CRESST REPORT 826

INSPIRING MINDS THROUGH A PROFESSIONAL ALLIANCE OF COMMUNITY TEACHERS (IMPACT): EVALUATION RESULTS OF THE COHORT 1 MATH AND SCIENCE APPRENTICE TEACHERS

MARCH, 2013

Jia Wang

Jon Schweig

Noelle Griffin

Michelle Baldanza

Nichole M. Rivera

Vivian Hsu

National Center for Research on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

Inspiring Minds Through a Professional Alliance of Community Teachers (IMPACT): Evaluation Results of the Cohort 1 Math and Science Apprentice Teachers

CRESST Report 826

Jia Wang, Jon Schweig, Noelle Griffin,
Michelle Baldanza, Nichole M. Rivera, and Vivian Hsu
CRESST/University of California, Los Angeles

March 2013

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported by grant number U336S090159:01 from California Department of Education with funding to National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

We would like to thank J. Felipe Martinez for his feedback and input to this evaluation project. We would also like to thank Ron Dietel for his valuable contributions to this final report.

TABLE OF CONTENTS

INSPIRING MINDS THROUGH A PROFESSIONAL ALLIANCE OF COMMUNITY TEACHERS (IMPACT): EVALUATION RESULTS OF THE COHORT 1 MATH AND SCIENCE APPRENTICE TEACHERS

Jia Wang, Jon Schweig, Noelle Griffin,
Michelle Baldanza, Nichole M. Rivera, and Vivian Hsu
CRESST/University of California, Los Angeles

Abstract

This evaluation reports findings from a study of a UCLA teacher education program called IMPACT, Inspiring Minds through a Professional Alliance of Community Teachers. To measure program quality and goal attainment, the evaluation team used a comprehensive, multiple measures approach which included instructional artifacts, classroom logs, measures of pedagogical content knowledge, performance assessments, and teaching attitudes and beliefs. The evaluation team found that math and science teacher apprentices who completed the IMPACT program generally had a positive opinion of the program and applied what they learned in the classroom to their teaching. However, the team also found that the program did not significantly increase the pedagogical content knowledge of teachers nor contribute to substantial changes in teacher instructional strategies across lessons. Differences found in the experience and practices of math and science teacher apprentices suggest different support needs between the two groups. Study limitations and recommendations are discussed.

Introduction

UCLA IMPACT, *Inspiring Minds through a Professional Alliance of Community Teachers*, is a teacher education program developed in partnership with UCLA's Center X, the Los Angeles Unified School District (LAUSD), and the Los Angeles Small Schools Center. With the goal of preparing highly qualified community teachers and urban school teacher-leaders, IMPACT is an 18-month teacher education program in the high-need subject areas of math, science, and early childhood education. Participants receive a $10,000 stipend through a U.S. Department of Education Teacher Quality Partnership grant plus field support for the first two years of teaching. During year one of the IMPACT program, students, called apprentices, engage in summer foundational coursework followed by a year-long residency with a mentor teacher in a school within the LAUSD East Educational Service Center. At the end of year one, students have earned their California Preliminary Teaching Credential and are ready to become full-time teachers. In the fall of year two, teachers are supported in their own classrooms by UCLA faculty and work to develop a master's inquiry project before their graduation in December. Field support from UCLA faculty continues for the rest of the school year. IMPACT funding is expected to continue through September 2014.

For the purposes of program evaluation, research, and data-driven improvement, a collaborative team of researchers and practitioners from UCLA's National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and Center X recently studied IMPACT using multiple measures of teaching practice. These measures included observational data, classroom discourse analysis, performance assessments, instructional artifacts, logs, and student achievement data. Additional measures of pedagogical content knowledge (PCK) and teaching attitudes and beliefs (surveys) were collected during the apprentices' university coursework.

This report focuses on the first cohort of math and science teacher apprentices who started their UCLA training in the Fall of 2010 and who became full-time teachers in 2011-12. Subsequent pages describe the evaluation questions, a brief overview of the use of multiple measures to evaluate teaching quality, and the IMPACT program measures of teaching quality. Next, we describe our evaluation methodology, report results based on the two teacher self-reported measures (teacher surveys and classroom logs), and provide results on teacher knowledge and practices measures. Finally, we share the summary of major findings, limitations of our study, and recommendations for both program and evaluation improvement.

Evaluation Questions

Based on the program and evaluation objectives, our partnership developed the following questions:

1. What are the teacher apprentices' perceptions about the IMPACT program in terms of (a) coursework, (b) mentorship, (c) school placement, (d) community-based organizational resources, and (e) satisfaction with the IMPACT program overall? Does participation in the IMPACT program change the teacher apprentices' (f) priorities in their employment decisions, (g) beliefs in the causes of achievement gaps, or (h) beliefs in their roles as educators (social justice)?

2. (a) Do different teachers have different approaches to formative assessment? (b) Are some specific strategies implemented more than others? (c) Is there variation in the type of assessments used across days of instruction? (d) Is there evidence that teachers are differentiating their questioning strategies with respect to their instructional goals?

3. Does participation in the IMPACT program lead to increases in apprentices' pedagogical content knowledge as measured by (a) the Mathematical Knowledge for Teaching (MKT) for math or (b) the Assessing Teacher Learning About Science Teaching (ATLAST) for science?
4. What is the quality of teacher apprentices as measured by: (a) the Performance Assessment for California Teachers, and (b) the Instructional Quality Assessment?

Use of Multiple Measures

Capturing the complexity of teaching quality requires using more than one measure. However, *how* to use multiple measures to capture teaching quality is a highly visible and contentious area of education policy research. Much of the policy debate centers on the use of one measure—value-added student achievement data—in high-stakes district evaluations of practicing teachers. Critics contend that the relationship between student test scores, a single measure, and teacher quality are low at best (Hinchey, 2010). Even the use of multiple measures for teacher evaluation purposes has been criticized. The Measures of Effective Teaching (MET) Project (Kane et al., 2012), for example, is an ambitious large-scale study that analyzes observation data, student surveys, and teachers' pedagogical content knowledge to understand whether and how these measures are predictive of students' achievement growth and teacher quality. But this approach too has been criticized (Rothstein, 2011), in part due to the central focus on test score growth with a high percentage of a teacher's evaluation still based on improved student performance on high stakes state tests.

The teacher education community faces similar pressure, given the policy debate surrounding the value of traditional university-based pre-service preparation, the heightened scrutiny on the effects of these programs, and the proliferation of alternative pathways into teaching, including urban teacher residency programs. Residency proponents contend that teacher residency programs are the ideal context for producing high quality graduates as well as the opportunity to explore improved evaluation methods (Berry et al., 2008; Solomon, 2009). Teacher residents work alongside mentor teachers, but are not subject to a high stakes school evaluation process and have more time to develop teaching skills compared to shorter or less structured programs. Additionally, residency programs provide opportunities to collect data that can be built into teacher education programs and fed back into long-term program improvement. Teacher educators and mentors conduct regular observations of residents and collect participants' reflections on practice as well as artifacts of their teaching as part of their coursework. Residency programs also have an opportunity to assess residents' pedagogical content knowledge over time and measure the extent to which participants use specific instructional strategies. In short, proponents believe that the residency context produces both high quality teachers while advancing thoughtful measurement and improvement of teacher training systems.

IMPACT Program Measures

Based on the preceding discussion, the IMPACT research team worked with LAUSD teachers and schools on an ongoing basis to collect data from eight measures, as detailed in the following short descriptions.

1. Observations of pre-service and in-service teachers. During the program's pre-service component (Year 1: September-June) and in-service component (Year 2: September-December), UCLA Teacher Education Program (TEP) faculty advisors are in schools observing and giving feedback to teacher apprentices when they are enrolled as IMPACT students. The faculty advisors take qualitative notes and use a UCLA-developed rubric to support and understand pre-service teacher learning and practice. Graduate student researchers conduct observations to understand how the ideas and strategies taught in math and science methods translate into teacher practice. Observations occur 2 to 7 times per quarter for each teacher apprentice.

After the IMPACT students graduate, faculty advisors continue their teacher observations in schools and provide feedback for two quarters (January-June of Year 2). Observations occur 2 to 5 times per quarter for each teacher.

2. Classroom discourse analysis papers. With the goal of helping pre-service teachers engage students in high-level discussion in their content areas (math and science), apprentice teachers are asked to use transcripts of classroom discussion and write a discourse analysis paper. This assignment of writing a discourse analysis paper is given once during the apprentice teachers' pre-service year of teaching and collected as evidence of the quality of the apprentice teacher's classroom discourse.

3. Surveys of teacher program experience, attitudes, and beliefs. During the 18-month IMPACT program, apprentice teachers complete a three-part survey. Part 1 covers apprentice demographic information, benchmark information about apprentices' attitudes and beliefs about social justice and equitable practice, and the apprentices' long-term professional plans. Upon finishing the first year of the program, apprentice teachers take the Part 2 survey, measuring satisfaction with their coursework and mentors as well as their feelings of preparedness for teaching. Part 3 of the survey, administered at the end of 18-month program, collects information on apprentices' experiences as a first year teacher, their involvement with the IMPACT program (coursework, master's portfolio, and field support), their current teaching position, and growth in specific domains that are benchmarked in the Part 1 survey.

4. Assessment of teacher content knowledge. To measure apprentice teachers' pedagogical content knowledge (PCK), two assessment measures were adopted, one for math teachers and one for science teachers. Math apprentices took a test adapted from the Mathematics Knowledge for Teaching (MKT) survey, which originated as a part of the Learning Mathematics for Teaching project at the University of Michigan in 2000.[1] This survey has become one of the most widely used measures of teacher pedagogical content knowledge (Shulman, 1986) in the United States. A version of the survey appeared as one measure of the Content Knowledge for Teaching in the Measuring Effective Teaching (MET) project, one of the largest attempts to understand and measure classroom practice.

Science apprentices took an assessment based on the Assessing Teacher Learning About Science Teaching (ATLAST), developed by Horizon Research. ATLAST consists of a series of instruments that measure science teachers' knowledge for teaching at the middle school level. The instruments measure knowledge of science content, a teacher's ability to use science content to analyze student thinking, and a teachers' ability to use science content to make instructional decisions (Smith & Banilower, 2006).

5. Performance Assessment for California Teachers (PACT). PACT is a teacher performance assessment which pre-service teachers must pass in order to earn their elementary or single subject credentials. This assessment uses video, classroom artifacts, and writing from pre-service teachers' lessons to assess their skills in planning, instruction, assessment, academic language and reflection. Used by 30 California universities, this performance assessment records the classroom activities over 3-5 lessons and is graded using 12 detailed rubrics based on the Teacher Performance Expectations (TPEs). PACT scores are often used to measure teacher training programs' strengths and weaknesses. Pre-service teachers create and teach the lessons and videos required by PACT in winter and spring quarters.

6. Instructional Quality Assessment (IQA). IQA is an assessment of teacher practice that analyzes student work, teacher assignments, and teachers' descriptions of the context of their specific lessons. The purposes and benefits of the IQA include: (a) serves as both a summative and formative assessment of teaching quality, (b) assesses the quality of instruction with minimal burden on teachers, and (c) provides a strong correlation between instrument results and student achievement. Collected during the second half of the teachers' first year of teaching, this tool is a dependable measure of program quality effectiveness. IQA has been adopted as a measure of

[1] http://sitemaker.umich.edu/lmt/history

teacher quality because it is an existing CRESST measure that has high validity and reliability (Matsumura et al., 2006; Silk, Silver, Amerian, Nishimura, & Boscardin, 2009). Supportive of reliability with ample theoretical and research bases, the IQA "is uniquely positioned to provide useful formative feedback to instructional leaders and teachers to support school improvement through professional growth" (Crosson et al., 2006; p.2).

7. Logs of classroom practice. Logs are short surveys designed to collect self-reported data about teachers' daily instructional practices. The logs were used to assess the range of formative assessment strategies that IMPACT teachers utilized in their math and science classes to reflect upon the university methods of instructional practice. The logs were piloted with six pre-service math teachers in the spring of 2011 and formally used in March and May 2012 by the math and science program graduates.

8. California Standards Test scores. California Standards Test (CST) scores were collected from the program graduates' classes each year beginning with the spring 2012 CST administration. At the time of this report preparation, the first year of student CST data was not yet available for analysis, but will be added at a later time. These scores, with value-added modeling, will be used to understand how our program graduates' students' achievement compares to other teachers' students' achievement.

Aligned with UCLA's framework for learning, the preceding eight measures capture different types of information about teaching practice and quality. Each measure tells a story about how an apprentice teacher is performing in one or more specific areas. For example, the observation rubric allows teacher educators to score the frequency and quality of apprentice teachers' questioning strategies—including the extent to which questions promote student reasoning and conceptual understanding. In contrast, the Instructional Quality Assessment is an artifact-based measurement examining the quality of assignments that apprentice teachers give to students, as well as the quality of the students' work. The pedagogical content knowledge assessments provide evidence of apprentice teachers' knowledge of how to approach particular instructional challenges; for example, how to anticipate common errors in mathematical understanding.

Please note that the data on the first two measures were collected and reviewed for program information and improvement only. The data on the next six measures were collected for both program information improvement and program evaluation. This report analyzes and presents results based on program measures 3-7. As previously mentioned, the CST results (Measure 8) will be added when they become available.

