY COMPONENT 1: PROGRESSIVE CORE CLASSROOM ROLES

Indicators	Definition	Unit of implementation	Data sources	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
All indicators					Resident-level component score: Number of indicators implemented with fidelity per year. Possible range (per year) = 0 - 3	Threshold = 2 (Resident meets fidelity on at least 2 indicators)	Component score ranges from 0-1: Fidelity not met: Y1: Less than 90% of residents meet fidelity on 2+ indicators Y2: Less than 80% of residents meet fidelity on 2+ indicators Y3: Less than 85% of residents meet fidelity on 2+ indicators Fidelity met = Y1: 90% or more of residents meet fidelity on 2+ indicators; Y2: 80% or more of residents meet fidelity on 2+ indicators; Y3: 85% or more of residents meet fidelity on 2+ indicators	Y1: 14 Cohort 3 residents + 18 Cohort 4 residents + 18 Cohort 5 residents = 50 total residents Y2: 10 Cohort 3 residents + 15 Cohort 4 residents + 10 Cohort 5 residents = 35 total residents Y2: 10 Cohort 3 residents + 15 Cohort 4 residents + 10 Cohort 5 residents = 35 total residents	Cohort 3: 2017–2018 (Y1), 2018–2019 (Y2), 2019–2020 (Y3); Cohort 4: 2018–2019 (Y1), 2019–2020 (Y2), 2020–2021 (Y3); Cohort 5: 2019–2020 (Y1), 2020–2021 (Y2), 2021–2022 (Y3)

TABLE G2. KEY COMPONENT 2: EQUITY-CENTERED PROFESSIONAL LEARNING

Indicators	Definition	Unit of implementation	Data sources	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 experienced educators to attend equity-centered PL 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		2017–2018, 2018–2019, 2019–2020, 2020–2021, 2021–2022
Indicator 2	Experienced Educators attend ECCF	Non-resident educators from CREATE schools	CREATE attendance rosters	Rosters emailed to evaluators from CREATE	0 = Educator attends 50% or fewer of days; 1 = Educator attends greater than 50% of 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. 40-80 non-resident educators	2017–2018, 2018–2019, 2019–2020, 2020–2021, 2021–2022
Indicator 3	Residents attend ECCF Institute or PPI by the end of Year 3 of their program. Residents may attend the ECCF Institute or PPI 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 50% or fewer of days 1 = Resident attends greater than 50% of days	Adequate implementation = score of 1	0 = Less than 30% of residents meet fidelity at the unit level 1 = 30% or more of residents meet fidelity at the unit level	10 residents (Cohort 3) 15 residents (Cohort 4) 10 residents (Cohort 5)	Cohort 3: 2017–2018, 2018–2019, 2019–2020 Cohort 4: 2018–2019, 2019–2020, 2020–2021 Cohort 5: 2019–2020, 2020–2021, 2021–2022

TABLE G2. KEY COMPONENT 2: EQUITY-CENTERED PROFESSIONAL LEARNING

Indicators	Definition	Unit of implementation	Data sources	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 ECCF meetings (during Together Time 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 0–4 meetings; 1 = Attend 5+ 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	Cohort 3: Y1: 14 residents Y2: 10 residents Y3: 10 residents Cohort 4: Y1: 18 residents Y2: 15 residents Y3: 15 residents Cohort 5: Y1: 18 residents Y2: 10 residents Y3: 10 residents	Cohort 3: 2017–2018, 2018–2019, 2019–2020 Cohort 4: 2018–2019, 2019–2020, 2020–2021 Cohort 5: 2019–2020, 2020–2021, 2021–2022

TABLE G2. KEY COMPONENT 2: EQUITY-CENTERED PROFESSIONAL LEARNING

Indicators	Definition	Unit of implementation	Data sources	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
All indicators							Years 1 and 2 (indicator 3 not measured) 0 = Fidelity not met= Fidelity not met for indicator 1, 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	Cohort 3: Y1: 14 residents Y2: 10 residents Y3: 10 residents Cohort 4: Y1: 18 residents Y2: 15 residents Y3: 15 residents Cohort 5: Y1: 18 residents Y2: 10 residents Y3: 10 residents Experienced Educators: 40-80 educators	Cohort 3: 2017–2018, 2018–2019, 2019–2020 Cohort 4: 2018–2019 2019–2020 2020–2021 Cohort 5: 2019–2020, 2020–2021, 2021–2022

TABLE G3. KEY COMPONENT 3: COMPASSION-BASED PROFESSIONAL LEARNING

Indicators	Definition	Unit of implementation	Data sources	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 compassion-based PL 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		2017–2018, 2018–2019, 2019–2020, 2020–2021, 2021–2022
Indicator 2	Residents attend monthly Together Time meetings that include PPI practices	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 0–4 meetings; 1 = Attend 5+ 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 3: Y1: 14 residents Y2: 10 residents Y3: 10 residents Cohort 4: Y1: 18 residents Y2: 15 residents Y3: 15 residents Cohort 5: Y1: 18 residents Y2: 10 residents Y3: 10 residents	Cohort 3: 2017–2018, 2018–2019, 2019–2020 Cohort 4: 2018–2019 2019–2020 2020–2021 Cohort 5: 2019–2020, 2020–2021, 2021–2022

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	Cohort 3: Y1: 14 residents Y2: 10 residents Y3: 10 residents Cohort 4: Y1: 18 residents Y2: 15 residents Y3: 15 residents Cohort 5: Y1: 18 residents Y2: 10 residents Y3: 10 residents	Cohort 3: 2017–2018, 2018–2019, 2019–2020 Cohort 4: 2018–2019 2019–2020 2020–2021 Cohort 5: 2019–2020, 2020–2021, 2021–2022

TABLE G4. KEY COMPONENT 4: MENTORING (CALCULATED IN YEAR 2 AND YEAR 3 ONLY)

Indicators	Definition	Unit of implementation	Data sources	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 90% of residents have mentors that meet fidelity at the unit level 1 = 90% or greater of residents have mentors that meet fidelity at the unit level	Cohort 3: Y2: 10 residents Y3: 10 residents Cohort 4: Y2: 15 residents Y3: 15 residents Cohort 5: Y2: 10 residents Y3: 10 residents	Cohort 3: 2017–2018, 2018–2019, 2019–2020 Cohort 4: 2018–2019, 2019–2020, 2020–2021 Cohort 5: 2019–2020, 2020–2021, 2021–2022
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	Cohort 3: Y2: 10 residents Y3: 10 residents Cohort 4: Y2: 15 residents Y3: 15 residents Cohort 5: Y2: 10 residents Y3: 10 residents	Cohort 3: 2017–2018, 2018–2019, 2019–2020 Cohort 3: 2018–2019, 2019–2020 Cohort 4: 2019–2020, 2020–2021 Cohort 5: 2020–2021, 2021–2022

TABLE G4. KEY COMPONENT 4: MENTORING (CALCULATED IN YEAR 2 AND YEAR 3 ONLY)

Indicators	Definition	Unit of implementation	Data sources	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	Y2 Residents: 0 = Resident attends 0–20 meetings; 1 = Resident attends 20–24 meetings; 2 = Resident attends 25 or more meetings Y3 Residents: 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	Cohort 3: Y2: 10 residents Y3: 10 residents Cohort 4: Y2: 15 residents Y3: 15 residents Cohort 5: Y2: 10 residents Y3: 10 residents	Cohort 3: 2018–2019, 2019–2020 Cohort 4: 2019–2020 2020–2021 Cohort 5: 2020–2021, 2021–2022

TABLE G4. KEY COMPONENT 4: MENTORING (CALCULATED IN YEAR 2 AND YEAR 3 ONLY)

Indicators	Definition	Unit of implementation	Data sources	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	Cohort 3: Y2: 10 residents Y3: 10 residents Cohort 4: Y2: 15 residents Y3: 15 residents Cohort 5: Y2: 10 residents Y3: 10 residents	Cohort 3: 2018–2019, 2019–2020 Cohort 4: 2019–2020, 2020–2021 Cohort 5: 2020–2021, 2021–2022
All indicators							Sum of indicator scores Component score ranges from 0–4: Fidelity not met = score of 0–3 Fidelity met = score of 4	Cohort 3: Y2: 10 residents Y3: 10 residents Cohort 4: Y2: 15 residents Y3: 15 residents Cohort 5: Y2: 10 residents Y3: 10 residents	Cohort 3: 2018–2019, 2019–2020 Cohort 4: 2019–2020, 2020–2021 Cohort 5: 2020–2021, 2021–2022

Note. SBM = school-based mentor. IM = instructional mentor.

TABLE G5. KEY COMPONENT 5: SUMMER RESIDENT ACADEMY (SRA)

Indicators	Definition	Unit of implementation	Data sources	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 SRA between Y1 and Y2	Residents	CREATE attendance rosters Resident surveys	Rosters emailed to evaluators from CREATE Surveys administered quarterly by evaluators	0 = Resident attends less than 15 days of SRA 1= Resident attends 15–20 days of SRA 2 = Resident attends more than 20 days of SRA	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	10 Cohort 3 residents + 15 Cohort 4 residents + 10 Cohort 5 residents = 35 total residents	Cohort 3: Summer 2018 Cohort 4: Summer 2019 Cohort 5: Summer 2020

Appendix H. 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 3, 4, and 5 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, Year 1 residents are placed in a classroom with a highly-skilled veteran teacher—known as a Cooperating Teacher (CT)—who serves as their mentor at a CREATE school. CREATE aims to have multiple CREATE residents placed in each school.

Upon graduating from the GSU CEHD teacher certification program, residents in Year 2 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 overwhelm 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, CREATE administrators consider the preferences of residents and determine who is 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 had the aforementioned scaffolded teaching experiences during the previous two years. 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 placed in the same building as another Y1 resident for the fall semester
    - 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 the classroom of an experienced 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. Column 3 indicates how many residents in Cohort 3 met the threshold for fidelity. Column 4 indicates how many residents in Cohort 4 met the threshold for fidelity. Column 5 indicates how many residents in Cohort 5 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 90% 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% of residents in Cohort 3 and 100% of residents in Cohort 4 and Cohort 5 met fidelity for this component in Year 1; all cohorts met fidelity for this component at the program level. More details about the number of residents who met each specific indicator are in the table below.

TABLE H1. DETAILS OF COMPONENT 1 IN YEAR 1

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 1: Resident is placed in the same building as another Y1 resident for fall semester	0 = Resident is not paired with a co-resident 1 = Resident is paired with a co-resident	12/14 (86%)	16/18 (89%)	16/18 (89%)
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%)	18/18 (100%)	16/18 (89%)
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%)	18/18 (100%)	18/18 (100%)
Program level	90% or more of residents meet fidelity on 2+ indicators	13/14 (93%) Fidelity Met	18/18 (100%) Fidelity Met	18/18 (100%) Fidelity Met

Result for Key Component 1 in Year 2

CREATE administrators intended to have 80% 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 residents in

Cohort 3, 87% of residents in Cohort 4, and 90% of residents in Cohort 5 met fidelity in Year 2; all cohorts met fidelity for this component overall. More details about the number of residents who met each specific indicator are in the table below.

TABLE H2. DETAILS OF COMPONENT 1 IN YEAR 2

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 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	9/10 (90%)	12/15 (80%)	7/10 (70%)
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	9/10 (90%)	13/15 (87%)	9/10 (90%)
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	10/10 (100%)	15/15 (100%)	10/10 (100%)
Program level	80% or more of residents meet fidelity on 2+ indicators	10/10 (100%) Fidelity Met	13/15 (87%) Fidelity Met	9/10 (90%) Fidelity Met

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 100% of Cohort 3 and Cohort 4 residents and 90% of Cohort 5 residents met fidelity in Year 3; all three cohorts met fidelity at the program level. More details about the number of residents who met each specific indicator are in the table below.

TABLE H3. DETAILS OF COMPONENT 1 IN YEAR 3

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 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	10/10 (100%)	13/15 (87%)	4/10 (40%)

TABLE H3. DETAILS OF COMPONENT 1 IN YEAR 3

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
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	10/10 (100%)	15/15 (100%)	8/10 (80%)
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	10/10 (100%)	14/15 (93%)	9/10 (90%)
Program level	85% or more of residents meet fidelity on 2+ indicators	10/10 (100%) Fidelity Met	15/15 (100%) Fidelity Met	9/10 (90%) Fidelity Met

Overall Results for Component 1

The CREATE program met fidelity on Key Component 1 for Cohorts 3, 4, and 5 in all three years of their participation. We can, therefore, conclude that CREATE successfully implemented the core progressive classroom roles from the 2017–18, 2018–19, and 2019–20 school years for all Cohorts 3, 4, and 5 residents.

Key Component 2: Equity-Centered Professional Learning

CREATE's equity-centered PL opportunities are designed to enhance pedagogical skills that interrupt the inequities that teachers encounter at their schools. ECCF—CREATE's core PL opportunity that is based on teaching equity-practices—equips educators to closely examine issues of oppression, power, and privilege present in school structures, and it address inequitable practices in education. While ECCF institutes are offered as a form of PL for all teachers at all CREATE schools and other educators, CREATE residents experience learning around ECCF practices at monthly Together Time meetings with other residents in their cohort. Each resident is also expected attend a four-day ECCF Institute or another CREATE PL opportunity in either Year 2 or Year 3 of their residency.

Prior to the 2019–20 school year, CREATE also offered Critical Friendship Institutes, another equity-centered PL opportunity. Critical Friendship Institutes fostered spaces for educators to discuss student work, educator work, and dilemmas of practice with groups of other educators. Critical Friendship Institutes, offered by CREATE, are counted as equity-centered PL for this component.

We present data on the following indicators related to equity-centered PL.

- Indicator 1 (Years 1, 2, and 3): CREATE administrators host 2 or more options for experienced educators to attend equity-centered PL each year (institutes were not held if fewer than 16 educators signed up).
- Indicator 2 (Years 1, 2, and 3): Experienced educators attend equity-centered PL.

- Indicator 3 (Years 2 and 3): Residents attend a CREATE PL (other than Together Time meetings) in Year 2 or Year 3 of their program.
- Indicator 4 (Years 1, 2, and 3): CREATE residents attend monthly ECCF meetings (during Together Time meetings).

Results of Key Component 2 for Year 1

CREATE administrators expect that Indicator 1 and at least one other indicator listed above will reach fidelity in Year 1. Cohorts 3, 4, and 5 met fidelity at the program level for this component. At the indicator level, Cohorts 3 through 5 met or exceeded fidelity thresholds of all indicators with the exception of Cohort 5 for Indicator 3; fewer than the expected 95% attended monthly ECCF trainings at Together Time meetings. 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, 4, and 5 correspond to the proper school year, rather than the residency cohort.

TABLE H4. DETAILS OF COMPONENT 2 IN YEAR 1

Indicator	Fidelity threshold	2017–18 school year	2018–19 school year	2019–20 school year
Indicator 1: **CREATE administrators host 2 or more options for experienced educators to attend an equity-centered PL each year**	0 = CREATE administrators host 0–1 institute 1= CREATE administrators host 2+ institutes	CREATE administrators host 2 institutes Fidelity met	CREATE administrators host 4 institutes Fidelity met	CREATE administrators host 3 institutes Fidelity met
Indicator 2: **Experienced educators attend equity-centered PL**	0 = Educator attends 50% of days of equity-centered PL 1 = Educator attends more than 50% of days of equity-centered PL Of the experienced educators who attend, 85% or more meet fidelity at the unit level	71/77 (92%) Fidelity met	69/78 (94%) Fidelity met	83/85 (98%) Fidelity met

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 4: **CREATE residents attend monthly ECCF meetings (during Together Time meetings)**	0 = Attend 0–7 meetings 1 = Attend 8–10 meetings 95% or more of residents meet fidelity at the unit level	14/14 (100%) Fidelity met	18/18 (100%) Fidelity met	10/18 (56%) Fidelity not met

TABLE H4. DETAILS OF COMPONENT 2 IN YEAR 1

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	Fidelity met

Note. Indicator 3 (Residents in year 2 or year 3 of their program attend a CREATE PL outside of Together Time) 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 expected that Indicator 1 and at least one other indicator would be met in Year 2, in order to reach FOI. The records from CREATE attendance rosters and survey data show that Cohorts 3, 4, and 5 met fidelity at the program level in Year 2 for this component. Fidelity was met for all indicators, with the exception of Indicator 4 for Cohort 4, when 67% of Cohort 4 residents attended at least six Together Time meetings during the 2019–20 school year, causing FOI calculations to fall below the fidelity threshold of 95% residents attending at least six Together Time meetings.

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

TABLE H5. DETAILS OF COMPONENT 2 IN YEAR 2

Indicator	Fidelity threshold	2018–19 school year	2019–20 school year	2020–21 school year
Indicator 1: **CREATE administrators host 2 or more options for experienced educators to attend equity-centered PL 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 3 institutes Fidelity met	CREATE administrators host 3 institutes Fidelity met
Indicator 2: **Experienced educators attend equity-centered PL**	0 = Educator attends 50% of days of equity-centered PL 1 = Educator attends more than 50% of days of equity-centered PL 85% or more meet fidelity at the unit level	54/62 (87%) Fidelity met	83/85 (98%) Fidelity met	45/45 (100%) Fidelity met

TABLE H5. DETAILS OF COMPONENT 2 IN YEAR 2

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 4: CREATE residents attend monthly ECCF meetings (during Together Time)	0 = Attend 0–5 meetings 1 = Attend 6+ meetings 95% or more of residents meet fidelity at the unit level	10/10 (100%) Fidelity met	10/15 (67%) Fidelity not met	10/10 (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	Fidelity met

Note. Indicator 3 (Residents in year 2 or year 3 of their program attend a CREATE PL other than Together Time) is included in the next (Y3) table.

Results of Key Component 2 for Year 3

CREATE administrators expected that Indicator 1, Indicator 3, and at least two other indicators in Year 3 would reach FOI. The records from CREATE attendance rosters and survey data show that both Cohort 3 and Cohort 4 met fidelity on Component 2 in Year 3. Cohort 5 did not meet fidelity for this component, because the fidelity threshold for Indicator 3, which measures the attendance of residents at CREATE PL outside of Together Time meetings. CREATE expected at least 30% of residents to attend a CREATE PL session by the end of Year 3 of their residency. Twenty percent of Cohort 5 residents attended PL during their residency, so fidelity was not met for this indicator and therefore the entire component for Cohort 5 in Year 3.

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

TABLE H6. DETAILS OF COMPONENT 2 IN YEAR 3

Indicator	Fidelity threshold	2019–20 school year	2020-21 school year	2021–22 school year
Indicator 1: CREATE administrators host 2 or more options for experienced educators to attend equity-centered PL	0 = CREATE administrators host 0–1 institute 1= CREATE administrators host 2+ institutes	CREATE administrators host 3 institutes Fidelity met	CREATE administrators host 3 institutes Fidelity met	CREATE administrators host 3 institutes Fidelity met

TABLE H6. DETAILS OF COMPONENT 2 IN YEAR 3

Indicator	Fidelity threshold	2019–20 school year	2020-21 school year	2021–22 school year
Indicator 2: Experienced teachers attend equity-centered PL	0 = Educator attends 50% of days of equity-centered PL 1 = Educator attends more than 50% of days of equity-centered PL 85% or more meet fidelity at the unit level	83/85 (98%) Fidelity met	45/45 (100%) Fidelity met	52/53 (98%) Fidelity met
Indicator	**Fidelity threshold**	**Cohort 3 met fidelity?**	**Cohort 4 met fidelity?**	**Cohort 5 met fidelity?**
Indicator 3: Residents attend a CREATE PL opportunity (Other than Together Time meetings) by the end of Year 3 of their program **Residents may attend the PL in Year 2 or Year 3**	0 = Resident attends 50% of PL days 1 = Resident attends more than 50% of PL days 30% or more meet fidelity at the unit level	7/10 (70%) Fidelity met	7/15 (67%) Fidelity met	2/10 (20%) Fidelity not met
Indicator 4: CREATE residents attend monthly ECCF meetings (during Together Time meetings)	0 = Attend 0–4 meetings 1 = Attend 5+ meetings 95% or more of residents meet fidelity at the unit level	10/10 (100%) Fidelity not met	15/15 (100%) Fidelity met	10/10 (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 two other indicators	Fidelity met	Fidelity met	Fidelity not met

Overall Results of Key Component 2

CREATE consistently met or exceeded the fidelity threshold for offering equity-centered PL, at which they achieved a high level of consistent attendance by experienced educators. Indicator 3—which measures CREATE residents' attendance at Together Time meetings, during which they receive equity-centered training—was not met by neither Cohort 5 in Year 3 nor Cohort 4 in Year 2; this was the 2019–20 school year during which the COVID-19 pandemic shifted programming to be virtual, and also affected the schooling, work, and personal lives of residents. Indicator 3, which is only measured in Year 3 (as it calculates residents' attendance at CREATE PL outside of Together Time meetings during the span of their residency), was met for Cohort 3 and Cohort 4 residents, but not for Cohort 5 residents. Similarly, we

expect the COVID-19 pandemic that affected all three years of the residency experience for Cohort 5 to have had an impact of residents' attendance at CREATE PL opportunities.

At the program level, fidelity was met for Cohorts 3, 4, and 5 for all but Year 3 (Cohort 5 did not meet fidelity for Indicator 3).

Key Component 3: Compassion-Based Professional Learning

Acknowledging that workplace and relationship stressors diminish a teacher's energy for teaching, all CREATE residents engaged in regularly scheduled compassion-based trainings throughout their residency. PPI—also offered as a PL opportunity for experienced educators—teaches techniques designed to help residents develop a greater awareness of their own attitudes and behaviors and how these attitudes and behaviors impact their relationships with others. Through becoming more aware of their thoughts and feelings as they occur in each moment, PPI teaches educators how to respond to stressful situations in a healthy manner and strengthen their ability to conjure empathy and compassion for themselves and others during stressful situations. Prior to the 2020–21 school year, CREATE offered Cognitively-Based Compassion Training (CBCT©), which used research-based strategies developed at Emory University to teach educators to focus their attention and become aware of their thoughts and feelings as they occur in each moment using progressive mental exercises and meditation. CBCT is included as compassion-based PL for this component.

We present data on the following indicators related to compassion-based PL.

- Indicator 1 (Years 1, 2, and 3): Program administrators offer at least one compassion-based PL opportunity per year to the general population of teachers at CREATE schools (and other educators)
- Indicator 2 (Years 1, 2, and 3): Residents attend monthly Together Time meetings that include compassion-based practices

Results of Key Component 3 for Year 1

In Year 1, CREATE expected to hold at least one compassion-based PL opportunity for the general population of teachers at CREATE schools and for other educators. CREATE exceeded this goal by offering two compassion-based institutes during the 2017–18 school year, two compassion-based institutes during the 2018–19 school year, and three compassion-based institutes during the 2019–20 school year. Additionally, they expected 95% or more of the residents to receive compassion-based training by attending at least eight Together Time meetings. They met this goal for Cohorts 3 and 4. Cohort 5 did not meet fidelity for this indicator, with 56% of residents having attended at least eight Together Time meetings. Therefore, Cohort 5 did not meet fidelity at the program level for this component. Cohort 3 and Cohort 4 met fidelity overall for Component 3.

TABLE H7. DETAILS OF COMPONENT 3 IN YEAR 1

Indicator	Fidelity threshold	2017–18 school year	2018–19 school year	2019–20 school year
Indicator 1: Program administrators offer at least one compassion-based PL for the 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	CREATE administrators host 3 institutes Fidelity met

TABLE H7. DETAILS OF COMPONENT 3 IN YEAR 1

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 2: Residents attend monthly Together Time classes that include compassion-based practices	0 = Attend 0–7 meetings 1 = Attend 8–10 meetings 95% or more of residents meet fidelity at the unit level	14/14 (100%) Fidelity met	18/18 (100%) Fidelity met	10/18 (56%) 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 met	Fidelity not met

Results of Key Component 3 for Year 2

In Year 2, CREATE expected to hold at least one compassion-based course for the general population of teachers at CREATE schools and for other educators. CREATE exceeded this goal by offering two compassion-based institutes each year during the 2018–19 and 2020–21 school years, and three compassion-based institutes during the 2019–20 school year. Additionally, they expected 95% or more of the residents to receive compassion-based training through attending at least six Together Time meetings in Year 2. CREATE met this goal for Cohort 3 and Cohort 5, but did not meet this goal for Cohort 4. At the program level, Cohort 3 and Cohort 5 met fidelity, but Cohort 4 did not meet fidelity for Component 3 in Year 2.

TABLE H8. DETAILS OF COMPONENT 3 IN YEAR 2

Indicator	Fidelity threshold	2018–19 school year	2019–20 school year	2020–21 school year
Indicator 1: Program administrators offer at least one compassion-based PL for the 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 3 institutes Fidelity met	CREATE administrators host 2 institutes Fidelity met

TABLE H8. DETAILS OF COMPONENT 3 IN YEAR 2

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Residents attend monthly Together Time classes that include compassion-based practices	0 = Attend 0–5 meetings 1 = Attend 6+ meetings 95% or more of residents meet fidelity at the unit level	10/10 (100%) Fidelity met	10/15 (67%) Fidelity not met	10/10 (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 met	Fidelity not met	Fidelity met

Results of Key Component 3 for Year 3

In Year 3, CREATE expected to hold at least one compassion-based PL opportunity for the general population of teachers at CREATE schools and for other educators. CREATE met this goal by offering three compassion-based institutes during the 2019–20 school year, two during the 2020–21 school year, and three during the 2021–22 school year. Additionally, they expected 95% or more of the residents to receive compassion-based training by attending at least five Together Time in Year 3. CREATE met this goal for Cohorts 3, 4, and 5. At the program level, all three cohorts met fidelity for Component 3 in Year 3.

TABLE H9. DETAILS OF COMPONENT 3 IN YEAR 3

Indicator	Fidelity threshold	2019–20 school year	2020–21 school year	2021–22 school year
Indicator 1: **Program administrators offer at least one compassion-based PL for the 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	CREATE administrators host 3 institutes Fidelity met

TABLE H9. DETAILS OF COMPONENT 3 IN YEAR 3

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 2: Residents attend monthly Together Time classes that include compassion-based practices	0 = Attend 0–4 meetings 1 = Attend 5+ meetings 95% or more of residents meet fidelity at the unit level	10/10 (100%) Fidelity met	15/15 (100%) Fidelity met	10/10 (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 met	Fidelity met	Fidelity met

Overall Results of Key Component 3

CREATE regularly offered more than one compassion-based PL opportunity a year for all years during which this indicator was measured. CREATE did not meet fidelity during the 2019–20 school year for Cohort 5 (Year 1) or Cohort 4 (Year 2) for Indicator 2, which measures residents' attendance at Together Time meetings at which they receive compassion-based training. The COVID-19 pandemic catalyzed many changes to schooling, the world of work, and personal life during this school year, so we suspect these changes and the need for adjustment to have directly affected the low attendance rates at Together Time meetings. Fidelity for this indicator was met for Cohort 3 in all three years of their residency, and for all cohorts in Year 3.