Evaluation Methodology

In this section, we describe the data collected and data collection timeline, the descriptions of Cohort 1 math and science apprentice teachers, and our analytic strategies for answering the evaluation questions.

Data Collected and Data Collection Timeline

The current report analyzed the following data for Cohort 1 IMPACT math and science apprentice teachers during their 18-month training and post graduation:

- Three-part apprentice surveys of program experience, attitudes, and beliefs: administered at the beginning of the program, end of the year, and end of the program.
- Pre- and post- Pedagogical Content Knowledge (PCK) Assessments—Mathematics Knowledge for Teaching (MKT) and Assessing Teacher Learning About Science Teaching (ATLAST): Administered at the beginning and end of the program.
- Performance Assessment for California Teachers (PACT): Teacher apprentices took the test in spring, close to the end of their pre-service year.
- Instructional Quality Assessment (IQA): Administered after apprentice graduation from IMPACT program, in March Year 2 teaching.
- Log of classroom practices: Collected after apprentice graduation from the IMPACT program, one-week in March and May of Year 2 teaching.

Table 1 shows the data collection timeline for the above data:

Table 1

Timeline of IMPACT Data Collection: Cohort 1 Apprentice Teachers

Pre-Service			In-service (Year 1 Teaching)			In-service (Year 2 Teaching)
Fall	Winter	Spring	Fall	Winter	Spring	
PCK				PCK		
		PACT		Logs	Logs	
Survey		Survey		Survey		
				IQA	CST scores	CST scores

Description of Cohort 1 Teacher Apprentices

Twenty-three teacher apprentices completed the intake surveys in the summer of 2010. Approximately 68% of math and science apprentices were female. The largest racial groups were

Asian (39%), Hispanic or Latino (26%), and White (26%). Upon entry into the program, teacher apprentices' ranged in age from 21 to 39 years. The majority of teacher apprentices (78%) were 21 to 25 years old. Approximately 58% of math apprentices majored in a math-related field including mathematics, economics, or computer science. Half of the science apprentices (50%) majored in a science-related field including biology, physics, or computer science. Approximately 43% of math and science apprentices majored in other fields including business, anthropology, women's studies, music, psychology, and Chicano studies.

Analytic Strategies

Multiple analytic strategies were employed to analyze the data for each of the five measures: survey of teacher program experience, attitudes, and beliefs; logs of classroom practices; PCK measures for math and science apprentice teachers; PACT; and IQA. The following provides a brief description of how each type of data was analyzed and whenever appropriate, how the data was processed before the analysis.

Analysis of survey data. Descriptive analysis was employed to analyze the survey data by reporting the frequency and/or means and standard deviations for each item.

Analysis of MKT data. Based on answer keys provided by the instrument developer, each MKT item was scored dichotomously—with 1 indicating correct answers and 0 indicating incorrect answers. Answers that were left blank were treated as incorrect responses and were scored 0. Scores for each teacher apprentice were calculated as a percent correct—the sum of the correct items divided by the total number of items. Group means were computed by averaging individual scores. Analysis was conducted by comparing group means, both overall and on each subdomain. Group means and standard deviations were reported as percentages.

Analysis of ATLAST data. Based on answer keys provided by the instrument developer, each ATLAST item was scored dichotomously—with 1 indicating correct answers and 0 indicating incorrect answers. Blank answers were treated as incorrect responses and were scored as 0. Scores for each individual were added up to individual sum scores. Group means were computed by averaging individual sum scores. Analysis was conducted by comparing group means. Group means and standard deviations were reported in terms of raw scores.

Analysis of PACT data. Analysis of PACT data was conducted on group means, averaging scores across individuals. Group means were used so that they would be on the same scale as the rubrics. For example, a mean score of 3.5 could be interpreted as suggesting that a teacher (or a cohort of teachers) scored between exceeding standards and exceptional performance. Consistent with practices in the PACT technical report (Pecheone & Chung, 2007), double scores for individuals were averaged for tasks that were scored more than once.

Analysis of IQA data. Descriptive analysis was employed to analyze the IQA data in terms of classroom context, the mean and mode of the IQA scores on the dimensions measured, and IQA scores by subject (math or science).

Analysis of log of classroom practice data. Three different approaches were used in the analysis of log data. Item means were inspected in order to understand which instructional strategies were, on average, more popular with the typical teacher apprentice. Next, three different generalizability studies were conducted to decompose variance in teacher responses into components associated with four main effects (teachers, items, instructional days, and administration wave) and their interactions. Technically speaking, the three designs were configured as follows:

- A $p\ x\ i\ x\ o$ design, run separately for each administration. This design looked separately at the first wave of administration and the second wave of administration, and looked at three main effects: teachers (p), items (i) and instructional days (o).
- A $p\ x\ i\ x\ o$ design, run over all 10 administrations averaging over wave. This design looked at all 10 administrations of the log, regardless of administrative wave. The analysis looked at three main effects: teachers (p), items (i) and instructional days (o).
- A $p\ x\ i\ x(o{:}w)$ design, which took each set of administrations as being nested in a particular wave, then treated those waves as sources of variance (the waves were treated as randomly sampled from a pool of possible waves).

All generalizability studies were conducted on three groups of items included in the log survey. These groups included items measuring three different dimensions of classroom practice. One set of items asked teachers about the types of feedback they provided to students; a second set of items asked teachers about the types of questioning strategies they employed in their classrooms; and a third set asked teachers about specific instructional strategies they used in their classroom practice. Because complete responses were not available for all nine mathematics and nine science participants, analyses were done using the available data.

Lastly, a series of cross-tabs were analyzed, looking at each of the three item groups described above. The purpose was to investigate whether apprentice teachers were using instructional and questioning strategies that supported specific instructional goals.

Evaluation Results

There were two teacher self-reported measures: 1) surveys of teacher training experiences, attitudes, and beliefs; and 2) instructional logs of classroom practices. There were four teacher knowledge and practices measures: Pedagogical Content Knowledge for math teachers, Pedagogical Content Knowledge for science teachers, PACT, and IQA. We analyzed and reported results from each measure individually, with survey results broken into two sections: experience with the program components and pre- and post-IMPACT.

Evaluation Question 1 (a-e)

What are the teacher apprentices' perceptions about the IMPACT program in terms of: (a) coursework, (b) mentorship, (c) school site placement, (d) community-based organizational resources, and (e) satisfaction with the IMPACT program overall?

As described previously, apprentice teachers completed a three-part survey over the course of their 18-month IMPACT program: intake survey at the beginning of the program in summer 2010, second survey at the end of Year 1 in June 2011, and a final survey at the end of the IMPACT program in December 2011. Twenty-three apprentice teachers completed the intake surveys.[2] Their perceptions about the IMPACT program are presented below and organized by the program components—coursework, mentorship, school site placement, community based organizations, and lastly the overall program.

1a) Perceptions related to coursework. Upon completion of the academic year, teacher apprentices were moderately satisfied with their IMPACT methods coursework. On a scale from 1 "Strongly Disagree" to 5 "Strongly Agree" the average rating was 3.7 (*SD* .92). Additionally, just over half (60%) of the teacher apprentices indicated that they agreed or strongly agreed with the statement, "Overall, I feel satisfied with my methods coursework." However, when asked to rate their level of agreement with various statements about the helpfulness of their IMPACT methods coursework for learning a range of skills related to teaching, the results showed increased variability, suggesting some courses may have been more helpful than others.

The first five skills measured were directly related to the five PACT domains, while others were aligned to the IMPACT program's social justice emphasis and other program goals. Math and science apprentices felt that their IMPACT methods coursework prepared them for building reflective teaching practices and developed their identity as a social justice educator; the average ratings were 4.5 (*SD* .61) and 4.3 (*SD* 1.03). However, apprentices indicated that their methods coursework was less helpful for preparing them to use assessment results formatively, 3.55 (*SD* .95), and provide equitable access to content 3.79 (1.03).

[2] Three teacher apprentices left the program during the 18-month period.

Compared to their methods coursework, math and science apprentices showed a slightly lower level of overall satisfaction with their teacher inquiry course during the fall of 2011. The average rating was 3.4 (*SD* 1.1). Also, less than one-half (45%) of teacher apprentices indicated that they agreed or strongly agreed with the statement, “Overall, I am satisfied with weekly seminar courses I attended at UCLA.”

Consistent with program goals, teacher apprentices felt that the teacher inquiry course helped them to reflect on their experiences in urban schools; with an average rating of 4.2 (*SD* .696)—while indicating that the same course was least helpful for learning to provide equitable access to content; with an average rating of 2.35 (*SD* .813). Qualitative responses from teacher apprentices supported these findings. For example, teachers stated that talking about their experience as a first-year teacher within a group context was particularly helpful for debriefing after a stressful day and for reflecting on their practice. As a result, they felt close to other members of the group and built strong relationships with them over time. A typical response follows:

> The most helpful aspects were having a place to de-stress, share experiences, ask questions and hear that I wasn’t the only one feeling overwhelmed.

When asked how the course could be improved, the most prevalent responses were requests for more structure and content. Several people stated that it didn’t feel like an academic class because it was based so heavily on sharing personal experiences. As a result, Cohort 1 apprentices felt as if they did not aquire practical strategies for use within their classrooms. A typical qualitative response follows:

> I appreciate checking in each week, but I had expected there to be more content & teaching strategies. Specific topics and focus like teaching strategies and ideas we can implement in the classroom each week.

Teacher apprentices showed similar levels of overall satisfaction with the master’s portfolio. The average rating was 3.2 (*SD* .95), and one-half (50%) of apprentices indicated that they agreed with the statement, “Overall, I am satisfied with the master’s portfolio.” The master’s portfolio included a professional perspective, reflection on their apprentice year, and resident experience.

Consistent with requirements, teacher apprentices felt that the master’s portfolio was helpful for building reflective teaching practices and reflecting on their experiences in urban schools; the average ratings were 4.35 (*SD* .587) and 4.3 (*SD* .657) respectively. However, apprentices felt that the portfolio was less helpful for developing strategies for working with familes; with an average rating of 2.55 (*SD* .999). Qualitative responses further support these survey responses. Specifically, the majority of apprentices felt that developing their teaching philosophy was the most helpful aspect of the master's portfolio. A typical response follows:

> I think having to reflect and write the philosophy of education was definitely helpful in understanding our positions as teachers, especially in urban communities.

When asked what were the least helpful aspects of the master's portfolio, apprentice responses emphasized not having enough time and feeling rushed, as well as a lack of organization and clear expectations. A typical response states:

> I feel like the master's portfolio kept changing, which made it rather confusing for me at times to know what was expected. The master's portfolio also felt rushed, so I feel like my mark is not as good as it could have been.

More specifically, several apprentices mentioned that they ran out of time to finish their classroom management project, which is part of their professional perspective. This may explain why, on average, 40% of apprentices disagreed or strongly disagreed that the master's portfolio helped them develop classroom management strategies. One such response follows:

> Classroom management plan—no follow up for summer deadline it didn't get written before school started in the fall.

1b) Perceptions related to mentorship. Upon completion of their residency, teacher apprentices generally indicated that they were satisfied with their mentor relationship. The average rating was 4.05 (*SD* 1.15). Additionally, the majority of apprentices (80%) indicated they agreed or strongly agreed with the statement, "Overall, I feel satisfied with my relationship with my current mentor." However, apprentices also felt strongly that they should be placed with a knowledgeable mentor with whom they can get along. For example, a typical apprentice response follows:

> Allow apprentices to meet mentors before the program starts to see if it's a good fit. Also, filter mentors better and/or provide more training for mentors with apprentices.

This concern for fit may explain the wide variance in apprentice responses to this item. Although the mean rating was high, responses ranged from 2 to 5, indicating that some apprentices may have been dissatisfied with their mentor experience. Almost unilaterally, apprentices voiced a strong opinion that working with someone you can relate to and share similar ideas about teaching with, influences what you get out of the experience. For example two typical apprentice responses follow:

> Getting to know your mentor at the beginning of the school year and being able to choose your mentor based on teaching style, personality, etc.

> I would spend more time building a relationship and expectations between the mentor and mentee. My mentor and I got along, but I never felt like I got everything out of our relationship that I wanted. I was afraid to ask for more.

Math and science apprentices were also asked to rate their level of agreement with various statements related to the helpfulness of their IMPACT mentorship experience (residency) for learning a range of skills related to teaching. Consistent with the residency model where apprentices spend the majority of time interacting with mentors within the context of a real-world classroom, teacher apprentices felt that their mentor helped them prepare and implement lesson plans. The average ratings were 3.7 (*SD* 1.3) and 3.65 (*SD* 1.23) respectively. Teacher apprentices felt that their mentor helped them the least in developing their identity as a social justice educator, using assessment results formatively, and developing strategies to improve their students' academic discourse; the average ratings were 3.05 (*SD* 1.57), 3.15 (*SD* 1.18), and 3.15 (*SD* 1.18) respectively.

When asked about which aspects of the program's mentorship component were the most meaningful, teacher apprentices overwhelmingly said it was the direct experience of working with their mentor in a classroom. Specifically, apprentices praised the usefulness of observing their mentor's teaching methods and developing their own set of teaching practices with guidance. One apprentice noted:

> Learning classroom practice and observing my mentor, and being able to implement the practices I saw.