At the program level, CREATE met fidelity for the Compassion-Based PL component for Cohorts 3 and 4 during Year 1, Cohorts 3 and 5 during Year 2, and Cohort 3, 4, and 5 during 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 (SBMs), who are veteran educators working at 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). IMs are members of the CREATE staff who are trained in supporting residents during their development of compassion-based, equitable, effective teaching practices. Component 4 only applies to years 2 and 3, the years in which residents are paired with SBMs and IMs. CREATE residents do not receive SBMs in Year 1, as they are still student teaching through GSU CEHD and paired with CTs, in whose classrooms they are placed.

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): SBMs attend training during the summer prior to mentoring
- Indicator 2 (Years 2 and 3): SBMs attend training during their mentoring year (at least 2 sessions)
- Indicator 3 (Years 2 and 3): Residents attend semi-monthly meetings with their mentors (SBM and IM)
- Indicator 4 (Years 2 and 3): Residents participate in mentor-resident observation cycles with their CREATE IMs

Results of Key Component 4 for Year 2

CREATE expected that at least 90% of mentors would attend a summer training prior to beginning their work as a mentor and that at least 90% of them would attend two or more trainings during the school year, while they served as mentors. Additionally, CREATE expected that 90% of residents would meet with their mentors at least 25 times during the year, and that 90% of them would complete at least 2 observation cycles with their mentors during the year. CREATE met these goals for Cohorts 3 and 4. Cohort 5 did not meet fidelity on Indicators 1 and 2. This may be due to changing expectations for mentors, as they remained mentors year after year (after they have already received CREATE training in previous mentoring years). Because all indicators must be met in order for Component 4 to reach fidelity at the program level, Cohort 5 did not meet fidelity on this component for Year 2.

TABLE H10. DETAILS OF COMPONENT 4 IN YEAR 2

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 1: Residents have mentors (SBM) who attend training prior to mentoring	0 = Resident has a mentor who did not attend training during the summer prior to the mentoring year 1 = Resident has a mentor who attended at least one training session during the summer prior to the mentoring year 90% of residents have mentors that meet fidelity at the unit level	10/10 (100%) Fidelity met	15/15 (100%) Fidelity met	2/10 (20%) Fidelity not met
Indicator 2: Residents' mentors (SBM) attend training during their mentoring year (2 sessions)	0 = Resident has a mentor who attends 0–1 sessions 1 = Resident has a mentor attends 2 or more sessions 90% or more of residents have mentors that meets fidelity at the unit level	10/10 (100%) Fidelity met	15/15 (100%) Fidelity met	0/15 (0%) Fidelity not met

TABLE H10. DETAILS OF COMPONENT 4 IN YEAR 2

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 3: Residents attend semi-monthly meetings with their mentor (School Based Mentor or Instructional Mentor)	0 = Resident attends 0–20 meetings 1 = Resident attends 20–24 meetings 2 = Resident attends 25 or more meetings 90% of residents earn a score of 2 at the unit level and no residents earn a score of 0	9/10 (90%) receive score of 2 1/10 (0%) receive score of 1 None receive score of 0 Fidelity met	15/15 (100%) receive score of 2 None receive score of 1 None receive a score of 0 Fidelity met	10/10 (100%) receive score of 2 None receive score of 1 None receive a score of 0 Fidelity met
Indicator 4: Residents participate in mentor-resident observation cycles with CREATE Instructional Mentor	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	10/10 (100%) Fidelity met	15/15 (100%) Fidelity met	10/10 (100%) Fidelity met
Program level	All indicators meet fidelity	Fidelity met	Fidelity met	Fidelity not met

Results of Key Component 4 for Year 3

As mentioned above, CREATE expected that 90% of mentors would attend a summer training prior to beginning their work as a mentor and that at least 90% of them would attend two or more trainings during the school year in which they serve as a mentor. In Year 3, CREATE expected that 90% of residents would meet with their mentors at least twelve times during the year, and that 80% of them would complete at least 2–3 observation cycles with their mentors during the year. Fidelity for Indicator 1 was met for Cohort 3, but not for Cohort 4 or Cohort 5 in Year 3. Fidelity was not met for any of the three cohorts for Indicator 2 during Year 3. As Indicators 1 and 2 address SBM trainings, it is possible that expectations were adjusted for the repeat mentors who had already received training in mentoring from CREATE in previous mentoring years. CREATE met the goals for Indicator 3 and Indicator 4, which measured resident-mentor meetings and observation cycles, for Cohort 3, Cohort 4, and Cohort 5 in Year 3.

TABLE H11. DETAILS OF COMPONENT 4 IN YEAR 3

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 1: Residents have a mentor (SBM) that attended training prior to mentoring	0 = Resident has a mentor that did not attend training 1 = Resident has a mentor that attended training 90% of residents have mentors that meet fidelity at the unit level	5/5 (100%) Fidelity met	8/11 (73%) Fidelity not met	6/7 (86%) 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 0–1 sessions 1 = Resident has a mentor that attends 2 or more sessions 90% or more of residents have mentors that meet fidelity at the unit level	4/5 (80%) Fidelity not met	0/11 (0%) Fidelity not met	5/7 (71%) 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 90% of residents earn a score of 2 at the unit level and no residents earn a score of 0	10/10 (100%) receive a score of 2 Fidelity met	15/15 (100%) receive a score of 2 Fidelity met	10/10 (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	9/10 (90%) Fidelity met	14/15 (93%) Fidelity met	10/10 (100%) Fidelity met
Program level	All indicators meet fidelity	Fidelity not met	Fidelity not met	Fidelity not met

Note. In Year 3, not all CREATE residents were assigned an SBM based on their preferences.

Overall Results of Key Component 4

For Year 2, Indicators 1 and 2—regarding mentors attending mentor trainings—were not met for Cohort 5. For Year 3, Indicator 2 was not met for Cohort 3, and neither Indicator 1 nor Indicator 2 were met for Cohorts 4 and 5. CREATE consistently met fidelity for all years of residency Cohorts 3, 4, and 5 for Indicators 3 and 4, which measure participation in resident-mentor meetings and observation cycles.

Key Component 5: Summer Resident Academy

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. Key Component 5 is measured in Year 2, after residents have attended SRA. 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 expected that at least 95% of residents would attend 15–20 days of SRA, and that at least 85% of residents would attend more than 20 days of SRA.

TABLE H12. DETAILS OF COMPONENT 5 IN YEAR 2

Indicator	Fidelity threshold	Cohort 3 met fidelity?	Cohort 4 met fidelity?	Cohort 5 met fidelity?
Indicator 1: Residents attend Summer Resident Academy between Y1 and Y2	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	10/10 (100%) of active residents receive a score of 2	15/15 (100%) of active residents receive a score of 2	10/10 (100%) of active residents receive a score of 2
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 met	Fidelity met	Fidelity met

Overall Results of Key Component 5

Cohorts 3, 4, and 5 all met fidelity at the program level for this component.

Appendix I. Technical Details for Exploratory Intermediate Impacts on Teachers

Appendix I offers technical details relating to Chapter 4 of the final report. These details elaborate on the intermediate impacts on teachers' self-compassion, levels of burnout, and stress management and empathy related to teaching.

Research Questions

We examined average impacts on five key potential mediators. The evaluation of impacts spanned three cohorts of study participants. We addressed the following question concerning the intermediate outcomes.

- Is there a positive impact of CREATE on self-reported levels of self-compassion, burnout, and stress management and empathy related to teaching, three years after entry into the residency program?

Scale Details

Below we describe the five intermediate outcomes on which we examined impacts. They include important potential mediators of the effects of CREATE on more distal outcomes, such as retention of teachers in the profession. A goal of CREATE implementation is to equip teachers with skills that give them strategies to cope effectively with challenges of the profession, including potential stressors. The survey measures are meant to capture the more immediate changes.

Self-compassion

Cronbach's Alpha = .93

The Self-compassion scale (SCS) is a 26-item scale that assesses an individual's levels of self-compassion according to three main components: 1) self-kindness (versus self-judgment), 2) a sense of common humanity (versus isolation), and 3) mindfulness (versus over-identification). Respondents are asked to respond to a series of statements about their thoughts and feelings regarding the above three components on a 5-point scale of *Almost Never* to *Almost Always*. A large body of research demonstrates strong construct validity; thus, SCS scores correlate with wellbeing. The SCS administration takes place in the spring of each study school year. The complete scale is below. For this and the other scales on which we assessed impacts, we recoded reverse-coded items before averaging responses across items to create a total self-compassion score. A higher score on the scale indicates more self-compassion (Neff, 2003).

1. I'm disapproving and judgmental about my own flaws and inadequacies.
2. When I'm feeling down I tend to obsess and fixate on everything that's wrong.
3. When things are going badly for me, I see the difficulties as part of life that everyone goes through.
4. When I think about my inadequacies, it tends to make me feel more separate and cut off from the rest of the world.
5. I try to be loving towards myself when I'm feeling emotional pain.
6. When I fail at something important to me I become consumed by feelings of inadequacy.
7. When I'm down, I remind myself that there are lots of other people in the world feeling like I am.
8. When times are really difficult, I tend to be tough on myself.
9. When something upsets me I try to keep my emotions in balance.

10. When I feel inadequate in some way, I try to remind myself that feelings of inadequacy are shared by most people.
11. I'm intolerant and impatient towards those aspects of my personality I don't like.
12. When I'm going through a very hard time, I give myself the caring and tenderness I need.
13. When I'm feeling down, I tend to feel like most other people are probably happier than I am.
14. When something painful happens I try to take a balanced view of the situation.
15. I try to see my failings as part of the human condition
16. When I see aspects of myself that I don't like, I get down on myself.
17. When I fail at something important to me I try to keep things in perspective.
18. When I'm really struggling, I tend to feel like other people must be having an easier time of it.
19. I'm kind to myself when I'm experiencing suffering.
20. When something upsets me I get carried away with my feelings.
21. I can be a bit cold-hearted towards myself when I'm experiencing suffering.
22. When I'm feeling down I try to approach my feelings with curiosity and openness.
23. I'm tolerant of my own flaws and inadequacies.
24. When something painful happens I tend to blow the incident out of proportion.
25. When I fail at something that's important to me, I tend to feel alone in my failure.
26. I try to be understanding and patient towards those aspects of my personality I don't like

Maslach Burnout Inventory for Educators

Cronbach's Alpha

Teachers, among others who frequently work with people, are at risk for burnout. The Maslach Burnout Inventory for Educators assesses the three components of burnout.

1. Emotional exhaustion: depletion of emotional resources (a 9-item subscale, Cronbach Alpha =.93)
2. Depersonalization: negative feelings about one's students (a 5-item subscale, Cronbach Alpha =.86)
3. Reduced personal accomplishment: tendency to evaluate oneself negatively with regard to work (an 8-item subscale; Cronbach Alpha =.92)

Respondents read a series of statements and respond with how frequently they have the stated feelings or attitudes on a 6-point scale ranging from *Never* to *Everyday* (Maslach et al., 2018). The scale is proprietary, and therefore, we are not including the specific items composing the scale in this report. Higher scores on the emotional exhaustion and depersonalization subscales correspond to greater degree of experienced burnout. Lower scores on the personal accomplishment subscale correspond to greater degree of experienced burnout.

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 perspectives of my students.
6. Since starting this school year, I feel more able to understand the perspectives 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.

Methods

Sample

After limiting the sample to teachers with survey outcomes and non-missing covariates, and after matching teachers within-cohort, there were 59 teachers remaining across all three cohorts (20 from Cohort 1, 25 from Cohort 2, and 14 from Cohort 3). Tables I1, I2, and I3 show teacher sample sizes combined, and by condition, for (1) the full samples available for analysis, (2) the limited samples including only cases with non-missing covariate data, and (3) the sample after matching cases across conditions to establish baseline equivalence, in terms of the covariates used in the analysis.

TABLE I1. SAMPLE SIZES FOR FULL AND MATCHED SAMPLES FOR ANALYSIS OF SURVEYS (CREATE AND COMPARISON CASES COMBINED)

Sample	2018–19 posttest	2019–20 posttest	2020–21 posttest	2021–22 posttest
Cohort 3 full sample	29	30		
Cohort 3 matched sample	19	20		
Cohort 4 full sample	46	36	30	

TABLE I1. SAMPLE SIZES FOR FULL AND MATCHED SAMPLES FOR ANALYSIS OF SURVEYS (CREATE AND COMPARISON CASES COMBINED)

Sample	2018–19 posttest	2019–20 posttest	2020–21 posttest	2021–22 posttest
Cohort 4 matched sample	25	24	25	
Cohort 5 full sample		35	17	16
Cohort 5 matched sample		14	13	14

TABLE I2. SAMPLE SIZES FOR FULL AND MATCHED SAMPLES FOR ANALYSIS OF SURVEYS (CREATE CASES ONLY)

Sample	2018–19 posttest	2019–20 posttest	2020–21 posttest	2021–22 posttest
Cohort 3 full sample	8	8		
Cohort 3 matched sample	8	8		
Cohort 4 full sample	15	12	12	
Cohort 4 matched sample	11	11	11	
Cohort 5 full sample		16	9	8
Cohort 5 matched sample		8	8	8

TABLE I3. SAMPLE SIZES FOR FULL AND MATCHED SAMPLES FOR ANALYSIS OF SURVEYS (COMPARISON CASES ONLY)

Sample	2018–19 posttest	2019–20 posttest	2020–21 posttest	2021–22 posttest
Cohort 3 full sample	21	22		
Cohort 3 matched sample	11	12		
Cohort 4 full sample	31	24	18	
Cohort 4 matched sample	14	13	14	
Cohort 5 full sample		19	8	8
Cohort 5 matched sample		6	5	6

Impact Model

Analyses are based on matched samples of residents combined across Cohorts 3–5. We conducted matching within-cohort.