Over the course of the year, math and science apprentice roles shifted in the classroom. Teacher apprentices spent more time observing their mentor at the beginning of the school year and more time leading instruction independently at the end of the school year. This indicates that as the year progressed, apprentices were given more instructional responsibility in the average classroom. Furthermore, math and science apprentices engaged in a range of instructional activities with their mentor. As expected, teacher apprentices indicated that they often engaged in receiving feedback about their in-class instructional practices and co-designing/planning lessons; on a scale from 1 "Never" to 4 "Often" the average ratings were 3.25 (*SD* .716) and 3.2 (*SD* .77) respectively. The activities they engaged in the least with their mentor were discussing how to leverage out-of-classroom resources and communicating with the parents of their students; the average ratings were 1.7 (*SD* .66) and 1.85 (*SD* .81) respectively. See Appendix B regarding the demographics and perspectives of IMPACT mentor participants.

Over the course of the academic year, math and science apprentices engaged in various types of communication with their mentor. Teacher apprentices preferred in-person communication over all other types. Specifically, when asked about communication with their mentor, apprentices gave the highest levels of endorsement to working alongside their mentor during instructional time in the classroom as well as to the mentor faculty-advisor debrief; the average ratings were 4.2 (*SD* .89) and 3.22 (*SD* 1.26) respectively. They gave the lowest levels of endorsement to the online journal and the dialogue journal; the average ratings were 1.5 (*SD* 1.1) and 1.7 (*SD* .98) respectively. In fact, apprentices said the dialogue journal impeded their communication due to time constraints, which encouraged quick responses by both apprentices and mentors. Two apprentices noted:

> The dialogue journal did not encourage dialogue with my mentor. It was mainly a last minute compilation of questions I came up with at the last minute.

> The dialogue journal was difficult to implement, my mentor was very busy and did not make this a priority.

1c) Perceptions related to school site placement. On average, teacher apprentices indicated that they were satisfied with their school site placement. The average rating was 4.05 (*SD* 1.05), and the majority of apprentices (80%) indicated that they agreed or strongly agreed with the statement, "Overall, I feel satisfied with my current school placement."

Math and science apprentices were asked to rate their level of agreement with various statements related to their school site placement. These questions were designed to better understand their experience as a member of the school community. The majority of math and science apprentices (75%) agreed or strongly agreed that their school placement supported the development of their identity as a teacher; the average rating was 4.1 (*SD* .91). They also agreed that observing teachers other than their mentor and discussing instructional and classroom issues with other members of the school community were useful. The average ratings were 4.3 (*SD* .8) and 3.94 (*SD* .83) respectively. However, when asked to rate their agreement with the statement, "If given the opportunity, I would choose to teach at my current school placement in the future," teacher apprentices showed large variability. On a 5-point scale, average responses ranged from 1 to 5; with approximately 20% of apprentices strongly disagreeing. Sixty percent agreed or strongly agreed (*SD* 1.47). This finding suggests that a subset of math and science apprentices had a less positive experience at their school placement compared to others.

1d) Perceptions related to Community Based Organizations. IMPACT apprentices were asked to rate their overall perception of Community Based Organizations (CBO) as a valuable resource. The average rating was 3.95 (*SD* 1.1). Further, the majority of teacher apprentices (75%) indicated that they agreed or strongly agreed with the statement, "Overall, I consider these community-based organizations to be valuable resources."

Math and science apprentices interacted with CBOs several times over the course of the IMPACT program and were asked to rate their level of agreement regarding the helpfulness of the CBOs for a range of outcomes. On average, apprentices' endorsement was relatively low. For example, the highest level of endorsement was given to the statement, "I feel the community-based organizations helped me understand the strengths of the community I teach in." However the average rating was 3.3 (*SD* 1.17) and only 50% of teacher apprentices agreed or strongly agreed with this statement. The lowest level of endorsement was for the statement, "I feel the community-based organizations helped me build relationships with families." The average rating was 2.6 (*SD* 1.14). In combination, the responses suggest that while the majority of teacher apprentices saw value in the CBOs, the specific interactions with CBOs may not have been very helpful.

1e) Perceptions related to the IMPACT program overall. Upon completion of their residency year, teacher apprentices indicated a moderate level of satisfaction with the IMPACT program. The average rating was 3.6 (*SD* .82), with 60% of teacher apprentices indicating that they agreed or strongly agreed with the statement, "Overall, I feel satisfied with the IMPACT program."

When the apprentices reflected on the overall IMPACT program, they indicated that the program helped them to build a sense of community between each other. They strongly endorsed the statement, "The IMPACT program helped me get to know the other students in my program well," with an average rating of 4.16 (*SD* 0.9). However, three questions having to do with relationships outside of the school—build partnerships with allies off campus, engage parents in student's education, and communicate with parents from diverse linguistic and cultural backgrounds—received the lowest levels of endorsement, nearly one scale point below the relationship-building item.

Overall, apprentices believed that the amount of time that was dedicated to the three main program components during their residency year—methods coursework, fieldwork, and other coursework, was reasonable. They reported spending approximately 60% of their time engaged in their residency and approximately 40% (split evenly) among their methods and non-methods coursework. When asked to indicate what they perceived to be the ideal time allocation, the split was nearly identical.

However, despite the fact that the balance of activities seemed appropriate, teacher apprentices voiced their concern that the course requirements were very demanding, and at times, very stressful. For example, one apprentice wrote:

> I sometimes had difficulty managing both coursework and fieldwork. It seemed like there was not enough communication with other TEP [UCLA Teacher Education Program] teachers about our expectations. I felt overwhelmed often.

During the fall of 2011 when teacher apprentices' began full-time teaching, they were asked to rate their level of agreement with statements regarding how well the IMPACT program had prepared them for teaching. The responses showed large variability across items. Specifically, they showed the highest level of agreement when asked if the IMPACT program helped them to build reflective teaching practices and to work as social justice educators. The average ratings were 4.5 (*SD* .607) and 4.2 (*SD* .89) respectively. However, average agreement was lower when asked if the program prepared them with practical classroom management strategies (2.65, *SD* 1.309), to work with special education students (2.65, *SD* 1.309), to engage parents in their student's education (2.95, *SD* .970), to communicate with parents from diverse backgrounds (2.9, *SD* 1.119), and to effectively manage teaching multiple class periods (2.75, *SD* 1.682).

Summary of Results. In summary, teacher apprentices indicated higher levels of satisfaction with multiple aspects of the IMPACT program (i.e., mentorship and school placement), relative to their levels of satisfaction with coursework and the IMPACT program overall. This may reflect the heavy burden of the program requirements including a large course load and many time constraints. Further evidence for this conjecture can be found in teacher apprentices' lower ratings of overall satisfaction with the various program components during the fall 2011 (i.e., teacher inquiry course, master's portfolio), when apprentices showed increased time constraints as they began working full-time as math and science teachers, attended classes at UCLA, and took steps toward completing their master's degree in education.

In general, the IMPACT program reached its goal of training and preparing teacher apprentices to teach at urban schools. Having received training and help from various program components, teacher apprentices indicated that they learned a range of teaching skills related to the Performance Assessment for California Teachers domains and social justice. The exceptions were their preparedness in the areas of communicating with parents, engaging out-of-classroom resources, and working with diverse populations of students including special education students. This was consistent with ratings related to the amount of time that teacher apprentices spent with their mentors, communicating (or discussing communications) with parents, and discussing how to leverage out-of-classroom resources compared to other activities.

Evaluation Question 1 (f-h)

Does participation in the IMPACT program change teacher apprentices:' (f) priorities in their employment decisions, (g) beliefs in the causes of achievement gaps, and (h) beliefs about their roles as educators (social justice)?

IMPACT math and science apprentices were asked several sets of questions concerning social justice and equitable practice when they initially entered the program. These same questions were asked in December 2011 at the conclusion of their formal involvement with UCLA. The following analyses focus only on items that showed the largest differences between pre- and post-surveys.

1f) Priorities in employment decisions. The first set of questions concerned apprentices' priorities in their employment decisions. Figure 1 shows that (a) choice of grade level, (b) schools in which the apprentices have field experiences, and (c) good conditions of facilities, were relatively more important (between one-third to one-half scale point) on the post survey compared to the pre-survey. Apprentices also indicated a stronger desire to work in an urban location and to work with high poverty and ELL students (about one quarter scale point) on the post-survey. Having friends working in the school and school size were relatively less important. The importance of supportive school leadership was also lower on the post-survey compared to the initial survey (about one-half scale point).

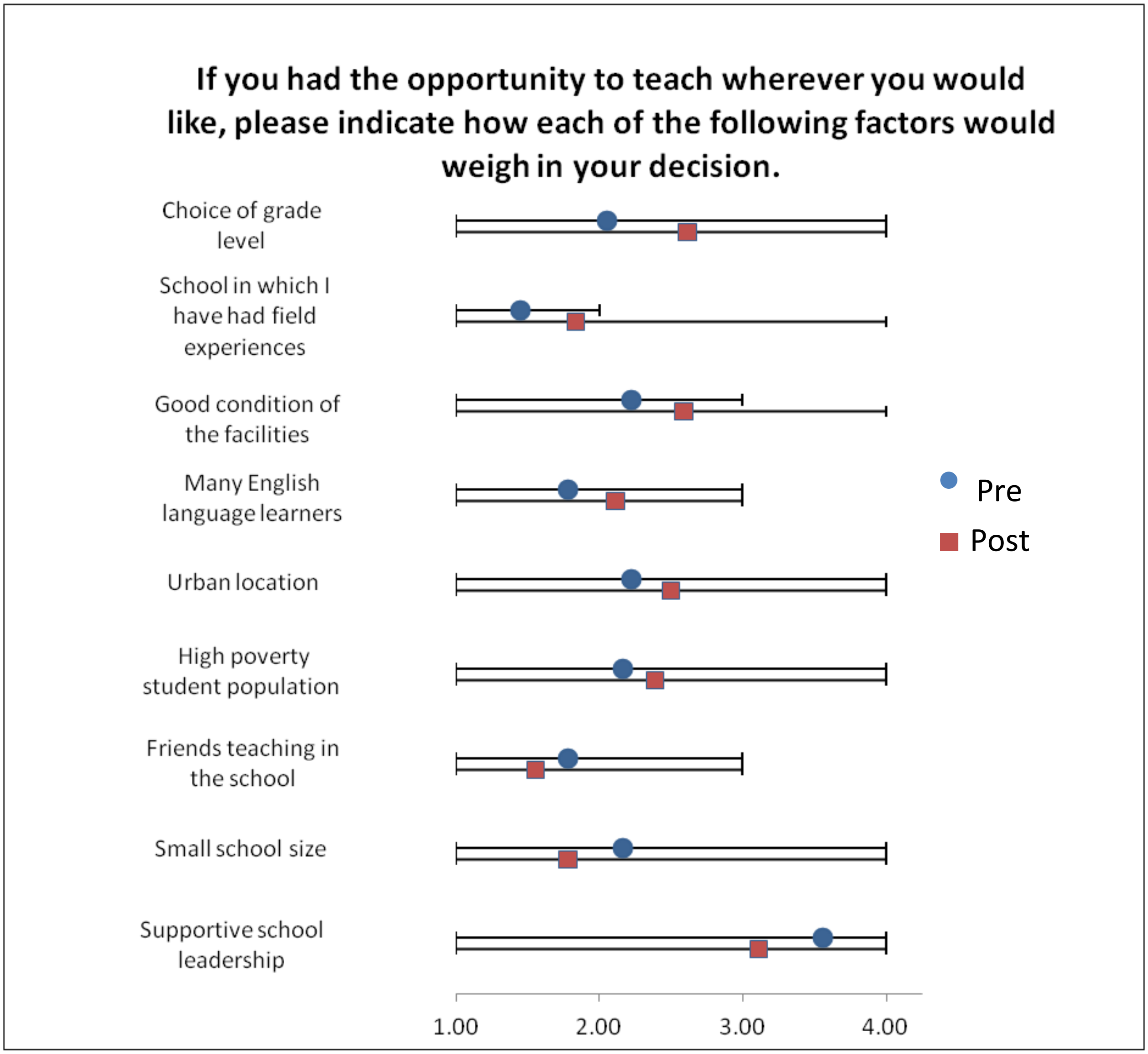

Figure 1. Cohort 1 math and science apprentice teachers' responses regarding factors that would influence where they teach pre (before IMPACT participation) and post (upon completion of the program).

1g) Beliefs in the causes of achievement gaps. Math and science apprentices were provided with a list of reasons that are commonly cited as causes of the achievement gap in K-12 schools (Figure 2). The pre- and post-survey responses showed a slight increase in apprentices beliefs that teacher quality is an important contributing factor to the achievement gap and, on average, there was a slight drop after program participation in apprentices beliefs that student behavior, mobility, and class-size (about one quarter scale point in all cases) are important causes of the achievement gap.

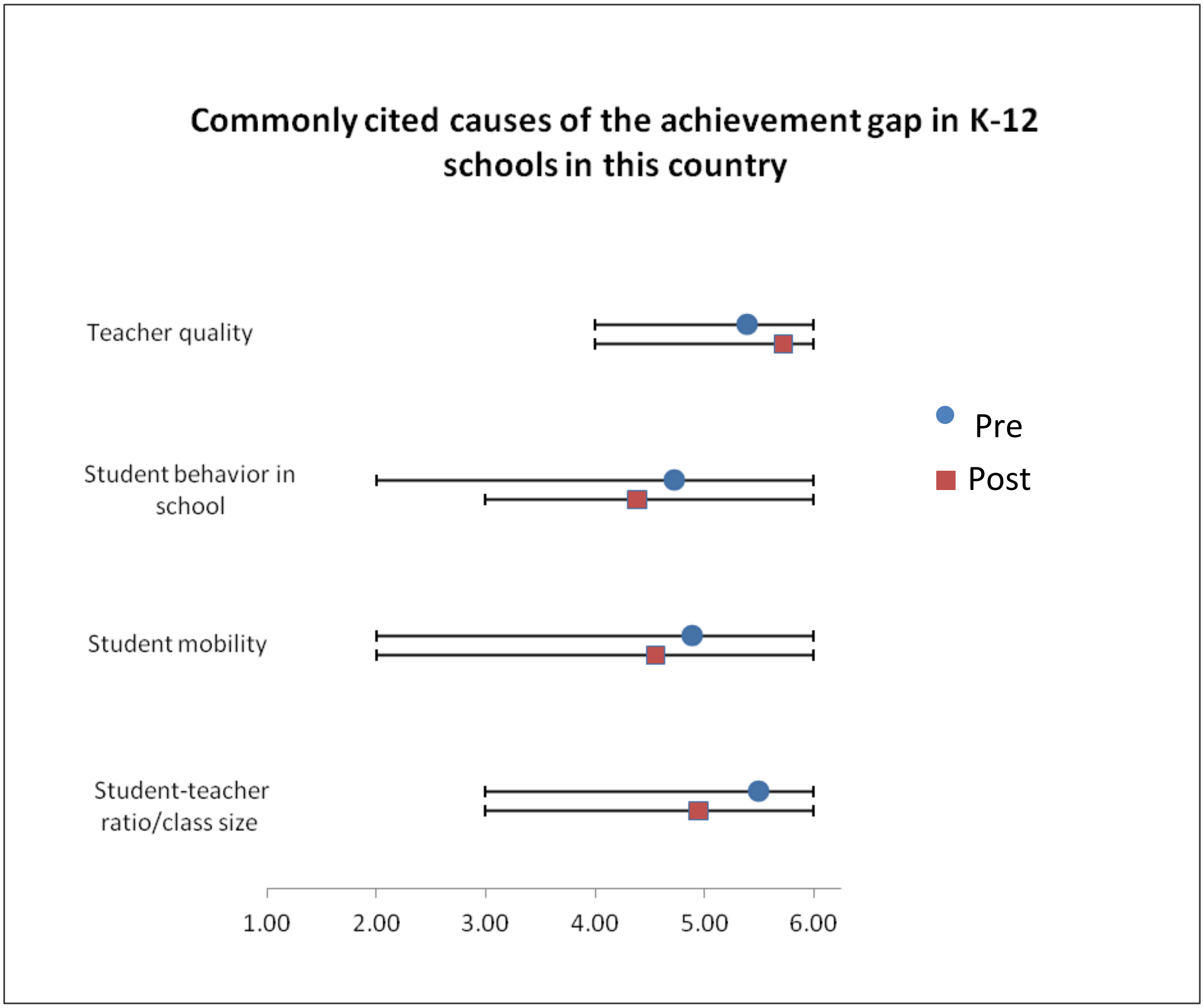

Figure 2. Cohort 1 math and science apprentice teachers' beliefs regarding commonly cited causes of the achievement gap pre (before IMPACT participation) and post (upon completion of the program).

1h) Beliefs about their roles as educators (social justice). Apprentices were asked to endorse a series of statements about their roles as educators. These statements were constructed to help gauge the social justice attitudes of the apprentices. Since social justice attitude is one of the apprentice selection factors, no significant changes were expected. However, the post-surveys showed a slight increase in social justice attitudes compared to the pre-survey, with apprentices more likely to say that teachers have an obligation to teach students to think critically about the government and more likely to think about how to incorporate diverse cultures into classroom instruction (Figure 3). Also, apprentices on the post-survey were slightly less likely to believe that student success is primarily related to how hard they work and less likely to believe that multicultural topics are only relevant in certain subject areas, such as social studies.

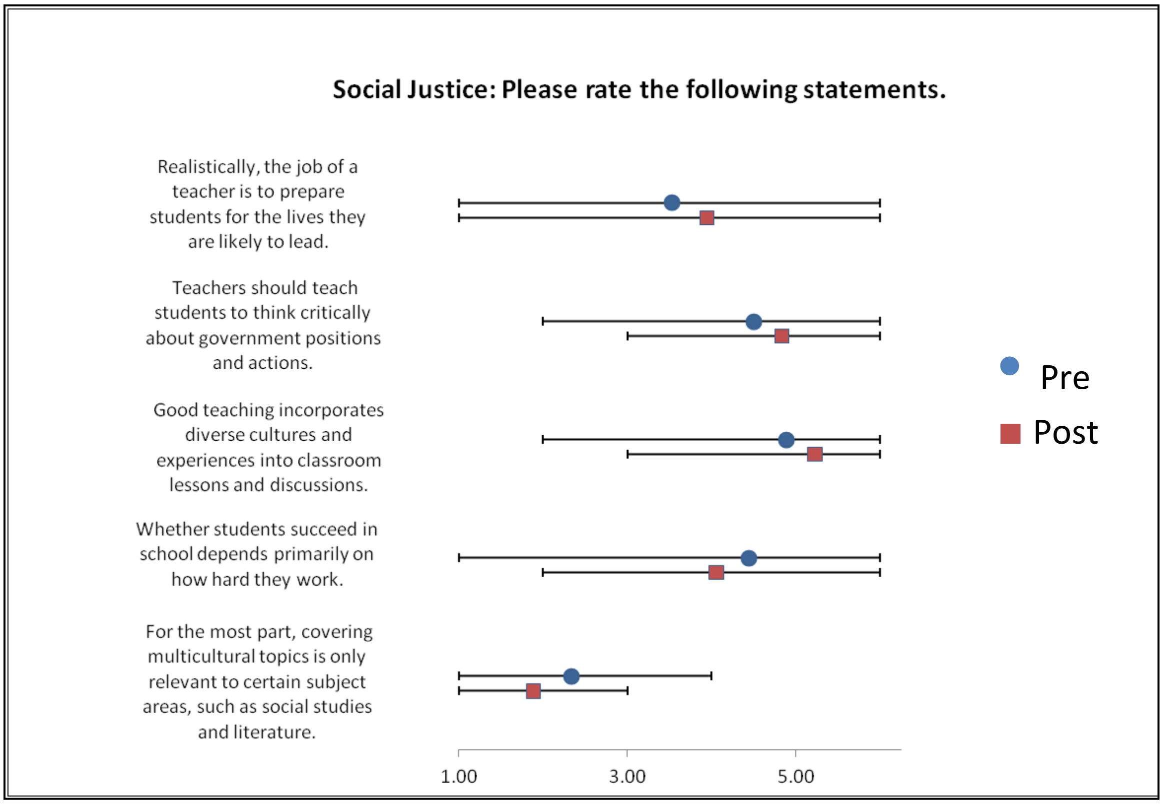

Figure 3. Cohort 1 math and science apprentice teachers' beliefs about social justice pre (before IMPACT participation) and post (upon completion of the program).

Math and science apprentices were asked to describe their career intentions, including how long they planned to work in education and what role they perceived themselves in during the next five years. Over 90% of the respondents saw themselves working in education for at least five years in both the pre- and post-survey. However, although the apprentices maintained their commitment to stay in the field of education, there was a change in what they perceived their future role to be in terms of their career path. On the pre-survey, nearly all apprentices saw themselves as being a classroom teacher in five years (89%, N=16). Post survey results showed much more variation with only 50% (N=10) saying that they would be a classroom teacher in five years. The other 50% either saw themselves in some other role in K-12 environments (6%) or in a role outside of K-12 schools (28%, N=5). This included individuals who wanted to return for a doctorate degree. Nearly 17% (N=3) saw themselves leaving education temporarily.

Summary of results. Our findings overall suggest that the IMPACT program contributed to small increases in apprentices' desire to work in an urban location and to work with high poverty and ELL students after their graduation. The program also appeared to contribute to small increases in apprentices' beliefs that teacher quality was an important factor in the achievement gap as well as their beliefs that teachers have an obligation to teach students to think critically about the government. It also appears to have contributed to a small increase in teachers' interests of incorporating diverse cultures into classroom instruction.

Evaluation Question 2

(a) Do different teachers have different approaches to formative assessment? (b) Are some [formative assessment] specific strategies implemented more than others? (c) Is there variation in the types of assessment across days of instruction? (d) Is there evidence that teachers are differentiating their questioning strategies with respect to their instructional goals?

To help answer evaluation questions 2a-d, all Cohort 1 math and science IMPACT graduates were asked to participate in the collection of instructional logs for two five-day periods in the spring of 2012. In each of these periods, teachers were asked to fill out short (approximately 10 minute) surveys capturing classroom information about a range of instructional practices related to formative assessment. IMPACT graduates were asked to complete the log in reference to one specific class. The first administration took place in early March before statewide standardized testing and the second administration took place in May after testing. It was hypothesized that there may be systematic differences in teacher instructional practices teachers before and after the testing period.

Nine mathematics teachers and nine science teachers participated in the first wave of data collection. Seven of the mathematics teachers and six science teachers completed logs for both waves. We found that IMPACT apprentices participating in the log collection taught a wide range of classes in a variety of instructional environments. Eighty-nine percent of mathematics teachers had their own classrooms. Class sizes ranged from 19 to 36 students with a mean class size of approximately 27 students. Class periods ranged from 55 to 95 minutes and the modal length was 95 minutes. Mathematics teachers also taught students across a wide range of ability levels. One teacher taught an honors level class. The average number of students with Individual Education Programs (IEP) was approximately one with a 0-5 range. The average number of reported English Learners was approximately three, with a range from 0-10.

Every science teacher had his or her own classroom. Class sizes ranged from 13 to 40 students, with a mean class size of approximately 27 students. Class periods ranged from 54 to 100 minutes and the modal lengths were 90 and 100 minutes. As with mathematics teachers, science teachers also taught students across a wide range of ability levels. One science teacher taught an honors level class. The average number of students with Individual Education Programs (IEP) was approximately two and ranged from 0-8. The average number of reported

English Learners was approximately five with a range from 0-9. The log results are organized by the specific evaluation questions.

2a) Evaluation question. Do different teachers have different approaches to formative assessment?

Table 2 shows the variance components from the combined p x I x o generalizability study. The results provide some insight into whether or not teachers had different approaches to formative assessment by looking at the relative magnitude of the p effect in the first row of this table. The p effect summarizes the extent to which teachers differed from each other, on average, in terms of their use of formative assessment practices over time. The variance of the p effect ranges between approximately 0% and 12%, showing that for some domains there were meaningful differences between teachers, but for others these differences were much smaller. In particular, there was substantially more variance between teachers in terms of their feedback strategies than there was in terms of their questioning and instructional strategies.

Table 2

Variance Components (VC) and G-coefficients: Combined p x i x o design

	Feedback				Questioning				Instruction			
	Math		Science		Math		Science		Math		Science	
Effect	VC	Pct	VC	Pct	VC	Pct	VC	Pct	VC	Pct	VC	Pct
p	.014	5.18%	.021	7.99%	0.004	1.54%	.000	.00%	.005	2.89%	.003	1.96%
o	.000	.00%	.000	.07%	0.001	0.32%	.004	1.57%	.000	.00%	.000	.00%
i	.044	16.12%	.053	20.28%	0.045	17.30%	.060	22.75%	.038	22.19%	.031	18.69%
po	.034	12.34%	.017	6.54%	0.021	8.25%	.005	1.75%	.003	1.84%	.005	3.22%
pi	.038	13.76%	.003	1.00%	0.048	18.52%	.059	22.42%	.030	17.19%	.035	21.59%
oi	.000	.00%	.000	.00%	0.000	0.00%	.001	.39%	.005	2.64%	.005	2.93%
poi	.144	52.59%	.168	64.12%	0.140	54.08%	.135	51.12%	.092	53.26%	.085	51.61%
G-coefficient		.536		.808		.243		.000		.706		.617

We used these variance components to estimate g-coefficients for each of the Feedback, Questioning, and Instruction strategies. G-coefficients can be thought of as “reliability-like” coefficients, helping to describe the accuracy of the relative standing of teachers. For the Feedback strategy, the g-coefficient is approximately .81 for science and .54 for math (see Table 2). This suggests that, with six items and 10 occasions, it is possible to distinguish between teachers based on their use of Feedback among science teachers but much less possible among mathematics teachers. For the Questioning strategy, since the p-effect is near zero for science teachers, there is no ability to distinguish among teachers. For math teachers, the g-coefficient is around .21. Thus, overall, there are no meaningful differences between teachers in terms of their questioning strategies. For the Instruction strategy, the g-coefficient is .71 for mathematics teachers and approximately .62 for science teachers. This means that based on the current number of items and occasions, there are small discernible differences among mathematics teachers and less discernible differences among science teachers based on their instructional strategies.