The impact model used has the following form.

$$Y_i = \beta_0 + \beta_{cohort4} C_{4i} + \beta_{cohort5} C_{5i} + \beta_T T_i + \sum_{p=1}^{P} X_{p,i} + \varepsilon_i \qquad (1)$$

The survey score of teacher i, Y_i, is expressed as the sum of an intercept term, β_0, an effect of cohort membership, $\beta_{cohort4}$, (C_{i4} being coded 1 if belonging to Cohort 4, and 0 otherwise), $\beta_{cohort5}$, (C_{i5} being coded 1 if belonging to Cohort 5, and 0 otherwise), β_T, an effect of being in treatment (T_i being coded 0 if belonging to comparison, and 1 if belonging to CREATE), a series of teacher-level covariates, $X_{p,i}$, and a term, ε_i, representing the random deviation of a person's score from the grand mean outcome, conditional on covariates in the model.

The reported standardized effect size consists of the regression-based impact estimate divided by the pooled standard deviation of the outcome variable.

Baseline Equivalence

To determine baseline equivalence, we regressed each of three measures used to test baseline equivalence against the indicator of treatment assignment status, dummy variables indicating membership in cohorts 4 or 5, and a random effect at the teacher level, parallel to the main impact model in Equation (1). Pre-intervention measures of the outcome variables were unavailable; therefore, we assess baseline equivalence on three covariates that we considered to be important in influencing survey outcomes.

Confidence in General Teaching Skills

Confidence in General Teaching Skills (α = .81) is 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.

Motivation for Entering Teaching

Motivation for Entering Teaching (α = .51) is 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 to enter into teaching.

Math Anxiety Scale

Math Anxiety Scale (α = .96) is 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.

All three scales achieved standardized mean differences of less than .25 (Table I4).

TABLE I4. SUMMARY OF BASELINE EQUIVALENCE TESTS FOR ANALYSIS OF IMPACTS ON SURVEY OUTCOMES

Scale	Sample size (CREATE)	Raw score mean (CREATE)	Raw score Standard Deviation (CREATE)	Sample size (comparison)	Raw score mean (comparison)	Raw score Standard Deviation (comparison)	Regression Adjusted Difference Estimate	Standardized Mean Difference
Confidence in general teaching skill	27	4.11	0.53	32	4.09	0.51	0.01	0.02
Motivation for entering teaching	27	2.36	0.98	32	2.57	1.01	-0.18	-0.18

TABLE I4. SUMMARY OF BASELINE EQUIVALENCE TESTS FOR ANALYSIS OF IMPACTS ON SURVEY OUTCOMES

Scale	Sample size (CREATE)	Raw score mean (CREATE)	Raw score Standard Deviation (CREATE)	Sample size (comparison)	Raw score mean (comparison)	Raw score Standard Deviation (comparison)	Regression Adjusted Difference Estimate	Standardized Mean Difference
Math anxiety	27	4.69	0.27	32	4.67	0.20	0.002	0.01

Note. N = 59. The standardized mean difference is the regression-adjusted difference estimate based on a model with dummy variables to indicate cohort membership and a variable indicating the treatment condition, divided by the pooled standard deviation of the pretest distribution.

Full Results from Impact Model

We report the full results of the analysis of the impact of CREATE on each of the survey outcomes.

TABLE I5. FIXED AND RANDOM EFFECTS OF SELF COMPASSION (SCALE 1)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	-1.99	1.86	51	-1.07	0.29
Impact of CREATE	0.34	0.18	51	1.93	0.06
Belongs to Cohort 5	0.002	0.21	51	0.01	0.99
Belongs to Cohort 4	0.29	0.20	51	1.43	0.16
Is a Black educator	0.004	0.19	51	0.02	0.98
Confidence in teaching	0.37	0.16	51	2.41	0.02
Motivation for teaching	0.73	0.40	51	1.82	0.07
Level of math anxiety	0.10	0.08	51	1.22	0.23
Random effects	**Estimate**	**Standard error**	**Z value**	**Pr > Z**	
Residual	0.31	0.06	5.43	<.0001	

TABLE I6. FIXED AND RANDOM EFFECTS OF BURNOUT (EMOTIONAL EXHAUSTION SCALE)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	11.56	4.63	51	2.50	0.02
Impact of CREATE	-0.29	0.44	51	-0.65	0.52

TABLE I6. FIXED AND RANDOM EFFECTS OF BURNOUT (EMOTIONAL EXHAUSTION SCALE)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Belongs to Cohort 5	0.54	0.53	51	1.01	0.32
Belongs to Cohort 4	-0.29	0.50	51	-0.59	0.56
Is a Black educator	-0.47	0.46	51	-1.02	0.31
Confidence in teaching	-0.96	0.39	51	-2.48	0.02
Motivation for teaching	-0.78	0.995	51	-0.78	0.44
Level of math anxiety	0.01	0.21	51	0.03	0.98
Random effects	**Estimate**	**Standard error**	**Z value**	**Pr > Z**	
Residual	1.94	0.36	5.43	<.0001	

TABLE I7. FIXED AND RANDOM EFFECTS OF BURNOUT (DEPERSONALIZATION SCALE)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	10.21	3.43	51	2.98	0.005
Impact of CREATE	-0.31	0.33	51	-0.96	0.34
Belongs to Cohort 5	-0.001	0.39	51	-0.00	0.998
Belongs to Cohort 4	-0.02	0.37	51	-0.05	0.96
Is a Black educator	-0.37	0.34	51	-1.07	0.29
Confidence in teaching	-0.77	0.29	51	-2.70	0.01
Motivation for teaching	-0.99	0.74	51	-1.34	0.19
Level of math anxiety	0.01	0.15	51	0.06	0.96
Random effects	**Estimate**	**Standard error**	**Z value**	**Pr > Z**	
Residual	1.06	0.20	5.43	<.0001	

TABLE I8. FIXED AND RANDOM EFFECTS OF BURNOUT (PERSONAL ACCOMPLISHMENT SCALE)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	1.47	2.45	51	0.60	0.55
Impact of CREATE	0.24	0.23	51	1.03	0.31
Belongs to Cohort 5	-0.16	0.28	51	-0.58	0.56
Belongs to Cohort 4	-0.01	0.27	51	-0.02	0.98
Is a Black educator	0.12	0.25	51	0.49	0.63
Confidence in teaching	0.64	0.20	51	3.11	0.003
Motivation for teaching	0.33	0.53	51	0.63	0.53
Level of math anxiety	0.07	0.11	51	0.65	0.52
Random effects	**Estimate**	**Standard error**	**Z value**	**Pr > Z**	
Residual	0.54	0.10	5.43	<.0001	

TABLE I9. FIXED AND RANDOM EFFECTS OF STRESS MANAGEMENT AND EMPATHY RELATED TO TEACHING

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	1.02	1.99	51	0.51	0.61
Impact of CREATE	0.22	0.19	51	1.16	0.25
Belongs to Cohort 5	-0.04	0.23	51	-0.17	0.87
Belongs to Cohort 4	0.16	0.22	51	0.73	0.47
Is a Black educator	0.27	0.20	51	1.35	0.18
Confidence in teaching	0.29	0.17	51	1.76	0.08
Motivation for teaching	0.35	0.43	51	0.83	0.41
Level of math anxiety	0.02	0.09	51	0.19	0.85
Random effects	**Estimate**	**Standard error**	**Z value**	**Pr > Z**	
Residual	0.36	0.07	5.43	<.0001	

Appendix J. Supplementary Results for Analysis of Impacts on Student ELA Achievement

Research Question

The impact evaluation of the CREATE teacher residency program addressed the following *confirmatory* research question regarding student achievement.

- Is there a positive impact of CREATE on student achievement in elementary and middle grades on average three years after the start of residency?

Measures

We collected student level data from the GaDOE: Georgia Milestones scores (as the outcome measure and pretest), student gender, age, grade level, race, ethnicity, special education status, and limited English proficiency status. More details about the data used in this analysis are available in the Appendix E. Schedule of Major Milestones.

Methods

Samples

We designed the impact evaluation to assess the confirmatory impacts of CREATE on **mathematics** and **ELA** achievement of students in grades 4–8, as measured by the Georgia Milestones Assessment System.

Given the availability of data, we evaluated impacts on students of novice teachers in the CREATE program across Cohorts 3–5, following the timeline below.

- For Cohort 3, in the third year of teaching (in the year after completion of the CREATE residency program)
- For Cohort 4, in the second year of teaching (in their third year of the CREATE residency program)
- For Cohort 5, in their first year of teaching (in their second year of the CREATE residency program)

We assessed student outcomes concurrently in both conditions.

Changes in Data Collection as a Result of COVID-19

The timing of outcomes data collection was changed from the original study plan as a result of COVID-19. In the original plan we stipulated that impacts on student achievement outcomes would be assessed three years after start of residency across all three cohorts. The adjusted plan assessed impacts three years after start of residency, on average (i.e., 4 years after start of residency for Cohort 3, 3 years after start of residency for Cohort 4, and 2 years after start of residency for Cohort 5). Refer to Table J1.

TABLE J1. SEED1 FINAL ANALYSIS OF IMPACTS ON STUDENT ACHIEVEMENT (SAMPLE AS PLANNED AND SAMPLE AS ANALYZED)

	2017–18	2018–19	2019–20	2020–21	2021–22
Cohort 3	Year 1	Year 2	Year 3	AS	AS
Cohort 4		Year 1	Year 2	Year 3	AS

TABLE J1. SEED1 FINAL ANALYSIS OF IMPACTS ON STUDENT ACHIEVEMENT (SAMPLE AS PLANNED AND SAMPLE AS ANALYZED)

	2017–18	2018–19	2019–20	2020–21	2021–22
Cohort 5			Year 1	Year 2	Year 3

Note. Confirmatory impacts on achievement were assessed in 20/21 for Cohorts 3 - 5 (shown in red font).

The proposed schedule was to assess confirmatory impact after three years across the three cohorts (marked with blue shading)

AS = After Study (i.e., after the three years of CREATE residency)

COVID-19 had three important effects on the impact study.

First, because no student outcomes data were available for the 2019–20 school year, we were unable to analyze impacts on Milestones outcomes for Cohort 3 in Spring 2020 as was planned in the original study proposal. Instead, we evaluated student outcomes in spring 2021 for Cohorts 3, 4, and 5 (on average in the second year as teacher of record).

Second, the lack of student achievement data from spring 2020 resulted in a two-year lag between when we collected baseline achievement measures (spring 2019) and when we assessed achievement outcomes (spring 2021). This reduced our sample size; we could not include Grade 4 posttests because Milestones pretests were not available from Grade 3 the year before, and Milestones is not tested in Grade 2, precluding use of pretests from two years before. In turn, the smaller sample reduced statistical power to detect impact. Statistical power was likely further reduced from the weaker predictiveness of a lagged pretest on posttest performance.

Third, COVID affected CREATE and BAU programming, so that impacts of the program reflect the unusual circumstances of schooling during the pandemic. For example, starting March 2020 through some of the 2020–21 school year, schooling, as well as CREATE PL experiences were virtual. Additionally, the first year of teaching for Cohort 5 and the second year of teaching for Cohort 4 began in a virtual setting. As described in the main report, several teachers in both conditions struggled with having to navigate their early career teaching experience during this disruption.

Matching

We matched students of CREATE teachers with those of comparison teachers on either the mathematics or ELA pretest within cohort and grade. We matched within cohort to ensure that the pretest scores we collected (Milestones assessments) were from the same assessment administration period. Because Milestones scale scores are not vertically scaled, it is not possible to compare scores across grades. To analyze effects combined across grades, we z-transformed scores within each grade and cohort.

In the process of defining the analysis sample, we took the following steps.

1. We removed fourth grade students because third grade pretests were unavailable for those students (a consequence of COVID-19).
2. We removed grade levels for which student outcomes were available only for treatment teachers or only for control teachers. (As noted above, Milestones is not a vertically scaled assessment, which requires a z-transformation of scores within each grade. If posttest scores are not available from both conditions, the z-transformed scores capture no information about impact.)

3. We matched cases within each cohort using the program *Matchit* in R (Ho, 2005; Ho et al., 2007), applying logit distances with nearest neighbor matching without replacement. We set the caliper, or standard deviation, of the propensity score within which comparison units were drawn to .25. The goal was to arrive at a sample of students of CREATE teachers who were close enough to their comparison counterparts to achieve equivalence on the pretest. If we could not find a comparison case that was sufficiently proximal to the CREATE case, we removed the CREATE case.