When looking at whether or not teachers, on average, have different approaches to formative assessment, three other variance components from Table 2 can give some additional perspective. These components are the *pi* effect, the *po* effect and the *poi* triple interaction.[3]

The *pi* effect is very large in most cases except for science feedback. This describes the extent to which teachers use specific practices in their classrooms over time. This variance is particularly large for the Instructional and Questioning strategies, accounting for between 17% and 22% of the total variance. Substantively this means that while some teachers may use Write-Pair-Share or a KWL chart most frequently, others report using Whip Around or Gestures most frequently.

The *po* effect describes the extent to which teachers differ based on the total number of strategies that they use across occasions. For example, a teacher who uses all six questioning strategies on the first day of instruction may only use one strategy on the second day and a teacher who used one strategy on the first day may use six on the second day. This variance component is largest for both math and science feedback items where it accounts for approximately 12% and 6% of the variance respectively. This suggests that there is some variation in the number of feedback strategies used by teachers over time, but that the number of Instruction strategies and Questioning strategies tends to be relatively stable. (The teacher that used six strategies on day 1 is also likely to use six strategies on day 2).

2b) Evaluation question. Are some specific [formative assessment] strategies implemented on average more than others?

To determine if specific strategies are implemented on average more than others, two different sources of information were considered. First, it was possible to examine the *i*-effect from the variance decomposition presented in Table 2. The *i* effect describes the extent to which certain practices are used more frequently than others, averaging over teachers and occasions.

[3] The poi effect is confounded with residual variance and it describes error not specifically accounted for in this design.

Using the *i* effect, for example, we found that apprentices asked Initiate-Respond-Evaluate (IRE)- type questions more frequently than Self-Reflection type questions. In all three domains, the item variance component was very large for both math and science teachers, approximately 20% of the variance, suggesting that some practices tend to be used far more than others, averaging over teachers and occasions.

Item means information is presented in Appendix A, Tables A25-A30, showing which instruction strategies are used, on average, most often. Some items had means close to .8 (80%), and others had means closer to 0%. Item means from the Feedback items show that the most commonly used Feedback strategy for both math and science teachers was providing oral feedback to students, either one-on-one or in a whole-class setting. Other more detailed forms of feedback, such as providing written commentary on homework assignments, were used less frequently. The 40% average for providing grades on student output means that on average teachers provided written feedback to students four out of 10 days. The 80% average for providing oral feedback to students in a whole class setting means that the average teacher provided oral feedback on eight out of 10 instructional days. These percentages are largely consistent with expectations for teacher interactions with students.

Item means are largely consistent across occasions for both math and science teachers. There are some differences between math and science teachers in terms of the feedback they provide (Table A25). A typical science teacher provided more written feedback (beyond grades) than a typical math teacher, but provided fewer opportunities for students to assess the work of others. This may be partly due to differences in assignments that are typical across the two disciplines. Science assignments may involve more writing, a lab report for example, that is conducive to written feedback.

Item means for the questioning item domain (Tables A26-A27) show that most of the questioning that was conducted during class periods was of an Initiate-Respond Evaluate (IRE) type. For a typical mathematics teacher, this type of question was asked in just under 9 out of 10 class periods. For a typical science teacher, this type of question was asked in just over 7 out of 10 class periods. For both mathematics and science teachers, questioning types with a range of cognitive demand were used. While less cognitively demanding questioning types (IRE) were used far more frequently than more cognitively demanding types (Mathematical Reasoning, or Evaluate questions), high cognitive demand types (Mathematical reasoning, Evaluate) still appear in a typical classroom on approximately 2.5 out of every 10 instructional days.

Item means for instructional strategies (Tables A27-A29) show that some strategies were used very often. In a typical mathematics classroom, homework was assigned 7 out of 10 instructional days. In a typical science classroom, homework was assigned 6 out of 10 instructional days. Gesturing was also a popular strategy. Other strategies, like role-play and jigsaw, were used very infrequently, if at all. Teachers cited procedural fluency/fact recognition as an instructional goal (Table A30) on approximately 7 out of 10 instructional days. Both math and science teachers devoted time to developing positive dispositions in approximately 3.5 out of 10 instructional days. There was a large difference between a typical mathematics teacher and a typical science teacher in terms of how many days they spent explicitly developing academic language. A typical science teacher spent very little time developing questions or designing experiments (less than one class out of 10 classes).

2c) Evaluation question. Is there variation in the types of assessment across days of instruction?

Variation across days for instruction can be ascertained by inspecting the *o* effect and the *oi* effect presented in the variance decompositions in Table 2. The *o effect* describes the extent to which there was variation in the number practices used across days of instruction. For example, on the first day of instruction, teachers may have used three questioning strategies. On the second day of instruction they may have used only one strategy. The *o effect* is very small (nearly 0) for the Feedback and Instruction strategies. It is slightly larger for the Questioning strategy, but still accounts for only approximately 1% of the variance. This suggests that the number of strategies used, averaged over teachers, is stable over days. Low occasion variance may be related to the fact that different individual teachers were at different points in their lesson trajectory.

Similarly, the *oi effect* describes the extent to which specific practices may be used more frequently on some days and less frequently on others. For example, it may be that a Problem of the Week Protocol is used to launch a concept, and therefore is frequently used early in a unit, while a quiz or test may be more common later in a sequence of instruction. Based on the analysis, the *oi effect* is almost zero in all three scales, which means that the use of specific practices is relatively stable across occasions.[4]

2d) Evaluation question. Is there evidence that teachers are differentiating their questioning strategies with respect to instructional goals?

Cross-tabs (Tables A31-A32) of formative assessment strategies and instructional goals may give more information about how these two dimensions of classroom practice relate to one

[4] This may also be because different teachers were in different instructional points of their units, and so differences in the use of strategies were essentially "washed out.". Had the logs been administered such that all teachers were instructed to begin administration on the first instructional day of a new unit, this may have yielded different results.

another. Ideally, there should be some evidence that teachers are modulating their instruction over time. In other words, different lessons show different amounts of cognitive demand and more cognitively demanding instructional goals are accompanied by activity structures and questioning types that support higher level student thinking.

The cross tabs show the relationship between Questioning type and Instructional goals. The percentages give the marginal occurrences of each questioning type, conditional on the stated instructional goal. For example, of the number of instructional periods (across all teachers) where Procedural Fluency was stated as a goal, 92% were instructional periods where IRE questions were used.

Again, consistent with the item means, Initiate-Respond-Evaluate (IRE) questions occurred with great frequency regardless of lesson goals in both math and science. This is not surprising because "quick" low demand questions are expected frequently during the instruction. However, there is little modulation in the appearance of other questioning strategies, relative to the instructional goals. For example, among math teachers, questions asking students to find patterns (Pattern finding) occur approximately 35% of the time while questions asking students to make connections show up approximately 67% of the time, both regardless of instructional goals.

In science, Understanding, Evaluating, and Analyzing type questions occur with approximately the same frequency regardless of lesson goal. This suggests that teachers are not differentiating their instruction relative to the cognitive demands of instruction. However, during laboratory work, when the instructional goal was designing or conducting investigations to test scientific questions/hypotheses, we observed a noticeable increase in the number of times that teachers indicated they asked Create-type questions or gave Create-type instructional prompts. One of the subtype instructional prompts was "design a lab…", Additionally, there was a noticeable increase in the number of Create questions related to the Mastery of lab skills goal. Thus when lab work is part of the class, questions that are closely related to the specific lab purpose are used more frequently than at other times.

Summary of results. Collectively, the log evidence suggests two main findings:

- Teachers had a consistent set of goals for instruction and these goals did not change over the course of instruction. Predominantly, these goals were for procedural and conceptual understanding as well as for developing academic language.
- Teachers used a consistent set of tools and strategies in the classroom both before and after state testing. While different teachers used different tools, they did not change the tools they used during the course of instruction. There was some evidence that teachers had a tendency to stick to "familiar" components of classroom instruction (taking notes, providing oral feedback, assigning homework) and shied away from more complex activity structures (such as a jigsaw) and cognitively demanding questions. While the

number of occurrences for low demand strategies was not necessarily surprising, there was little evidence, either from the cross-tabs or from the variance components analysis, that teachers varied their instructional practices systematically to align with their goals.

Evaluation Question 3

Does participation in the IMPACT program lead to increases in apprentice's pedagogical content knowledge as measured by (a) the Mathematical Knowledge for Teaching (MKT) for math and (b) the Assessing Teacher Learning About Science Teaching (ATLAST) for science? We analyze and report results from each measure individually.

3a) Increases in mathematics pedagogical content knowledge. Hill, Rowan, and Ball (2005) define Mathematical Knowledge for Teaching (MKT) as, "explaining terms and concepts to students, interpreting students' statements and solutions, judging and correcting textbook treatments of particular topics, using representations accurately in the classroom and providing students with examples of mathematical concepts and proofs" (p. 373). Multiple-choice measures for mathematical knowledge for teaching were found by Hill, Schilling and Ball (2004) to, "reliably discriminate among teachers and meet basic validity requirements for measuring teachers' mathematical knowledge for teaching" (as cited in Hill et al., 2005). Additionally, teachers' scores have been found to predict increases in student achievement. Consequently, we administered the MKT to IMPACT apprentices as a pre- and post measure. Our hypothesis was that by participating in a subject-specific internship program, teacher apprentices would increase their mathematical knowledge for teaching.

Instrument development. MKT items have been developed for a variety of content domains over the K-8 mathematics curriculum. Because it was anticipated that IMPACT apprentices would be teaching upper-middle school and high school grade levels, items were selected from five of these domains for inclusion on a pre- and post-test administered to teacher apprentices at program intake and again at the conclusion of the apprentices 18-month formal training program.

Table 3 reports the five domains included on the IMPACT MKT tests, along with the number of items measuring each domain. The largest set of items (19) measured apprentices' skills in the Patterns, Functions and Algebra domain, anticipating that most of the IMPACT apprentices would be teaching algebra either during their apprenticeship or as classroom teachers after program completion.

Table 3

Test Composition (scored items)

Domain	Pre-test	Post-test
Geometry	4	4
Data, Probability and Statistics	6	11
Patterns, Functions and Algebra.	20	19
Proportional Reasoning	1	0
Rational Numbers	6	11
Total	37	45

Test items were obtained for each domain from the University of Michigan, along with Item Response Theory (IRT) parameters for item difficulty and item discrimination. Based on this information, along with expert content analysis offered by the IMPACT mathematics program director, a set of 45 items was selected for the pre-test MKT. The items were selected both because of their measurement properties (items over a range of difficulties that were sufficiently discriminatory) and because of their alignment with the type of mathematics that apprentices were likely to encounter during their training program. In IRT, item difficulty can be interpreted as representing the ability level required in order to have a 50/50 chance of answering an item correctly. In other words, a difficulty level of zero means that an individual of average ability has a 50% chance of giving a correct response. A difficulty level of two means an individual with an ability of two standard deviations above the mean has a 50% change of answering correctly (a very difficult item).

Table 4

IRT Parameters, Pre-Test Items

Item	Domain	Discrimination	Difficulty	Item Mean
1	Geometry	0.743	0.383	0.42
2	Geometry	0.704	-0.232	0.5
3	Geometry	0.738	-0.39	0.83
4	Geometry	0.854	-0.424	0.17
5a	Data, Probability and Statistics	1.116	-0.789	1
5b	Data, Probability and Statistics	1.114	-0.684	1
5c	Data, Probability and Statistics	0.445	2.3	0.58
5d	Data, Probability and Statistics	0.867	0.024	0.92
5e	Data, Probability and Statistics	0.844	0.131	0.83
6a	Data, Probability and Statistics	0.833	0.109	0.75
6b	Data, Probability and Statistics	1.831	-0.077	1
6c	Data, Probability and Statistics	1.711	-0.05	0.83
6d	Data, Probability and Statistics	1.231	-0.02	0.58
6e	Data, Probability and Statistics	0.982	0.005	0.83
6f	Data, Probability and Statistics	0.708	0.628	0.67
7a	Data, Probability and Statistics	0.634	0.356	0.58
7b	Data, Probability and Statistics	0.634	0.356	0.42
8	Data, Probability and Statistics	0.989	-0.05	0.58
9	Patterns, functions and Algebra.	1.105	0.015	0.67
10	Patterns, functions and Algebra.	0.946	0.467	0.5
11	Patterns, functions and Algebra.	0.565	0.453	0.75
12	Patterns, functions and Algebra.	0.347	1.864	0.58
13	Patterns, functions and Algebra.	0.74	0.669	0.5
14a	Patterns, functions and Algebra.	0.653	1.003	0.42
14b	Patterns, functions and Algebra.	1.063	-0.926	0.92
15	Patterns, functions and Algebra.	0.582	0.906	0.42
16a	Patterns, functions and Algebra.	0.777	0.409	0.42
16b	Patterns, functions and Algebra.	1.102	-0.678	0.75

Item	Domain	Discrimination	Difficulty	Item Mean
16c	Patterns, functions and Algebra.	0.652	-0.293	0.58
16d	Patterns, functions and Algebra.	0.598	0.974	0.17
17a	Patterns, functions and Algebra.	0.845	-0.492	0.83
17b	Patterns, functions and Algebra.	0.843	-0.318	0.75
18	Patterns, functions and Algebra.	0.331	2.039	0.17
19	Patterns, functions and Algebra.	0.949	-0.573	0.92
20	Patterns, functions and Algebra.	0.51	2.2	0.17
21	Patterns, functions and Algebra.	0.946	0.467	0.58
22a	Patterns, functions and Algebra.	0.751	0.507	0.25
22b	Patterns, functions and Algebra.	0.872	-0.222	0.42
23	Proportional Reasoning	0.804	1.76	0.58
24	Rational Numbers	1.05	0.035	0.67
25a	Rational Numbers	0.993	-0.571	0.83
25b	Rational Numbers	0.78	-0.717	0.75
26a	Rational Numbers	1.496	0.282	0.75
26b	Rational Numbers	0.839	-0.296	0.83
27	Rational Numbers	0.538	0.531	0.25

The IMPACT MKT assessment was administered to 41 individuals from three training programs (IMPACT, Joint mathematics/Education Program and Teacher Education Program) for calibration and item piloting purposes. All 45 items were scored. In addition to the IRT parameters provided by the test developers, we estimated an item mean for each item. This item mean can be interpreted as item difficulty (Crocker & Algina, 1986) in a Classical Test Theory (CTT) framework. Items with higher means have lower difficulty because a higher proportion of individuals answered these items correctly. An item with a mean of one is an item that every individual answered correctly while an item with a mean of zero is an item that no one answered correctly.