After matching, 29 students remained in the CREATE condition and 52 in the comparison condition. After applying these steps, for the ELA outcome, only students in sixth grade for five teachers were available for the impact analysis. No grade level remained to conduct analysis of impacts on mathematics.

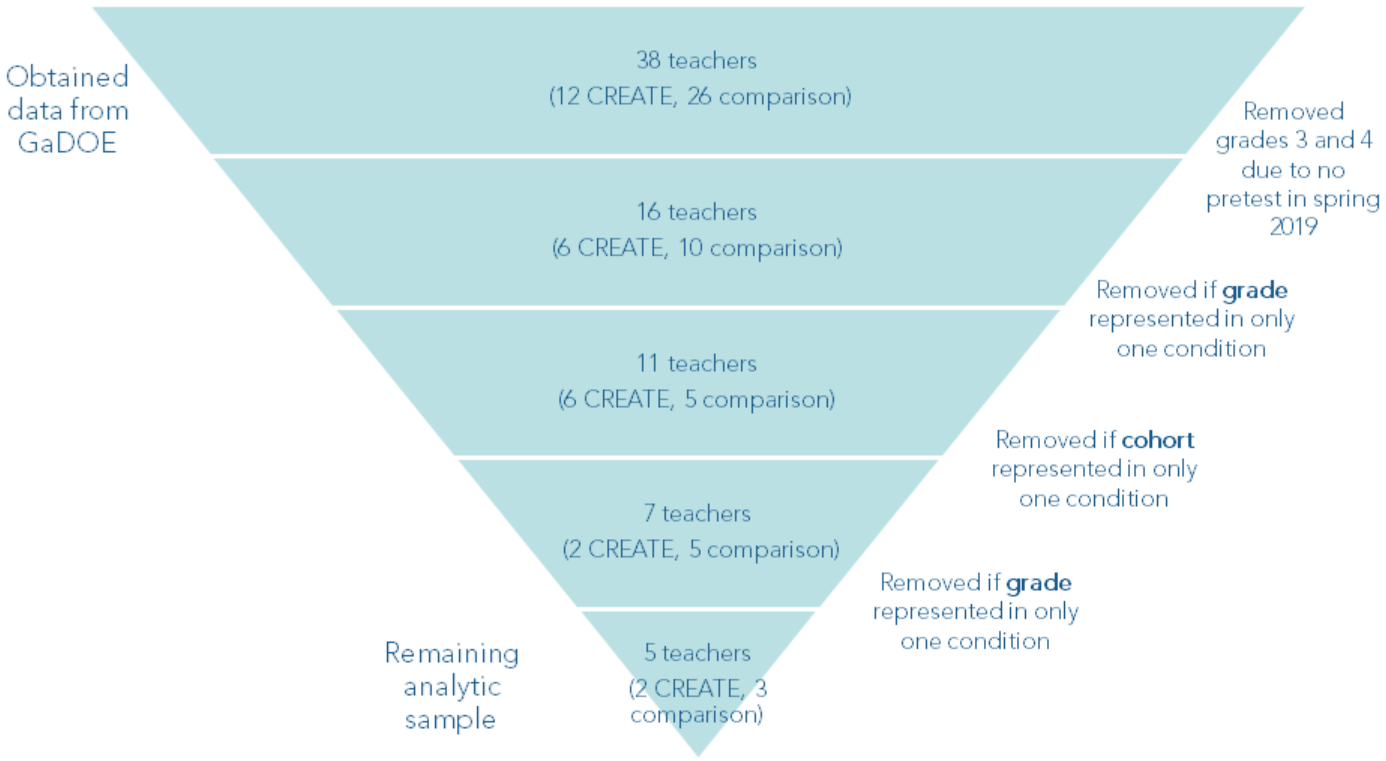

FIGURE J1. SAMPLE SIZE REDUCTION OF ELA MILESTONES OUTCOME

Impact Model

After matching students within grade and cohort (for the confirmatory analysis), we analyzed impacts on ELA. The impact model used to assess impacts on ELA had the following form.

$$Y_{ij} = \beta_0 + \beta_{cohort3} C_{3j} + \beta_T T_j + \sum_{p=1}^{P} X_{p,ij} + e_{0j} + \varepsilon_{ij} \quad (2)$$

We express the z-transformed posttest score for student i in the class of teacher j, Y_{ij}, as the sum of:

- an intercept term, β_0,

- an effect of cohort, $\beta_{cohort3}$, (C_{3j} being coded 1 if belonging to Cohort 3, and 0 if belonging to Cohort 4, with no cases present from Cohort 5),
- an effect of being in treatment (T_j being coded 0 if belonging to comparison, and 1 if belonging to CREATE),
- a series of student-level covariates $X_{p,ij}$ (the covariates included the pretest, gender, ethnicity, special education status, and ELL status), and
- terms for random deviations of scores at the teacher level from the grand mean outcome conditional on covariates in the model, e_{0j}, and for random deviation of scores at the student level from the respective teacher average conditional on covariates in the model, ε_{ij}.

Baseline Equivalence

To establish baseline equivalence, we regressed the pretest against the indicator of treatment status, a dummy variable indicating cohort, 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 = -.22 SD) expressed in units of the pooled (across conditions) standard deviation. The estimate associated with treatment was -10.55 scale score units, with a pooled SD of 48.29.

Full Impact Model

We report the results of the benchmark impact model in Table J2.

TABLE J2. FIXED AND RANDOM EFFECTS ESTIMATES

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	168.00	54.21	2	3.10	0.09
Impact of CREATE	-10.78	14.38	2	-0.75	0.53
Belongs to Cohort 3	34.98	14.95	68	2.34	0.02
Pretest	0.84	0.09	68	8.95	<.0001
Is Male	-9.33	4.35	68	-2.14	0.03
Race: Black	-97.60	6.83	68	-14.29	<.0001
Race: White	-62.61	12.47	68	-5.02	<.0001
Race: two or more	-71.03	31.94	68	-2.22	0.03
Ethnicity: Hispanic	-74.93	7.39	68	-10.14	<.0001
Has a Special Education designation	-29.42	5.64	68	-5.21	<.0001
Is English Proficient [a]	-32.51	9.21	68	-3.53	0.001

TABLE J2. FIXED AND RANDOM EFFECTS ESTIMATES

Random effects	Estimate	Standard error	Z Value	Pr > Z
Teacher effects	473.83	622.06	0.76	0.22
Residual	1245.22	213.45	5.83	<.0001

Note. N = 81 students. Reference category for ethnicity is Asian. Reference cohort is Cohort 5.

[a] We verified that the direction of this effect is correct, in spite of the counter-intuitive direction. The reference category for cohort is Cohort 4.

Exploratory Impact Findings

We examined impact using the same matching procedure as for the benchmark impact analysis, but using Restricted Maximum Likelihood Estimation. The estimated impact for this model was also -19.87 scale score units (p = .251).

Appendix K. Technical Details for Impacts on Early Career Teacher Retention

Research Questions

We examined average impacts on uninterrupted retention, two years after graduation from GSU. We evaluated the impacts across three cohorts of study participants. We addressed the following questions.

- Is there a positive impact of CREATE, compared to the GSU BAU program, on teacher retention three years after start of residency?
- Is there a positive impact of CREATE, compared to the GSU BAU program, on teacher retention three years after start of residency among Black educators?
- Is there a positive impact of CREATE, compared to the GSU BAU program, on teacher retention through the third year after start of residency, among teachers who completed teaching through the second year of residency?
- What is the impact of CREATE on completion of the teacher preparation program at GSU CEHD and teacher retention into the first year of teaching for the overall sample?
- Are impacts on teacher retention after 1, 2, and 3 years increasing by cohort for the overall sample and among Black educators?

Measures

We rely on a variety of sources to determine the status for each study participant at the three time points (graduation from GSU CEHD (Year 1 of CREATE/the study), completion of teaching through the first year (Year 2 of CREATE/the study), and completion of teaching through the second year (Year 3 of CREATE/the study)). For graduation from GSU CEHD, we relied on data from our participant tracker,[2] 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 (Georgia Professional Standards Commission, 2022). For teaching in Year 2 and teaching in Year 3 we triangulated data received on teacher surveys, data received from the CREATE program team, data from GaDOE, and teaching records from the Open Georgia: Transparency in Government travel and salary database (Open Georgia, 2008).

Each participant has a record indicating their early career trajectory for the first three years, with the first year covering graduation from GSU CEHD, the second indicating teaching status through the first year after graduation, and the third indicating teaching status through the second year after graduation. For each of the three years, we code outcomes for participants with a 0 (not graduated or not teaching), 1 (graduated or teaching), or 2 (unknown status).

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

[2] The participant tracker is a database of all study participants and their contextual information, including demographic characteristics, teacher preparation program, practicum and teaching placements (e.g., district, school, grade), data collection completion, and notes from any communication with or about the participant with GSU CEHD and the CREATE program team.

For Teaching through Year 2 (first year of teaching) and Teaching through Year 3 (second year of teaching), we first triangulated data received from the teacher placement and quarterly surveys, database received from the CREATE program team, communication with the CREATE program team, and 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 (where data was available), we designated the participant as teaching, and assigned them a 1 for teaching in the Georgia Public Schools system through 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 teaching in the Georgia Public School system through 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 through 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. We 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 the codes in Table K1.

TABLE K1. CODES ASSIGNED TO PARTICIPANTS

Code	Meaning
1	The participant had a successful match on the Open Georgia records. *Successful match* was defined as having one unique entry in Open Georgia 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 Open Georgia (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 Open Georgia (e.g., based on a non-unique name match in Open Georgia).
3	The participant did not have a match in Open Georgia but either graduated from GSU CEHD or taught in their second year.
5	The participant did not have a match in Open Georgia, and did not graduate from GSU CEHD or did not teach in the previous year (if applicable).

The team reasoned that not having a match in Open Georgia 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 this report, we collapsed the three unknown categories (2, 3, and 5) into one single category for unknown (i.e., lost to follow-up).

METHODS

Sample

Table K2 shows the years through which we examined impacts on graduation (G in Year 1), teaching through the first year (Year 2) and teaching through the second year (Year 3) for Cohorts 3–5. We evaluated confirmatory impacts on retention through completion of Year 3 (i.e., retention status at the point of completion of the second year of teaching in spring 2020 for Cohort 3, spring 2021 for Cohort 4, and spring 2022 for Cohort 5).

TABLE K2. TIMELINE FOR ANALYSIS OF IMPACTS ON RETENTION

	2017–18	2018–19	2019–20	2020–21	2021–22
Cohort 3	G in Year 1	Year 2	Year 3	AS	AS
Cohort 4		G in Year 1	Year 2	Year 3	AS
Cohort 5			G in Year 1	Year 2	Year3

Note. AS stands for After Study. G stands for Graduation.

This section about the sample displays sample sizes for analysis of impacts of CREATE on teacher retention. They are displayed for the full sample and for the sample of Black educators. We break down the numbers by cohort. We provide sample sizes for (a) all teachers recruited into study, (b) teachers with non-missing covariates (baseline measures for math anxiety, confidence in teaching, and motivation to teach), and (c) teachers matched on those baseline covariates. We matched cases within each cohort both by removing individuals with obviously extreme scores, and then by using the program *Matchit* in R (Ho, 2005; Ho et al., 2007), applying logit distances with nearest neighbor matching without replacement. The sample sizes in this section support the main analysis of impacts on retention and teaching through first and second years.

We conducted a separate analysis to evaluate the impact of CREATE on retention through the second year of teaching, among individuals who taught through the first year. The sample selection for that analysis is different. **To support WWC review, the sample includes teachers for whom we have available class composition variables, including class averages of student pretests, and class composition based on ethnicity. That analysis is described in Appendix L.**

Sample of all Educators

TABLE K3. ALL EDUCATORS COHORT 3 COMPLETE SAMPLE

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
CREATE group			
***n* with retention outcome**	14	11	11
***n* missing retention outcome**	0	3	3
***n* graduated/teaching**	14	10	10
Percent graduated/teaching among non-missing	100%	90.91%	90.91%
Percent missing	0.00%	21.43%	21.43%
Comparison group			
***n* with retention outcome**	47	39	26
***n* missing retention outcome**	1	9	22
***n* graduated/teaching**	43	30	23
Percent graduated/teaching among non-missing	91.49%	76.92%	88.46%
Percent missing	2.08%	18.75%	45.83%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	8.51%	13.99%	2.45%
***p* value**	0.57	0.42	1.00
Cox.index (effect size)	Inf	0.65	0.16
Difference in percent missing	-2.08%	2.68%	-24.40%
Total percent missing	1.61%	19.35%	40.32%

TABLE K4. ALL EDUCATORS COHORT 3 COMPLETE SAMPLE WITH NON-MISSING COVARIATES

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
CREATE group			
***n* with retention outcome**	14	11	11
***n* missing retention outcome**	0	3	3
***n* graduated/teaching**	14	10	10
Percent graduated/teaching among non-missing	100.00%	90.91%	90.91%
Percent missing	0.00%	21.43%	21.43%

TABLE K4. ALL EDUCATORS COHORT 3 COMPLETE SAMPLE WITH NON-MISSING COVARIATES

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
Comparison group			
***n* with retention outcome**	46	39	26
***n* missing retention outcome**	0	7	20
***n* graduated/teaching**	42	30	23
Percent graduated/teaching among non-missing	91.30%	76.92%	88.46%
Percent missing	0.00%	15.22%	43.48%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	8.70%	13.99%	2.45%
***p* value**	0.56	0.42	1.00
Cox.index (effect size)	Inf	0.65	0.16
Difference in percent missing	0.00%	6.21%	-22.05%
Total percent missing	0.00%	16.67%	38.33%

TABLE K5. ALL EDUCATORS COHORT 3 COMPLETE SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
CREATE group			
***n* with retention outcome**	14	11	11
***n* missing retention outcome**	0	3	3
***n* graduated/teaching**	14	10	10
Percent graduated/teaching among non-missing	100.00%	90.91%	90.91%
Percent missing	0.00%	21.43%	21.43%
Comparison group			
***n* with retention outcome**	43	37	25
***n* missing retention outcome**	0	6	18
***n* graduated/teaching**	39	28	22
Percent graduated/teaching among non-missing	90.70%	75.68%	88.00%
Percent missing	0.00%	13.95%	41.86%

TABLE K5. ALL EDUCATORS COHORT 3 COMPLETE SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	9.300%	15.23%	2.91%
***p* value**	0.56	0.42	1.00
Cox.index (effect size)	Inf	0.69	0.18
Difference in percent missing	0.00%	7.48%	-20.43%
Total percent missing	0.00%	15.79%	36.84%

TABLE K6. ALL EDUCATORS COHORT 4 COMPLETE SAMPLE

	Graduated in Year 1 (2018–19)	Teaching in Year 2 (2019–20)	Teaching in Year 3 (2020–21)
CREATE group			
***n* with retention outcome**	16	16	16
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	16	14	14
Percent graduated/teaching among non-missing	100.00%	87.50%	87.50%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	42	44	44
***n* missing retention outcome**	9	7	7
***n* graduated/teaching**	40	39	36
Percent graduated/teaching among non-missing	95.24%	88.64%	81.82%
Percent missing	17.65%	13.73%	13.73%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	4.76%	-1.14%	5.68%
***p* value**	1.00	1.00	0.72
Cox.index (effect size)	Inf	-0.06	0.26
Difference in percent missing	-17.65%	-13.73%	-13.73%
Total percent missing	13.43%	10.45%	10.45%

TABLE K7. ALL EDUCATORS COHORT 4 COMPLETE SAMPLE WITH NON-MISSING COVARIATES

	Graduated in Year 1 (2018–19)	Teaching in Year 2 (2019–20)	Teaching in Year 3 (2020–21)
CREATE group			
***n* with retention outcome**	15	15	15
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	15	13	13
Percent graduated/teaching among non-missing	100.00%	86.67%	86.67%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	37	39	39
***n* missing retention outcome**	8	6	6
***n* graduated/teaching**	36	35	32
Percent graduated/teaching among non-missing	97.30%	89.74%	82.05%
Percent missing	17.78%	13.33%	13.33%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	2.70%	-3.08%	4.62%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	Inf	-0.18	0.21
Difference in percent missing	-17.78%	-13.33%	-13.33%
Total percent missing	13.33%	10.00%	10.00%

TABLE K8. ALL EDUCATORS COHORT 4 COMPLETE SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2018–19)	Teaching in Year 2 (2019–20)	Teaching in Year 3 (2020–21)
CREATE group			
***n* with retention outcome**	15	15	15
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	15	13	13
Percent graduated/teaching among non-missing	100.00%	86.67%	86.67%
Percent missing	0.00%	0.00%	0.00%

TABLE K8. ALL EDUCATORS COHORT 4 COMPLETE SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2018–19)	Teaching in Year 2 (2019–20)	Teaching in Year 3 (2020–21)
Comparison group			
***n* with retention outcome**	37	39	39
***n* missing retention outcome**	8	6	6
***n* graduated/teaching**	36	35	32
Percent graduated/teaching among non-missing	97.30%	89.74%	82.05%
Percent missing	17.78%	13.33%	13.33%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	2.70%	-3.08%	4.62%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	Inf	-0.18	0.21
Difference in percent missing	-17.78%	-13.33%	-13.33%
Total percent missing	13.33%	10.00%	10.00%

TABLE K9. ALL EDUCATORS COHORT 5 COMPLETE SAMPLE

	Graduated in Year 1 (2019–20)	Teaching in Year 2 (2020–21)	Teaching in Year 3 (2021–22)
CREATE group			
***n* with retention outcome**	17	17	17
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	15	14	14
Percent graduated/teaching among non-missing	88.24%	82.35%	82.35%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	24	24	23
***n* missing retention outcome**	1	1	2
***n* graduated/teaching**	22	15	16
Percent graduated/teaching among non-missing	91.67%	62.50%	69.57%
Percent missing	4.00%	4.00%	8.00%

TABLE K9. ALL EDUCATORS COHORT 5 COMPLETE SAMPLE

	Graduated in Year 1 (2019–20)	Teaching in Year 2 (2020–21)	Teaching in Year 3 (2021–22)
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	-3.43%	19.85%	12.79%
***p* value**	1.00	0.30	0.47
Cox.index (effect size)	-0.23	0.61	0.42
Difference in percent missing	-4.00%	-4.00%	-8.00%
Total percent missing	2.38%	2.38%	4.76%

TABLE K10. ALL EDUCATORS COHORT 5 COMPLETE SAMPLE WITH NON-MISSING COVARIATES

	Graduated in Year 1 (2019–20)	Teaching in Year 2 (2020–21)	Teaching in Year 3 (2021–22)
CREATE group			
***n* with retention outcome**	17	17	17
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	15	14	14
Percent graduated/teaching among non-missing	88.24%	82.35%	82.35%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	24	24	23
***n* missing retention outcome**	1	1	2
***n* graduated/teaching**	22	15	16
Percent graduated/teaching among non-missing	91.67%	62.50%	69.57%
Percent missing	4.00%	4.00%	8.00%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	-3.43%	19.85%	12.79%
***p* value**	1.00	0.30	0.47
Cox.index (effect size)	-0.23	0.61	0.42
Difference in percent missing	-4.00%	-4.00%	-8.00%
Total percent missing	2.38%	2.38%	4.76%

TABLE K11. ALL EDUCATORS COHORT 5 COMPLETE SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2019–20)	Teaching in Year 2 (2020–21)	Teaching in Year 3 (2021–22)
CREATE group			
***n* with retention outcome**	17	17	17
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	15	14	14
Percent graduated/teaching among non-missing	88.24%	82.35%	82.35%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	17	17	17
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	16	13	14
Percent graduated/teaching among non-missing	94.12%	76.47%	82.35%
Percent missing	0.00%	0.00%	0.00%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	-5.88%	5.88%	0.00%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	-0.45	0.21	0.00
Difference in percent missing	0.00%	0.00%	0.00%
Total percent missing	0.00%	0.00%	0.00%

TABLE K12. ALL EDUCATORS COHORTS 3, 4, AND 5 COMBINED, COMPLETE SAMPLE

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
CREATE group			
***n* with retention outcome**	47	44	44
***n* missing retention outcome**	0	3	3
***n* graduated/teaching**	45	38	38
Percent graduated/teaching among non-missing	95.74%	86.36%	86.36%
Percent missing	0.00%	6.38%	6.38%

TABLE K12. ALL EDUCATORS COHORTS 3, 4, AND 5 COMBINED, COMPLETE SAMPLE

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
Comparison group			
***n* with retention outcome**	113	107	93
***n* missing retention outcome**	11	17	31
***n* graduated/teaching**	105	84	75
Percent graduated/teaching among non-missing	92.92%	78.50%	80.65%
Percent missing	8.87%	13.71%	25.00%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	2.82%	7.86%	5.72%
***p* value**	0.73	0.36	0.48
Cox.index (effect size)	0.33	0.33	0.25
Difference in percent missing	-8.87%	-7.33%	-18.62%
Total percent missing	6.43%	11.70%	19.88%

TABLE K13. ALL EDUCATORS COHORTS 3, 4, AND 5 COMBINED, WITH NON-MISSING COVARIATES

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
CREATE group			
***n* with retention outcome**	46	43	43
***n* missing retention outcome**	0	3	3
***n* graduated/teaching**	44	37	37
Percent graduated/teaching among non-missing	95.65%	86.05%	86.05%
Percent missing	0.00%	6.52%	6.52%
Comparison group			
***n* with retention outcome**	107	102	88
***n* missing retention outcome**	9	14	28
***n* graduated/teaching**	100	80	71
Percent graduated/teaching among non-missing	93.46%	78.43%	80.68%
Percent missing	7.76%	12.07%	24.14%

TABLE K13. ALL EDUCATORS COHORTS 3, 4, AND 5 COMBINED, WITH NON-MISSING COVARIATES

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	2.19%	7.62%	5.37%
***p* value**	0.73	0.36	0.63
Cox.index (effect size)	0.26	0.32	0.23
Difference in percent missing	-7.76%	-5.55%	-17.62%
Total percent missing	5.56%	10.49%	19.14%

TABLE K14. ALL EDUCATORS COHORTS 3, 4, AND 5 COMBINED, POST-MATCHING

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
CREATE group			
***n* with retention outcome**	46	43	43
***n* missing retention outcome**	0	3	3
***n* graduated/teaching**	44	37	37
Percent graduated/teaching among non-missing	95.65%	86.05%	86.05%
Percent missing	0.00%	6.52%	6.52%
Comparison group			
***n* with retention outcome**	97	93	81
***n* missing retention outcome**	8	12	24
***n* graduated/teaching**	91	76	68
Percent graduated/teaching among non-missing	93.81%	81.72%	83.95%
Percent missing	7.62%	11.43%	22.86%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	1.84%	4.33%	2.10%
***p* value**	1.00	0.63	1.00
Cox.index (effect size)	0.22	0.19	0.10
Difference in percent missing	7.62%	4.91%	16.34%
Total percent missing	5.30%	9.93%	17.88%

Sample of Black Educators

TABLE K15. BLACK EDUCATORS COHORT 3 COMPLETE SAMPLE

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
CREATE group			
***n* with retention outcome**	6	4	4
***n* missing retention outcome**	0	2	2
***n* graduated/teaching**	6	4	4
Percent graduated/teaching among non-missing	100.00%	100.00%	100.00%
Percent missing	0.00%	33.33%	33.33%
Comparison group			
***n* with retention outcome**	9	8	4
***n* missing retention outcome**	0	1	5
***n* graduated/teaching**	9	7	4
Percent graduated/teaching among non-missing	100.00%	87.50%	100.00%
Percent missing	0.00%	11.11%	55.56%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	12.50%	0.00%
***p* value**	1.00	1.00	1.00
Cox.index (effect size) [a]			
Difference in percent missing	0.00%	22.22%	22.22%
Total percent missing	0.00%	20.00%	46.67%

[a] Given the very small cell sizes, the logistic regression did not produce an estimate for the odds ratio.

TABLE K16. BLACK EDUCATORS COHORT 3 COMPLETE SAMPLE WITH NON-MISSING COVARIATES

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
CREATE group			
***n* with retention outcome**	6	4	4
***n* missing retention outcome**	0	2	2
***n* graduated/teaching**	6	4	4
Percent graduated/teaching among non-missing	100.00%	100.00%	100.00%
Percent missing	0.00%	33.33%	33.33%

TABLE K16. BLACK EDUCATORS COHORT 3 COMPLETE SAMPLE WITH NON-MISSING COVARIATES

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
Comparison group			
***n* with retention outcome**	9	8	4
***n* missing retention outcome**	0	1	5
***n* graduated/teaching**	9	7	4
Percent graduated/teaching among non-missing	100.00%	87.50%	100.00%
Percent missing	0.00%	11.11%	55.56%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	12.50%	0.00%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	Nan	Inf	Nan
Difference in percent missing	0.00%	22.22%	22.22%
Total percent missing	0.00%	20.00%	46.67%

TABLE K17. BLACK EDUCATORS COHORT 3 SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
CREATE group			
***n* with retention outcome**	6	4	4
***n* missing retention outcome**	0	2	2
***n* graduated/teaching**	6	4	4
Percent graduated/teaching among non-missing	100.00%	100.00%	100.00%
Percent missing	0.00%	33.33%	33.33%
Comparison group			
***n* with retention outcome**	6	6	3
***n* missing retention outcome**	0	0	3
***n* graduated/teaching**	6	5	3
Percent graduated/teaching among non-missing	100.00%	83.33%	100.00%
Percent missing	0.00%	0.00%	50.00%

TABLE K17. BLACK EDUCATORS COHORT 3 SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2017–18)	Teaching in Year 2 (2018–19)	Teaching in Year 3 (2019–20)
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	16.67%	0.00%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	Nan	Inf	Nan
Difference in percent missing	0.00%	33.33%	-16.67%
Total percent missing	0.00%	16.67%	41.67%

TABLE K18. BLACK EDUCATORS COHORT 4 COMPLETE SAMPLE

	Graduated in Year 1 (2018–19)	Teaching in Year 2 (2019–20)	Teaching in Year 3 (2020–21)
CREATE group			
***n* with retention outcome**	12	12	12
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	12	10	10
Percent graduated/teaching among non-missing	100.00%	83.33%	83.33%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	10	12	12
***n* missing retention outcome**	4	2	2
***n* graduated/teaching**	10	10	9
Percent graduated/teaching among non-missing	100.00%	83.33%	75.00%
Percent missing	28.57%	14.29%	14.29%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	0.00%	8.33%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	Nan	0.00	0.30
Difference in percent missing	-28.57%	-14.29%	-14.29%
Total percent missing	15.38%	7.69%	7.69%