Item means from this pilot sample were correlated with the IRT difficulty parameters provided by the MKT developers. That correlation was significant and strong—roughly -.59. A scatterplot showing item means plotted against IRT difficulty parameters is shown in Figure 4.

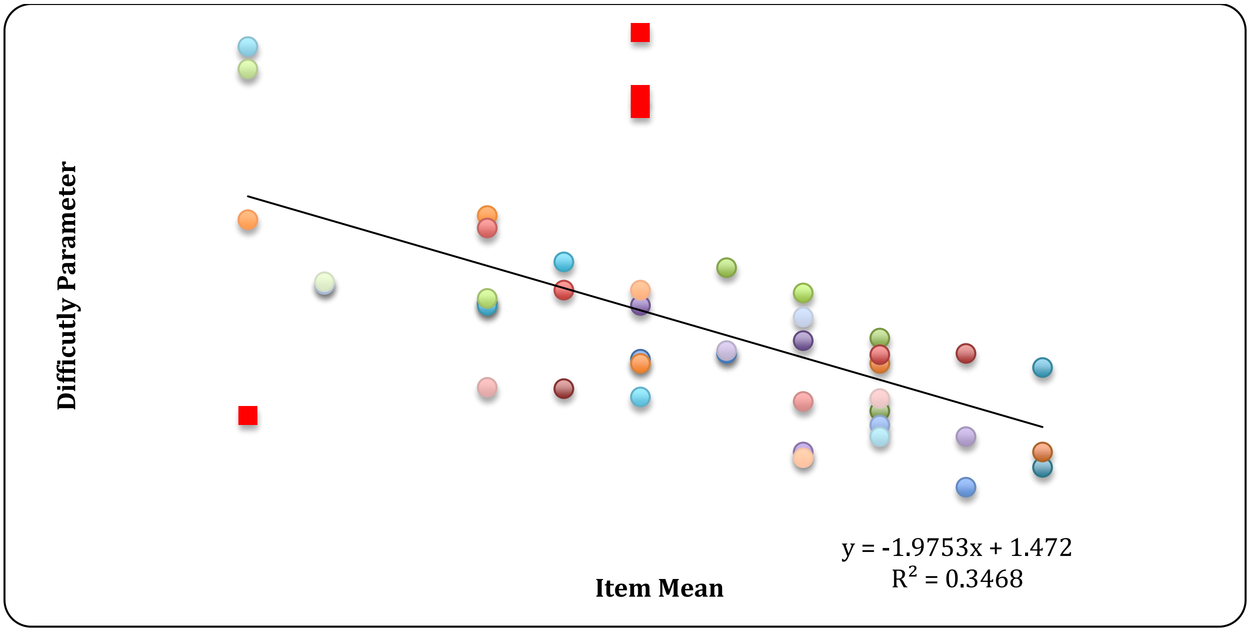

Figure 4. Scatterplot showing relationship between CTT item means and IRT difficulty parameters

This scatterplot shows CTT item means along the x axis, and IRT location parameters along the y axis. As expected, when the location parameter becomes smaller, the item mean becomes higher. A trendline is plotted to show the overall relationship between item mean and item difficulty. A negative difficulty parameter corresponds to higher difficulty and therefore a lower item mean. An item mean of .5 should theoretically correspond to a location parameter of zero. Four items are marked with (red) squares on this scatterplot because they show a slight lack of alignment between the item mean and the item difficulty. Three of these items were easier than anticipated for the pilot group and one of these items was more difficult than anticipated. Ultimately, these items were retained for scoring but were excluded from the post-test form. Removing those items resulted in a -.78 correlation between item location and item mean.

In addition to these four items, there was a small subset of items that were very easy for IMPACT apprentices. This subset consisted of 8 items that had means above .9. Consequently, these items were excluded from scoring on the IMPACT MKT pre- and post-test.

Of the 45 items on the pilot test, 37 items were scored, and 33 were candidates for inclusion on the post-test. Twenty-two of the 33 items were eventually chosen for the post-test. Additional items were selected in order to preserve the content domain representation and item difficulty to the fullest extent possible. These items were selected by considering both their measurement properties and their substantive relevance. Decisions about which items to add to the post-test were made in collaboration with IMPACT program directors. The number of items representing Geometry as well as Patterns, Functions, and Algebra remained essentially the same across test forms. Several additional items in the Data, Probability and Statistics domain as well as the Rational Numbers domain were added to the post-test, giving those domains slightly higher representation.

Results. Table 5 shows the mean item difficulties by domain for each pre- and post-test. Overall, item difficulties are comparable across forms. With item parameters only available for one item in the Proportional Reasoning domain, this item was not included in the post-test.

Table 5

Pre/Post Difficulty by Domains: Scored Items

Test	Geometry	Data, Probability and Statistics	Patterns, functions and Algebra	Proportional Reasoning	Rational Numbers	Overall
Pre-test	-0.166	-0.148	0.424	1.760	-0.123	0.142
Post-test	-0.181	0.012	0.370	NA	-0.209	0.215

The following analysis is based on results from 10 IMPACT mathematics teachers who taught a wide variety of mathematics content at various grade levels. The majority of teachers taught upper high school while only one teacher taught middle school mathematics. Three of the nine high school teachers taught only grades 10, 11 and 12. In total, 70% of the teachers taught at least one Algebra II class and 50% taught Algebra I. A much smaller proportion of teachers (20%) taught Geometry.

IMPACT apprentices showed the greatest overall score gains in the Geometry subdomain. However, this was based on only four items. The other three domains (Table 6) show either very small, statistically non-significant increases, or even a small decrease in the case of Rational Numbers. Overall scores on the post-test were essentially the same as the scores on the pre-test.

Table 6

MKT Scores

		Geometry	Data, Probability and Statistics	Patterns, functions and Algebra	Rational Numbers	Overall
Pre	Mean	47.50%	60.00%	56.50%	70.00%	61.35%
		(24.86%)	(14.05%)	(19.73%)	(24.6%)	(16.9%)
Post	Mean	77.50%	60.91%	61.05%	64.00%	63.18%
		(27.51%)	(23.49%)	(20.34%)	(12.65%)	(14.76%)

There are several possible hypotheses about these findings. The first may be related to our small sample size, which compromises statistical power. The test developers indicate that the MKT has sufficient power to detect moderate effects in groups of around 60.[5] In our case, the instrument was administered to a group of 10. Additionally, the test content might not be well-aligned to the program aims. Both of these issues compromise the validity of these findings. Subsequent evaluations will explore whether or not these findings are replicable.

Summary of results. Teacher apprentices scored similarly at the beginning and the end of the program. As measured by the IMPACT MKT, there was no significant change in apprentices' content knowledge. Despite teaching a wide variety of mathematics content and different grade levels, IMPACT apprentices showed a large score gain in the Geometry subdomain (based on only 4 items).

(b) Increases in science pedagogical content knowledge. Developed by Horizon Research, *Assessing Teacher Learning About Science Teaching* (ATLAST) uses a series of instruments to measure knowledge and skills for teaching at the middle school level. The instruments measure teachers' (a) knowledge of science content, (b) ability to use science content to analyze student thinking, and (c) ability to use science content to make instructional decisions (Smith & Banilower, 2006). Two instruments, *Force and Motion* plus *Flow of Matter and Energy in Living Systems*, were used for this study. Each instrument consisted of 29 multiple choice items. The same test form was administered for both pre- and post-test administrations.

The two assessments have shown strong test-retest reliability, .93 for the *Flow of Matter and Energy in Living Systems* assessment, and .88 for the *Force and Motion* assessment. These reliabilities were based on a pilot sample of 100 middle grades science teachers.[6]

[5] http://sitemaker.umich.edu/lmt/files/LMT_summary_tech_info.pdf
[6] http://www.horizon-research.com/atlast/?page_id=78

Results. The science teacher apprentices had a wide variety of science backgrounds and taught a variety of classes. Notably, only two taught middle school and only two taught physics. This may be important to consider since the content domains of the ATLAST instruments are very specific and one of the assessments, *Force and Motion*, is strongly connected to physics.

Group means from our nine science apprentices showed that performance was essentially unmoved over the course of the 18-month IMPACT apprenticeship (Table 7). Overall, apprentices performed slightly better on the *Force and Motion* assessment than the *Flow of Matter and Energy in Living Systems* assessment and there was some slight evidence of improvement between pre- and post-tests on the *Flow of Matter and Energy in Living Systems* assessment.

Table 7

ATLAST Scores

		Flow of Energy and Matter	Force and Motion
Post	Mean	14.78	16.78
	SD	3.27	5.17
Pre	Mean	13.11	16.33
	SD	2.76	2.69

There are several possible hypotheses about the lack of significant differences on the science tests, paralleling those discussed for the MKT. These include our small sample size, the relatively short time (18 months) for teachers to improve their science knowledge, and the short length (3-months) of full-time teaching experience during the program. Finally, as with our mathematics findings, the IMPACT apprenticeship program does not necessarily align with the science knowledge measured by ATLAST.

Summary of results. As measured by ATLAST, science teacher apprentices scored similarly at the beginning and the end of the program. The apprentices performed slightly better on the *Force and Motion* assessment compared to the *Flow of Matter and Energy in Living Systems* assessment and apprentices showed minor, but non-significant improvement between the pre- and post-test of the *Flow of Matter and Energy in Living Systems* assessment.

Evaluation Question 4

What is the quality of teacher apprentices as measured by (a) the Performance Assessment for California Teachers and (b) the Instructional Quality Assessment?

4a) Quality of teacher apprentices - Performance assessment for California teachers. The Performance Assessment for California Teachers (PACT) is a teacher performance assessment developed by a consortium of teacher preparation programs in California. The consortium consists of 30 universities, one district internship program, and one charter school network.[7] PACT was developed in response to a California state mandate that required teacher certification programs to use performance assessments as one component in making credentialing decisions for new teachers (Pecheone & Chung, 2006). PACT assesses preparedness for teaching using multiple sources of evidence including lesson plans, artifacts, video, and reflections on practice. All teacher candidates are scored on four dimensions: Planning, Instruction, Assessment, and Reflection. Candidates are also scored on their support for the development of Academic Language. Scores range from 1 to 4, with 1 indicating failure to meet the passing standard of performance. A score of 2 represents the minimum score for passing, and a score of 4 represents exceptional performance (Pecheone & Chung, 2006). Every IMPACT apprentice takes the PACT at the end of his or her first year of training.

Results. Table 8 presents IMPACT apprentices mean scores for each of five scoring categories, planning, instruction, assessment, reflection, and support for development of academic language support (Academic Language). Results are across six teacher groups, IMPACT Science, IMPACT Math, IMPACT All – combining science and math teachers, TEP, California Pilot Math, and California Pilot Science. Mean scores are reported to align with the rubrics. For example, a mean score of 3.5 indicates that a teacher (or a cohort of teachers) is between *exceeding standards* and *exceptional performance*.

[7] http://www.pacttpa.org/

Table 8

PACT Scores from IMPACT Apprentices, TEP teachers, and California Pilot Group

Group	Planning *(SD)*	Instruction *(SD)*	Assessment *(SD)*	Reflection *(SD)*	Academic Language *(SD)*	Overall *(SD)*	*N*
IMPACT Science	2.96	2.81	2.70	2.58	2.61	2.75	9
	(0.200)	(0.300)	(0.351)	(0.433)	(0.546)	(0.204)	
IMPACT Math	2.80	2.43	2.58	2.33	1.98	2.47	10
	(0.740)	(0.800)	(0.562)	(0.472)	(0.299)	(0.426)	
IMPACT All	2.88	2.61	2.64	2.45	2.28	2.60	19
	(0.547)	(0.631)	(0.466)	(0.461)	(0.533)	(0.361)	
TEP	2.61	2.39	2.42	2.46	2.19	2.43	20
	(0.658)	(0.73)	(0.517)	(0.635)	(0.667)	(0.529)	
CA Pilot Science	2.83	2.67	2.53	2.53	2.49	2.67	72
	(0.650)	(0.653)	(0.684)	(0.611)	(0.650)	(0.569)	
CA Pilot Math	2.52	2.34	2.3	2.27	1.84	2.35	50
	(0.719)	(0.679)	(0.672)	(0.497)	(0.710)	(0.554)	

Consistent with practices in the PACT technical report (Pecheone & Chung, 2007), double scores were averaged for tasks that were scored more than once. IMPACT teacher apprentices (2010-2011 cohort) had the highest performance in the Planning category and the lowest performance in Academic Language. This pattern of scores is consistent with the trends found by Pecheone and Chung (2007) in a pilot sample of California teachers (CA Pilot Math and CA Pilot Science) included for reference in Table 8. Pecheone and Chung found that California candidates tended to perform relatively better on Planning and Instruction tasks than on Assessment and Reflection. Pecheone and Chung also found that the lowest performance for California teachers was in the category of support for the development of academic language (Academic Language).