TABLE K19. BLACK EDUCATORS COHORT 4 SAMPLE WITH NON-MISSING COVARIATES

	Graduated in Year 1 (2018–19)	Teaching in Year 2 (2019–20)	Teaching in Year 3 (2020–21)
CREATE group			
***n* with retention outcome**	12	12	12
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	12	10	10
Percent graduated/teaching among non-missing	100.00%	83.33%	83.33%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	10	12	12
***n* missing retention outcome**	4	2	2
***n* graduated/teaching**	10	10	9
Percent graduated/teaching among non-missing	100.00%	83.33%	75.00%
Percent missing	28.57%	14.29%	14.29%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	0.00%	8.33%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	Nan	0.00	0.30
Difference in percent missing	-28.57%	-14.29%	-14.29%
Total percent missing	15.38%	7.69%	7.69%

TABLE K20. BLACK EDUCATORS COHORT 4 SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2018–19)	Teaching in Year 2 (2019–20)	Teaching in Year 3 (2020–21)
CREATE group			
***n* with retention outcome**	12	12	12
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	12	10	10
Percent graduated/teaching among non-missing	100.00%	83.33%	83.33%
Percent missing	0.00%	0.00%	0.00%

TABLE K20. BLACK EDUCATORS COHORT 4 SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2018–19)	Teaching in Year 2 (2019–20)	Teaching in Year 3 (2020–21)
Comparison group			
***n* with retention outcome**	10	12	12
***n* missing retention outcome**	4	2	2
***n* graduated/teaching**	10	10	9
Percent graduated/teaching among non-missing	100.00%	83.33%	75.00%
Percent missing	28.57%	14.29%	14.29%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	0.00%	8.33%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	Nan	0.00	0.30
Difference in percent missing	-28.57%	-14.29%	-14.29%
Total percent missing	15.38%	7.69%	7.69%

TABLE K21. BLACK EDUCATORS COHORT 5 COMPLETE SAMPLE

	Graduated in Year 1 (2019–20)	Teaching in Year 2 (2020–21)	Teaching in Year 3 (2021–22)
CREATE group			
***n* with retention outcome**	10	10	10
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	9	8	8
Percent graduated/teaching among non-missing	90.00%	80.00%	80.00%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	10	10	10
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	9	8	8
Percent graduated/teaching among non-missing	90.00%	80.00%	80.00%
Percent missing	0.00%	0.00%	0.00%

TABLE K21. BLACK EDUCATORS COHORT 5 COMPLETE SAMPLE

	Graduated in Year 1 (2019–20)	Teaching in Year 2 (2020–21)	Teaching in Year 3 (2021–22)
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	0.00%	0.00%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	0.00	0.00	0.00
Difference in percent missing	0.00%	0.00%	0.00%
Total percent missing	0.00%	0.00%	0.00%

TABLE K22. BLACK EDUCATORS COHORT 5 SAMPLE WITH NON-MISSING COVARIATES

	Graduated in Year 1 (2019–20)	Teaching in Year 2 (2020–21)	Teaching in Year 3 (2021–22)
CREATE group			
***n* with retention outcome**	10	10	10
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	9	8	8
Percent graduated/teaching among non-missing	90.00%	80.00%	80.00%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	10	10	10
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	9	8	8
Percent graduated/teaching among non-missing	90.00%	80.00%	80.00%
Percent missing	0.00%	0.00%	0.00%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	0.00%	0.00%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	0.00	0.00	0.00
Difference in percent missing	0.00%	0.00%	0.00%
Total percent missing	0.00%	0.00%	0.00%

TABLE K23. BLACK EDUCATORS COHORT 5 SAMPLE WITH COVARIATES POST-MATCHING

	Graduated in Year 1 (2019–20)	Teaching in Year 2 (2020–21)	Teaching in Year 3 (2021–22)
CREATE group			
***n* with retention outcome**	10	10	10
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	9	8	8
Percent graduated/teaching among non-missing	90.00%	80.00%	80.00%
Percent missing	0.00%	0.00%	0.00%
Comparison group			
***n* with retention outcome**	10	10	10
***n* missing retention outcome**	0	0	0
***n* graduated/teaching**	9	8	8
Percent graduated/teaching among non-missing	90.00%	80.00%	80.00%
Percent missing	0.00%	0.00%	0.00%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.00%	0.00%	0.00%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	0.00	0.00	0.00
Difference in percent missing	0.00%	0.00%	0.00%
Total percent missing	0.00%	0.00%	0.00%

TABLE K24. BLACK EDUCATORS COHORTS 3, 4, AND 5 COMBINED, COMPLETE SAMPLE

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
CREATE group			
***n* with retention outcome**	28	26	26
***n* missing retention outcome**	0	2	2
***n* graduated/teaching**	27	22	22
Percent graduated/teaching among non-missing	96.43%	84.62%	84.62%
Percent missing	0.00%	7.14%	7.14%

TABLE K24. BLACK EDUCATORS COHORTS 3, 4, AND 5 COMBINED, COMPLETE SAMPLE

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
Comparison group			
***n* with retention outcome**	29	30	26
***n* missing retention outcome**	4	3	7
***n* graduated/teaching**	28	25	21
Percent graduated/teaching among non-missing	96.55%	83.33%	80.77%
Percent missing	12.12%	9.09%	21.21%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	-0.12%	1.28%	3.85%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	-0.02	0.06	0.16
Difference in percent missing	-12.12%	-1.95%	-14.07%
Total percent missing	6.56%	8.20%	14.75%

TABLE K25. BLACK EDUCATORS COHORTS 3, 4, AND 5 COMBINED, WITH NON-MISSING COVARIATES

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
CREATE group			
***n* with retention outcome**	28	26	26
***n* missing retention outcome**	0	2	2
***n* graduated/teaching**	27	22	22
Percent graduated/teaching among non-missing	96.43%	84.62%	84.62%
Percent missing	0.00%	7.14%	7.14%
Comparison group			
***n* with retention outcome**	29	30	26
***n* missing retention outcome**	4	3	7
***n* graduated/teaching**	28	25	21
Percent graduated/teaching among non-missing	96.55%	83.33%	80.77%
Percent missing	12.12%	9.09%	21.21%

TABLE K25. BLACK EDUCATORS COHORTS 3, 4, AND 5 COMBINED, WITH NON-MISSING COVARIATES

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	-0.12%	1.28%	3.85%
***p* value**	1.00	1.00	1.00
Cox.index (effect size)	-0.02	0.06	0.16
Difference in percent missing	-12.12%	-1.95%	-14.07%
Total percent missing	6.56%	8.20%	14.75%

TABLE K26. BLACK EDUCATORS COHORTS 3, 4, AND 5 COMBINED, POST-MATCHING

	Graduated in Year 1	Teaching in Year 2	Teaching in Year 3
CREATE group			
***n* with retention outcome**	28	26	26
***n* missing retention outcome**	0	2	2
***n* graduated/teaching**	27	22	22
Percent graduated/teaching among non-missing	96.43%	84.62%	84.62%
Percent missing	0.00%	7.14%	7.14%
Comparison group			
***n* with retention outcome**	26	28	25
***n* missing retention outcome**	4	2	5
***n* graduated/teaching**	25	23	20
Percent graduated/teaching among non-missing	96.15%	82.14%	80.00%
Percent missing	13.33%	6.67%	16.67%
Differences between CREATE and Comparison			
Difference in percent retained among non-missing	0.27%	2.47%	4.62%
***p* value**	1.00	1.00	0.73
Cox.index (effect size)	0.05	0.11	0.19
Difference in percent missing	-13.33%	0.48%	-9.52%
Total percent missing	6.90%	6.90%	12.07%

Impact Model

We based the analyses on samples of residents matched within-cohort and combined across Cohorts 3–5.

The impact model we used has the following form.

$$Y_i = \beta_0 + \beta_{cohort4} C_{4i} + \beta_{cohort5} C_{5i} + \beta_T T_i + \sum_{p=1}^{P} X_{p,i} + \varepsilon_i \quad (3)$$

The retention outcome for teacher i, Y_i, coded 1 for retained and 0 for not retained (for the linear probability model) is expressed as the sum of an intercept term, β_0, an effect of cohort membership, $\beta_{cohort4}$, (C_{i4} being coded 1 if belonging to Cohort 4, and 0 otherwise), $\beta_{cohort5}$, (C_{i5} being coded 1 if belonging to Cohort 5, and 0 otherwise), β_T, an effect of being in treatment (T_i being coded 0 if belonging to comparison, and 1 if belonging to CREATE), a series of teacher-level covariates, $X_{p,i}$, (the three variables used to test baseline equivalence and an indicator of whether a teacher is a Black educator) and a term, ε_i, representing the random deviation of a person's score from the grand mean outcome, conditional on covariates in the model.

The model for analysis of impact on Black educators is the same as the one above except without the covariate indicating whether a teacher is a Black educator.

The logistic regressions had the same form but with the log odds of retention as the outcome.

Baseline Equivalence

To determine baseline equivalence, we regressed each of three measures used to test baseline equivalence against the indicator of treatment assignment status, dummy variables indicating membership in cohorts 4 or 5, and a random effect at the teacher level, parallel to the main impact model in Equation (3). We assess baseline equivalence on three scales: (a) confidence in general teaching skills, (b) motivation to enter teaching, and (c) self-reported levels of math anxiety. All three scales achieved standardized mean differences of less than .25 (Tables K27 and K28).

TABLE K27. BASELINE EQUIVALENCE FOR ANALYSIS SAMPLE (FULL SAMPLE)

	Confidence in general teaching skill	Math anxiety	Motivation for entering teaching
Sample size (CREATE)	42	42	42
Score mean (CREATE)	3.96	2.35	4.56
Score standard deviation (CREATE)	0.71	0.95	0.64
Sample size (comparison)	79	79	79
Score mean (comparison)	3.93	2.67	4.41
Score standard deviation (comparison)	0.61	1.12	0.58
Raw difference	0.03	-0.32	0.15
Difference estimate	0.004	-0.29	0.12
Standardized mean difference	0.01	-0.28	0.20

Note. N = 121

TABLE K28. BASELINE EQUIVALENCE FOR ANALYSIS SAMPLE (BLACK EDUCATORS)

	Confidence in general teaching skill	Math anxiety	Motivation for entering teaching
Sample size (CREATE)	26	26	26
Score mean (CREATE)	4.02	2.43	4.47
Score standard deviation (CREATE)	0.54	0.95	0.78
Sample size (comparison)	25	25	25
Score mean (comparison)	3.90	2.52	4.38
Score standard deviation (comparison)	0.69	1.04	0.53
Raw difference	0.12	-0.10	0.10
Difference estimate	0.10	-0.09	0.13
Standardized mean difference	0.15	-0.09	0.21

Note. N = 51

Full Impact Model

Results of the benchmark impact (linear probability) model are in Tables K29 and K30.

TABLE K29. IMPACT OF CREATE ON RETENTION FOR THE FULL SAMPLE (LINEAR PROBABILITY MODEL)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	0.76	0.31	113	2.44	0.02
Impact of CREATE	0.02	0.07	113	0.21	0.83
Belongs to Cohort 3	0.06	0.09	113	0.71	0.48
Belongs to Cohort 4	0.03	0.08	113	0.44	0.66
Confidence in general teaching skills	-0.06	0.05	113	-1.15	0.25
Motivation for entering teaching	0.08	0.06	113	1.28	0.20
Level of math anxiety	-0.01	0.03	113	-0.33	0.75
Is a Black educator	-0.03	0.07	113	-0.44	0.66
Random effects	**Estimate**	**Standard error**	**Z Value**	**Pr > Z**	
Residual	0.12	0.02	7.78	<.0001	

Note. Reference category for Cohort is Cohort 5. N = 121 (three teachers were removed from the full sample of 124 because they had missing data on ethnicity).

TABLE K30. IMPACT OF CREATE ON RETENTION FOR THE SAMPLE OF BLACK EDUCATORS (LINEAR PROBABILITY MODEL)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	-0.03	0.47	44	-0.06	0.96
Impact of CREATE	0.03	0.10	44	0.27	0.79
Belongs to Cohort 3	0.22	0.16	44	1.43	0.16
Belongs to Cohort 4	0.02	0.11	44	0.14	0.89
Confidence in general teaching skills	-0.03	0.09	44	-0.33	0.74
Motivation for entering teaching	0.19	0.08	44	2.25	0.03
Level of math anxiety	0.03	0.05	44	0.61	0.54
Is a Black educator	-0.03	0.47	44	-0.06	0.96
Random effects	**Estimate**	**Standard error**	**Z Value**	**Pr > Z**	
Residual	0.13	0.03	5.05	<.0001	

Note. Reference category for Cohort is Cohort 5.

Appendix L. Impacts on Retention Through the Second Year of Teaching Among Teachers Who Were Retained Through the First Year of CREATE

We report impacts of CREATE on retention through the second year of teaching among teachers who were retained through the first year of teaching.

Why Are We Answering This Question?

After graduating from GSU CEHD, teachers in the CREATE and comparison groups can enter the teaching profession.