While the small sample sizes preclude making any substantive claims about relative performance of IMPACT apprentices relative to other groups, IMPACT apprentice teachers tended to have slightly higher scores than math and science apprentices that participated in the UCLA TEP program. In the Planning and Instruction tasks, this gap was nearly a quarter of a scale point. There were also differences within the IMPACT program participants. Science apprentices generally scored higher than math apprentices, with overall scores nearly a quarter point higher. In particular, the group mean on the Planning task for science apprentices (2.96) nearly reached the *exceeding standards* level. Math apprentices struggled with Academic Language; the group mean for this task was 1.98; slightly below meeting the standard and one-half scale point lower than the group-mean for Science apprentices (2.61).

Summary of results. Given our small sample sizes, we cautiously note that the overall IMPACT teacher apprentice performances tended to have slightly higher scores than the traditional math and science apprentices who participated in the UCLA TEP program. In the Planning and Instruction tasks, this gap was nearly a quarter of a scale point.

There were differences within the IMPACT program, as well. In general, Science apprentices performed slightly better than Math apprentices, with overall scores nearly a quarter of a scale point higher. In particular, the group mean on the Planning task for science apprentices was nearly at the level of *exceeding standards*. Math apprentices struggled with Academic Language.

4b) Quality of teacher apprentices: Instructional Quality Assessment. This section describes a novel adaptation of the Instructional Quality Assessment (IQA), developed to evaluate graduates of the IMPACT program. The IMPACT program centers on three principles: to integrate course and fieldwork to prepare integrative professionals; to support teachers in acquiring the skills to provide students a rigorous standards-based education; and to arm new teachers with the capacity, commitment, capacity, and resilience to promote social justice, caring and anti-racism in urban schools (IMPACT: Urban Teacher Residency Program "Inspiring Minds through Professional Alliance of Community Teachers" Program Handbook, 2012-2013, p. 3). Focusing on the role of IMPACT teacher apprentices as social justice educators, researchers developed an *Equitable Teaching* rubric as a supplement to the existing IQA tool.

Instrument development. The first two rubrics of the IQA, *Academic Rigor* and *Clear Communications* have been previously found to be reliable and valid with experienced teachers in various subjects and grade levels ranging from later elementary to high school students (Matsumura et al., 2006; Silk, Silver, Amerian, Nishimura, & Boscardin, 2009). *Academic Rigor* includes four dimensions: (1) Potential of the Task (2) Implementation of the task (3) Rigor in Student's Responses and (4) Academic Rigor in Teacher's Expectations. *Clear Communications* includes two dimensions (1) Clarity and Detail of Expectations and (2) Communications of Expectations.

Researchers made initial refinements of the IQA based on research related to four Equitable Teaching dimensions: (1) Participation Structures (2) Differentiation (3) Academic Language and (4) Relevance. This third rubric was pilot tested in spring 2011. Based on this initial pilot, final revisions were made to the IQA rubric to better align with program expectations and increase reliability. See Appendix C for the complete IQA rubric.

Teacher participants. Of the 20 IMPACT graduates, 15 returned the Teachers Assignment Checklist (TAC) and accompanying assignments, including seven science and eight math submissions. Science topics included Chemistry, Physics, Biology, Life Sciences, and a non-standards-based elective. Math assignments spanned remedial Algebra, Algebra, Algebra II, and Trigonometry.

Materials and procedure. Materials were distributed to the teacher apprentice graduates, subsequently referred to as teachers, approximately one month before returning them to researchers, giving participants time to gather relevant materials. Teachers had the opportunity to complete the TAC in a regular program meeting. However, most participants, about 67%, chose to submit the document later electronically which allowed them extra time to complete the instruments.

Distributed instruments included the Teacher Assignment Checklist (TAC) and a copy of the complete IQA rubric for reference. To supplement the completed TAC, teachers submitted one sample each of low, medium, and high quality of student work on a classroom assignment, in addition to any assignment instructions, handouts, or other materials.

The TAC provided an opportunity for teachers to provide explanation and context for their selected assignment. Teachers filled in open-ended questions regarding targeted content standards, assignment description, reading materials, learning routines, metacognitive practices, instructional practices, as well as monitoring and assessment. The TAC also gave teachers the opportunity to self-report other relevant characteristics of their students, including percentages of high and low performers, struggling readers, English Language Learners (ELLs), and perceived assignment challenges.

Analysis of the TAC. Four CRESST staff members served as raters to score the teacher assignments. Three of these raters were involved with the pilot of Equitable Teaching in spring 2011. Initially, raters began by reviewing the rubric and discussing the criteria for each dimension. Raters considered all aspects of the submitted materials, including the TAC, student work, and teacher comments. That is, although certain portions of the TAC directly related to specific dimensions, researchers were not confined to certain questions to find evidence of effective teaching practices.

Based on previous findings on the IQA for mathematics and science (Silk, Silver, Amerian, Nishimura, & Boscardin, 2009), the CRESST in collaboration with Center X established a score of three as a satisfactory performance on individual dimensions. Because of reliability concerns and to ensure consistent scoring, all raters scored each submission. Once scoring was underway, raters discussed disagreements and then came to consensus on a final score. Two estimates of reliability were calculated. Within one-point reliability was calculated across each dimension. If any one of the four raters disagreed by more than one point, this was counted as a disagreement. In addition, a two-way random model with measures for consistency was run to estimate intra-class correlation (Nichols, 1998).

Overall, levels of internal consistency were acceptable to good with an overall within 1-point reliability of 81% and intra-class correlation of .683. Table 9 shows the reliability of the raters' scores prior to reaching a consensus score. It should be noted that where there was a large discrepancy (high agreement with a lower intra-class correlation coefficient [ICC]), raters were generally divided between two similar scores (i.e., 2 and 3, rather than 1 and 4).

Table 9

Inter-rater reliability (*N* = 15)

Dimension	Within 1-point	ICC
1-1 Rigor-Potential of the task	100%	.554
1-2 Rigor-Implementation	66.6%	.585
1-3 Rigor-Students' Responses	100%	.727
1-4 Rigor-Teacher's Expectations	86.7%	.567
2-1 Expectations-Clarity and detail	86.7%	.808
2-2 Expectations-Communication	60.0%	.818
3-1 Equitable Teaching-Participation structures	100%	.505
3-2 Equitable Teaching-Differentiation	66.6%	.563
3-3 Equitable Teaching-Academic language	80.0%	.865
3-4 Equitable Teaching-Relevance	60.0%	.844
Mean Overall	81.0%	.683

Results on classroom context. To describe classroom context, quantitative portions of the TAC were entered into SPSS to run descriptive statistics. The assignments were largely project-based assignments (mean = 6.33 number of similar assignments) and completed in class (mean = 77.67), while working in groups (mean = 54%). Only one assignment (math) was reported to be completely individual work. See Table 10 for an additional description.

Table 10

Characteristics of the assignment (*N* = 15)

Assignment	Minimum	Maximum	Mean	Std. Deviation
Minutes spent (*n* = 11)[a]	20	600	223.64	238.75
Similar each year	2	30	6.33	7.108
Percent completed in class	15	100	77.67	27.518
Percent completed at home	0	85	17.73	24.700
Percent completed individually	0	100	31.60	33.217
Percent completed in pairs	0	90	22.67	31.332
Percent completed in groups	0	100	54.00	40.054

[a]Not all teachers reported this number.

Table 11 shows that there were a significant minority of ELLs (23%) and struggling readers in teachers' classrooms (about 26%). Teachers described their students as mostly medium performers (about 42%) with fewer low (about 36%) and high performers (about 21%). Although teachers described their students as good/excellent (about 40%) or adequate on their submitted assignments (about 34%), they also reported that a large minority (about 32%) did not perform adequately.

Table 11

Characteristics of participants' students (*N* = 15, except where noted)

Student characteristics	Minimum	Maximum	Mean	Std. Deviation
ELLs	0	80	23.00	26.003
Struggling readers	0	90	25.53	28.005
Generally high performing students	5	60	21.13	15.675
Generally medium performing students	20	80	41.93	17.454
Generally low performing students	15	60	35.60	14.652
Good/Excellent on this assignment (*n* = 14) [a]	0	80	40.07	26.534
Adequate on this assignment (*n* = 14) [a]	0	70	33.86	20.214
Not yet adequate on this assignment (*n* = 14) [ab]	10	90	32.36	26.792

[a]Not all teachers reported this number. [b]One teacher noted this included students who had not turned in assignments.

Results on IQA scores. The overall mode for all scores across all dimensions was "3", given on 58% of all ratings (Figure 5). Scores of "4" were awarded on 24% of ratings, scores of "2" were awarded on 12.7% of ratings, and scores of "1" were awarded on 5.3% of ratings.

Potential of the Task (1-1), *Clarity and Detail of Expectations* (2-1*)*, and *Communications of Expectations* (2-2) had a mode of 4. Although the modes were the same, there was greater variability for the latter two dimensions; both components of *Clear Expectations*. *Relevance* had the lowest overall mode of 2. See Figure 5 for the mean and mode for each rubric. For a detailed explanation of dimensions and rubrics, please consult the IQA Rubric in Appendix C.

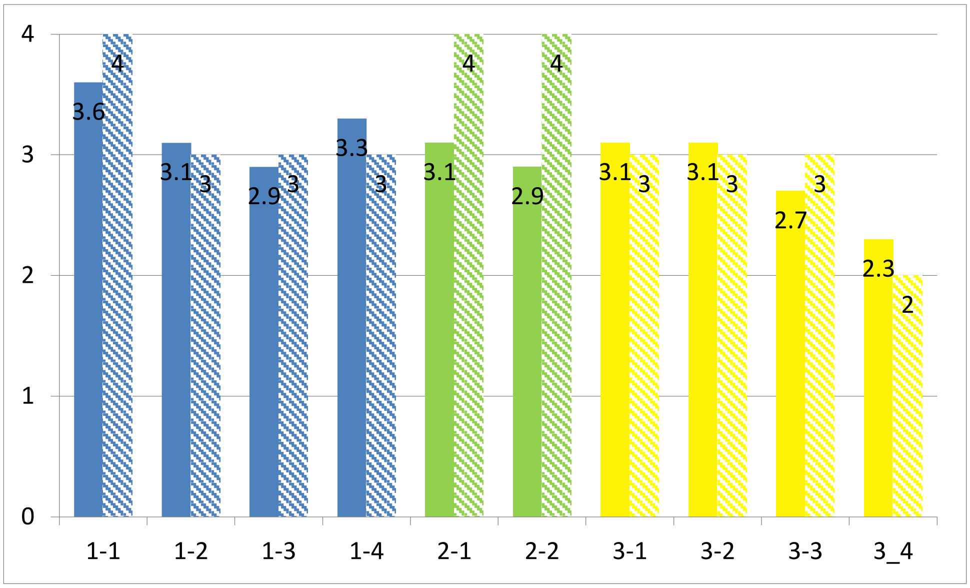

Figure 5. Mean (solid) and modes (stripes) scores for all apprentices (*n* =15).

Results by teaching subject. In terms of mode, math and science scores were similar; math and science teachers had the same mode on six of 10 dimensions (Table 12). Taking mean and mode into account, both math and science teachers scored highest on *Rigor - Potential of the task* and lowest for *Equitable Teaching - Relevance.* Although the collected assignments offered the potential for complex thinking and the opportunity to solve challenging problems, teachers often had difficulty integrating connections to real-world problems or experiential knowledge into their assignments.

In terms of means, overall, math and science scores were similar, with means of 2.9 and 3.1, respectively. While science and math teachers did not significantly differ for their overall mean scores, there were interesting differences observed at the dimension level. The modes differed for math and science teachers for *(1) Rigor-Potential of the task*, (2) *Expectations-Clarity,* (3) *Expectations-Communication,* (4) *Equitable Teaching-Relevance* (Table 12).

On closer inspection, science teachers were more consistent than their math counterparts. All modes for science were “3” and eight of ten science means were “3” or above. While math teachers had a higher mode “4” for *Potential of the task*, they were less successful for *Clarity*, *Communication,* and *Relevance.* Three math teachers scored a "1" on *Expectations-Communication*, indicating that their assignments did not effectively communicate to students about how they would be evaluated.