In their first year of teaching, CREATE participants, through co-teaching during their first year, begin a slower progression into the role of sole teacher-of-record. Teachers in the comparison group normally transition to a full teaching load right away. CREATE residents are offered additional supports during their first year as a classroom teacher, including mentorship from a CREATE-trained SBM and IM and monthly meetings with other CREATE residents (see Appendix A).

A contrast of these two models allows a test of the question of whether slower progression and additional supports impact longer retention, at least in the short horizon through their second year of teaching. The answer to this question may have value for the field generally (and it can support evaluation of whether longer retention offsets the added cost of co-teaching and other supports that CREATE offers).

In the context of this study, the question and corresponding analysis allows a test of a contrast reviewable according to WWC evidence standards, given the availability of pretest and demographic data for students in classes of teachers in their first year of teaching. These data allow an evaluation of the baseline equivalence of classes for teachers in both conditions in their first "baseline" year of teaching, with impacts assessed concurrently in CREATE and comparison groups on retention through the second year of teaching.

Method

We addressed this question with a view to WWC review using the Protocol for Teacher Excellence Version 4.0. CREATE belongs to the categories of "teacher preparation" and "teacher induction" programs. The eligible teacher outcome domain is "teacher retention in the state".

To achieve adequate statistical power, the study design draws on CREATE and comparison group participants from three consecutive cohorts. We analyzed impacts on retention for Cohort 3 through 2019–20 among individuals teaching in 2018–19, for Cohort 4 through 2020–21 among individuals teaching in 2019–20, and for Cohort 5 through 2021–22 among individuals teaching in 2020–21.

We used multiple sources to determine retention status in the base year and the outcome year. (See Appendix K for a complete description.)

Teachers were matched within cohort on the following variables, with goal of achieving baseline equivalence as per WWC requirements:

- Average years of teacher experience
- Standardized test scores of the teachers' students. We assessed equivalence at the student level using scores on the state Milestones achievement test.

- Student race/ethnicity. We assessed equivalence at the student level, using indicators of whether a student is a Black or Hispanic student.

Teachers are automatically matched on the first measure; that is, they all had the same years of teaching experience.

All teachers for whom student baseline data were available were matched on student pretests automatically (i.e., there was no need to limit the sample further). Additionally, a simple examination of the distribution of proportion Black or Hispanic students in classes of CREATE and comparison teachers revealed that the extreme of the lower-tail of the distribution (with the lowest-proportions of Black or Hispanic students), were all in the comparison conditions. We simply removed those cases. The remaining sample achieved balance on baseline achievement and the proportion of Black or Hispanic students.

Results are reported in four ways.

1. In terms of counts of teachers retained or not retained by condition
2. Using three linear probability models
    i. retention status (coded 1 for retained, and 0 as not retained) regressed against a treatment variable (coded 0 for comparison, and 1 for CREATE) and a random effect at the teacher level
    ii. Like (i) but with inclusion of dummy variables to indicate cohort
    iii. Like (ii) but with the inclusion of baseline covariates, teacher-averages of student pretests, and proportions minority students in teacher classes (student minority status was coded 1 for Black or Hispanic students, and 0 otherwise)
3. We report the results of Fisher's exact test of a difference between conditions in the proportion of teachers retained.
4. We also considered conducting a logistic regression and reporting both the Log Odds Ratio (LOR), as well as LOR(COX) as per WWC; however, a zero cell in the 2×2 table of counts precluded this option. We did explore calculation of log odds with application of a continuity correction.

We calculated student pretests and proportion of Black or Hispanic students whom the teacher inherited in his/her first year as teacher of record. The pretests and proportion of Black or Hispanic students were determined before the students joined the classes of teachers in their first year as teachers of record. For example, if the outcome for a teacher was retention status through 2020–21 (i.e., in their second year of teaching), we evaluated the composition of that teacher's class the year before, in their first year as a teacher of record in 2019–20. We typically obtained pretests on the Milestones state test in spring of the prior school year (spring 2019 in our example).

We z-transformed pretests within grade and subject area. We then averaged student scores across subject areas. Baseline equivalence (described below) is calculated at the student level for pretest and for minority status.

RESULTS

Impact

Counts

Of the 48 teachers (16 in CREATE and 32 in the comparison group) that we matched within cohort, based on student pretest and race in the base year, 16 CREATE teachers (100%), and 30 comparison teachers (94%), remained as teachers of record through the second year of teaching (Table L1).

TABLE L1. RETENTION RATES THROUGH THE SECOND YEAR OF TEACHING AMONG A MATCHED SAMPLE OF CREATE AND COMPARISON TEACHERS

		Retained through the second year of teaching	
		0	1
Comparison	0	2 (6.25%)	30 (93.75%)
CREATE	1	0 (0.00%)	16 (100.00%)

Note. Row percentages are in parentheses.

Analysis Using Linear Probability Models

Results of the three probability models are displayed in Tables L2, L3, and L4. The impact of CREATE on probability of retention in teaching through the second year, among teachers retained through the first year is .063 (p = .307), .075 (p = .224), and .080 (p = .184) for models a, b, and c, respectively, with the last considered the benchmark result.

TABLE L2. RESULTS USING LINEAR PROBABILITY MODEL (A)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	0.94	0.03	46	26.83	<.0001
Impact of CREATE	0.06	0.06	46	1.03	0.31
Random effects	**Estimate**	**Standard error**	**Z Value**	**Pr > Z**	
Residual	0.04	0.01	4.90	<.0001	

Note. N = 48 teachers.

TABLE L3. RESULTS USING LINEAR PROBABILITY MODEL (B)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	0.89	0.06	44	14.76	<.0001
Impact of CREATE	0.08	0.06	44	1.23	0.22

TABLE L3. RESULTS USING LINEAR PROBABILITY MODEL (B)

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Cohort 3	0.09	0.07	44	1.21	0.23
Cohort 4	0.04	0.07	44	0.52	0.60
Random effects	**Estimate**	**Standard error**	**Z Value**	**Pr > Z**	
Residual	0.04	0.01	4.90	<.0001	

Note. N = 48 teachers. Cohort 5 is the reference cohort.

TABLE L4. RESULTS USING LINEAR PROBABILITY MODEL (C), THE BENCHMARK MODEL

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	1.37	0.26	42	5.32	<.0001
Impact of CREATE	0.08	0.06	42	1.35	0.18
Cohort 3	0.07	0.07	42	0.93	0.36
Cohort 4	-0.001	0.07	42	-0.02	0.98
Pretest	-0.09	0.06	42	-1.33	0.19
Proportion of Black or Hispanic students	-0.54	0.28	42	-1.93	0.06
Random effects	**Estimate**	**Standard error**	**Z Value**	**Pr > Z**	
Residual	0.04	0.01	4.90	<.0001	

Note. N = 48 teachers. Cohort 5 is the reference cohort.

Fisher's Exact Test

A two-sided Fisher's exact test yielded a *p* value of .546.

Logistic Regression with Continuity Correction

Table L5 shows the expected cell frequencies based on the observed counts.

TABLE L5. EXPECTED CELL FREQUENCIES OF NUMBERS RETAINED

	Retained through the second year of teaching	
	no	**yes**
Comparison	1.33	30.66

TABLE L5. EXPECTED CELL FREQUENCIES OF NUMBERS RETAINED

	Retained through the second year of teaching	
	no	yes
CREATE	1.08	24.92

Pearson's Chi-squared test with Yates' continuity correction yielded a result of $\chi^2(1) = .047$, $p = .828$

Baseline Equivalence

Equivalence is satisfied for the base period. The outcome is teacher retention through the second year of teaching, and we demonstrate that the students taught by CREATE and comparison teachers during their first year of teaching are equivalent on academic performance and the distribution of race/ethnicity.

Teaching Experience

Among the 48 teachers in the analysis of impacts described in this section, all teachers had exactly the same number of years of experience. Within each cohort, they entered the teacher induction process at the same time, and they all taught in the first year.

Student Race/Ethnicity

We compared the difference between conditions in the log odds of students in teachers' classes being Black or Hispanic during teachers' first year of teaching (during the base period) (Table L6). LOR(COX) = .230.

TABLE L6. DIFFERENCE BETWEEN CONDITIONS IN LOG ODDS OF STUDENTS BEING BLACK OR HISPANIC IN TEACHERS' FIRST YEAR OF TEACHING

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	2.57	0.36	44	7.05	<.0001
Difference (CREATE-comparison)	0.38	0.36	3648	1.05	0.29
Cohort 3	-0.46	0.44	3648	-1.04	0.30
Cohort 4	-0.96	0.42	3648	-2.30	0.02
Random effect	**Subject**	**Estimate**	**Standard error**		
Intercept	teacher	1.04	0.28		

Note. Cohort 5 is the reference cohort. N (students) = 3,696. N (teachers) = 48.

TABLE L7. COUNTS OF TEACHERS AND STUDENTS IN THE SAMPLE USED TO ASSESS BASELINE EQUIVALENCE ON ETHNICITY

	N (teachers)	N (students)	N (students) Black or Hispanic = 1
CREATE	16	1204	1085
Comparison	32	2492	2073
Total	48	3696	3158

Student Pretest

We examined the difference between conditions in students' incoming academic achievement on the state standardized test in the classes of teachers in their first year of teaching (during the base period) (Table L8).

We z-transformed the pretests. The standard deviation in pretest in the treatment group is 1.042 and in the comparison group, it is 0.908. The standardized effect size is 0.018.

TABLE L8. DIFFERENCE BETWEEN CONDITIONS IN STUDENT BASELINE ACHIEVEMENT

Fixed effects	Estimate	Standard error	*df*	t value	Pr > |t|
Intercept	-0.22	0.15	44	-1.48	0.14
Difference (CREATE-comparison)	0.02	0.15	3648	0.11	0.91
Cohort 3	0.05	0.18	3648	0.27	0.79
Cohort 4	0.23	0.17	3648	1.38	0.17
Random effects	**Subject**	**Estimate**	**Standard error**	**Z Value**	**Pr > Z**
Intercept	teacher	0.20	0.04	4.41	<.0001
Residual		0.72	0.02	42.70	<.0001

Note. Cohort 5 is the reference cohort. N (students) = 3,696. N (teachers) = 48.

TABLE L9. COUNTS OF TEACHERS AND STUDENTS IN THE SAMPLE USED TO ASSESS BASELINE EQUIVALENCE ON BASELINE ACHIEVEMENT

	CREATE	Comparison	Total
N (teachers)	16	32	48
N (students)	1204	2492	3696
Pretest Mean	-0.09	-0.08	-0.08

TABLE L9. COUNTS OF TEACHERS AND STUDENTS IN THE SAMPLE USED TO ASSESS BASELINE EQUIVALENCE ON BASELINE ACHIEVEMENT

	CREATE	Comparison	Total
Pretest SD	1.04	0.91	Pooled SD: 0.95
Regression-adjusted difference		0.02	
Standardized ES based on regression-adjusted difference		0.02	

References

Elder, T., Wang, J., & Cramer, S. E. (n.d.). *An Evaluation of the Validity and Reliability of the Intern Keys Assessment*. University of Georgia College of Education. https://gaedassessment.files.wordpress.com/2018/05/intern-keys-validation-project-final-report-9-23-15.pdf

Elfers, A. M., Plecki, M. L., & Knapp, M. S. (2006). Teacher Mobility: Looking More Closely at "The Movers" Within a State System. *Peabody Journal of Education 81*(3):94-127. https://www.researchgate.net/publication/240519657_Teacher_Mobility_Looking_More_Closely_at_The_Movers_Within_a_State_System

Georgia Department of Education (GaDOE). (2020). *Teacher Keys Effectiveness System*. https://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Pages/Teacher-Keys-Effectiveness-System.aspx

Georgia Department of Education (GaDOE). (2019). *An Assessment & Accountability Brief: 2018-2019 Georgia Milestones Validity and Reliability*. https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Milestones/2018-19_Georgia_Milestones_Validity_and_Reliability_Brief.pdf

Georgia Professional Standards Commission (2022). *Check Certification Status*. https://www.gapsc.com/Certification/Lookup.aspx

Georgia Professional Standards Commission (GaPSC). (2023). *Induction*. https://www.gapsc.com/Certification/TieredCertification/induction.aspx

Ho, D. (2005). Matchit: *Matching Software for Causal Inference* [Computer software]. https://www.rdocumentation.org/packages/MatchIt/versions/1.0-1/topics/matchit

Ho, D., Imai, K., King, G., & Stuart, E. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis 15*, 199-236. http://gking.harvad.edu/files/abs/matchp-abs.shtml

Maslach, C., Jackson, S. E., & Leiter, M. P. (2018). *Maslach Burnout Inventory Manual: Fourth Edition*. Mind Garden, Inc.

Neff, K. D. (2003). Development and validation of a scale to measure self-compassion. *Self and Identity, 2*, 223-250. https://self-compassion.org/wp-content/uploads/publications/empirical.article.pdf

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

This appendix is a supplement to a report that was prepared for the Office of Elementary and Secondary Education (OESE) under Contract U423A170027 funded by the SEED grant program.

The full report, appendices, and research summary are at https://www.empiricaleducation.com/create/.

Reference the report: Jaciw, A. P., Lau, S., Zacamy, J., & Lin, L. (2023). Evaluation of Georgia State University's CREATE Teacher Residency Program – Cohorts 3 through 5: A Final Report of a Quasi-Experiment in Georgia. (Empirical Education Rep. No. Empirical_GSU-7035-FR-2023-O.2). Empirical Education Inc. https://www.empiricaleducation.com/create/