Table 12
Mean, modes, and range for math (n = 8) and science (n = 7)

Dimension	Math (n = 8)		Science (n = 7)	
	Mode	Mean	Mode	Mean
1-1 Rigor-Potential of the task	4	3.8	3	3.4
1-2 Rigor-Implementation	3	3.0	3	3.3
1-3 Rigor-Students' Responses	3	2.9	3	3.0
1-4 Rigor-Teacher's Expectations	3	3.4	3	3.1
2-1 Expectations-Clarity and detail	2	2.8	3	3.4
2-2 Expectations-Communication	1; 4	2.5	3	3.4
3-1 Equitable Teaching-Participation structures	3	3.1	3	3.0
3-2 Equitable Teaching-Differentiation	3	3.0	3	3.3
3-3 Equitable Teaching-Academic language	3	2.5	3	2.9
3-4 Equitable Teaching-Relevance	1; 2	2.0	3	2.6
All teachers (N = 15)		2.9		3.1

Summary of results. Across all dimensions and both topics, the mode for the submitted TACs was 3. Teachers scored well, especially in the *Rigo*r rubric, with the highest scores for *Potential of the task*. The submitted assignments offered the potential for complex thinking and the opportunity to solve challenging problems. Both science and math teachers scored well on *Participation structures*, which is bolstered by their self-reports of students having ample time to work individually, in pairs and in groups. However teachers still have room for growth in the *Equitable teaching* rubric. Math teachers could additionally improve in *Communication.*

Taking mean and mode into account, both math and science teachers scored the lowest for *Relevance*. That is, teachers had more difficulty integrating connections to real-world problems or experiential knowledge into assignments. In addition, despite having classrooms with significant numbers of ELLs and struggling readers, teachers scored lower in *Academic language* (see Table 12). Raters found that teachers had ample room to increase support for developing academic language in their classrooms.

In terms of mode, math and science scores were similar; science and math teachers had the same mode for six of 10 dimensions. However, there was greater variability in scores for math teachers than for science teachers. *Expectations-Communication* and *Equitable Teaching-Relevance*, which were bi-modal, had lower means. This suggests that teachers were generally skilled in preparing lessons but not always successful in connecting it to real-world or student experiences and that students may have been unsure about teacher expectations. In particular, the bimodal nature of math scores in *Expectations-Communication* and *Equitable Teaching-Relevance* suggests that some teachers were very proficent or poor in these areas, with few teachers scoring even moderately high.

Conclusions

Summary of Findings

Reflecting on our initial evaluation questions and based on the analysis of the collected data, we feel that the IMPACT program succeeded in training and preparing their teacher apprentices for teaching in urban schools. The first cohort of graduates were generally satisfied with their learning experience, especially with mentorship, school placement, and their faculty advisor. The IMPACT program and training did not have a significant effect on increasing apprentices' mathematical and science content knowledge, but had a small positive effect in helping graduates to increase their scores and pass the PACT assessment.

While IMPACT participants set consistent instructional goals for procedural and conceptual understanding as well as developing academic language in their classrooms, each apprentice also adopted a different yet consistent set of tools for their classrooms. To a certain degree, there was evidence that apprentices had a tendency to stick to "familiar" components of classroom instruction (taking notes, providing oral feedback, assigning homework) and shied away from more complex activity structures (such as a jigsaw) and cognitively demanding questioning types.

IQA data indicates that while most assignments offered the potential for complex thinking and the opportunity to solve challenging problems, many teachers had more difficulty in connecting them to meaningfully real-world situations or student experiences. They also were challenged in their efforts to make students aware of their expectations and to support academic language. In particular, the bimodal nature of math scores in *Expectations-Communication* and *Equitable Teaching-Relevance* suggests that teachers were either very proficient or very poor in these areas, with few teachers scoring in the middle range.

The following is a brief summary of our results:

- Overall, teacher apprentices had a positive opinion of the IMPACT program and had a successful experience.
- Teacher apprentices showed small changes in their beliefs about social justice, consistent with program goals.
- The program did not significantly increase the pedagogical content knowledge of teachers.
- Although showing strength in some areas, the program did not significantly contribute to substantial changes in major teacher instructional strategies across lessons.
- Apprentices showed evidence of applying what they had learned during their "classroom" year in their post-IMPACT placements.
- Differences were found in the experience and practices of math and science teacher apprentices, suggesting different support needs between the two groups.

Study Limitations

Like all studies, our analyses were limited by program constraints and available data, which suggest careful use of our findings. For example, participants were given access to the IQA rubric in advance of our administration and may have prepared or submitted assignments with this rubric in mind. Additionally many teachers did not adhere to the original IQA timeline and the rubric itself may be slightly biased towards project-based learning; consequently, some of our findings may be skewed in a more positive direction than would otherwise be the case. Future studies should include assignments given at separate times during the school year to evaluate consistency and growth.

The low number of program participants reduces the validity of inferences that can be made from our findings and restricts our ability to provide more detailed results. For example, many of the teachers in our study taught a wide variety of subjects, making it more difficult to identify systematic and meaningful differences and interactions by subject matter, such as science or math instruction. We have essentially “averaged over” these subject differences in the current study. This is in addition to the limited statistical power considerations that come with such a small sample size and multiple hypotheses.

Furthermore, our analyses were limited because of the time constraints of the IMPACT program. For example, the program prescribed that the data collection of classroom logs would occur in March and May. We believe that this narrow window of time between data collections, together with the closeness to the end of the school year, may have hindered our ability to detect significant changes in instructional strategies. An initial data collection before the winter break, which was not logistically possible for this group, would have provided a longer interval for detecting instructional changes.

Recommendations

Considering the above findings and study limitations, we recommend the following:

We feel that our findings point to the need for improvements in training teachers in using academic language as indicated by the data from the teacher apprentice surveys, IQA, and PACT. We encourage IMPACT to investigate and follow up on this recommendation. Given that we found that teacher apprentices used a wide range of classroom practices and strategies, we recommend that the IMPACT program provide participants more opportunities to try out these practices during their 18-month training period.

IMPACT may also wish to address apprentices' lower level of satisfaction in the areas of communicating with parents, engaging out-of-classroom resources, and working with diverse populations of students, including special education students, per our survey results. Additionally, given the heavy time demands on teacher apprentices, we encourage IMPACT to compare their program goals to the actual program requirements and decide if some streamlining may be possible. Furthermore, we encourage IMPACT managers to consider providing differentiated support aligned to the specific needs of math and science teacher apprentices.

We also strongly suggest that the IMPACT program consider having the initial collection of teacher instructional log data in October or November instead of March. This would provide teachers a much longer window of time to try out newly learned and more complex classroom strategies that are part of the IMPACT program.

References

Berry, B., Montgomery, D., Curtis, R., Hernandez, M., Wurtzel, J., & Snyder, J. (2008). Creating and sustaining urban teacher residencies: A new way to recruit, prepare, and retain effective teachers in high-needs districts. Chapel Hill, NC: The Aspen Institute and Center for Teaching Quality.

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Belmont, CA: Wadsworth Group.

Crosson, A., Boston, M., Levison, A., Matsumura, L., Resnick, L. B., Wolf, M. K., & Junker, B. (2006). *Beyond summative evaluation: The Instructional Quality Assessment as a professional development tool* (CSE Tech. Report 691). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Hinchey, P. H. (2010). *Getting Teacher Assessment Right: What Policymakers Can Learn from Research*. Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/publication/getting-teacher-assessment-right.

Kane, T. J., Staiger, D. O., McCCaffrey, D., Cantrell, S., Archer, J., Buhayar, S., et al. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Seattle: Bill & Melinda Gates Foundation.

Matsumura, L.C., Slater, Junker, B., Peterson, M. B., Steele, M., & Resnick, L. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment.* (CSE Report 681). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Nichols, D.P. (1998). Choosing an intraclass correlation coefficient. *SPSS Keywords, 67.* Retrieved from http://www.ats.ucla.edu/stat/spss/library/whichicc.htm

Pecheone, R., & Chung, R. (2006). Evidence in teacher education: The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education, 57,* 22-36.

Pecheone, R., & Chung, R. R. (2007). *Technical report of the Performance Assessment for California Teachers (PACT): Summary of validity and reliability Studies for the 2003-04 Pilot Year.* Retrieved from PACT website: http://www.pacttpa.org/_files/Publications_and_Presentations/PACT_Technical_Report_March07.pdf

Rothstein, J. (2011). Review of learning about teaching [Review of report *Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project,* by Bill and Melinda Gates Foundation]. Retrieved July 20, 2012 from http://nepc.colorado.edu/thinktank/review-learning-about-teaching.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2):4-14.

Silk, Y., Silver, D., Amerian, S., Nishimura, C., & Boscardin, C.K. (2009). *Examining the effectiveness and validity of Read-Aloud and Glossary accommodations for English language learners in a math assessment.* (CRESST Report 761). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Smith, P.S., & Banilower, E. R. (2006). *Measuring teacher's knowledge for teaching force and motion concepts.* Presented at the 2006 NARST Annual Conference, San Francisco, CA.

UCLA Center X, Urban Teacher Residency (2012). IMPACT: Urban Teacher Residency Program Inspiring Minds through Professional Alliance of Community Teachers Program Handbook 2012- 2013. Retrieved from: http://centerx.gseis.ucla.edu/teacher-education/resources/tep-handbook-2011-2012/Program%20Handbook%2012-13%20FINAL.pdf

Appendix A:
Accompanying Tables

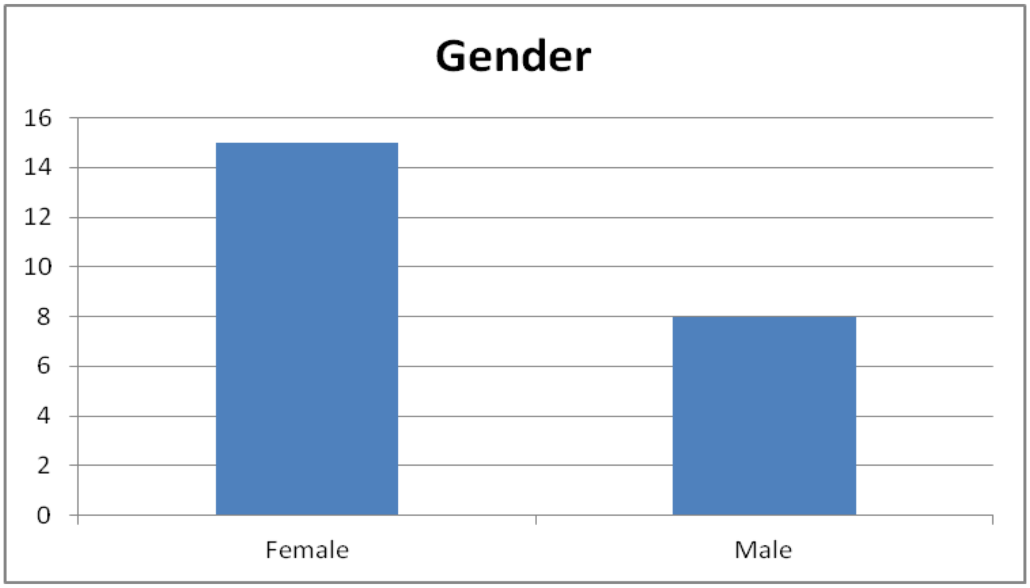

Figure A1.Cohort 1 math and science apprentice teachers responses to the question: What is your gender?

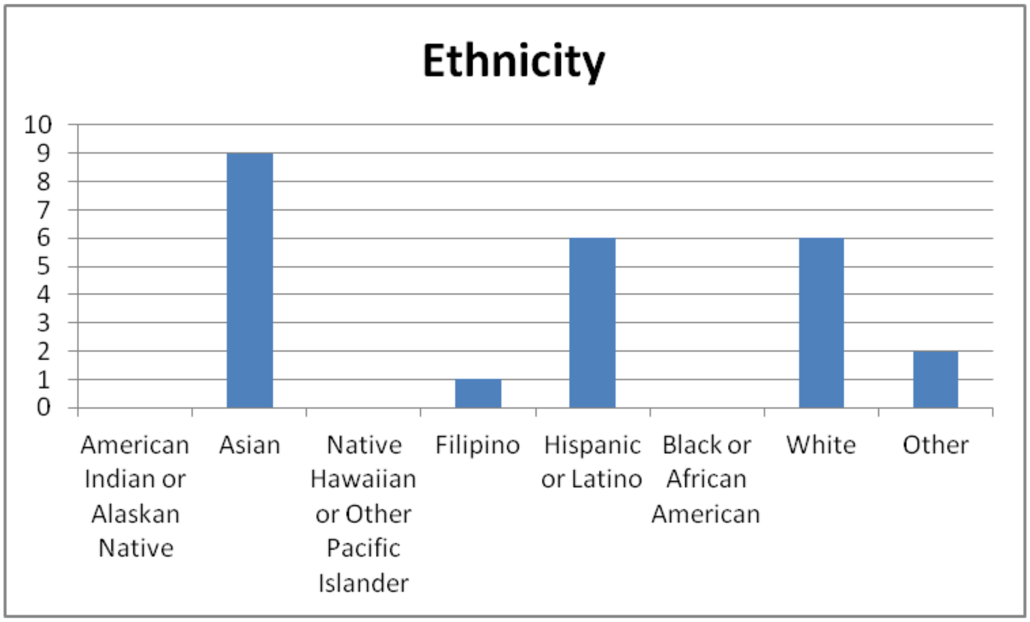

Figure A2. Cohort 1 math and science apprentice teachers' responses to the question: What is your ethnicity?

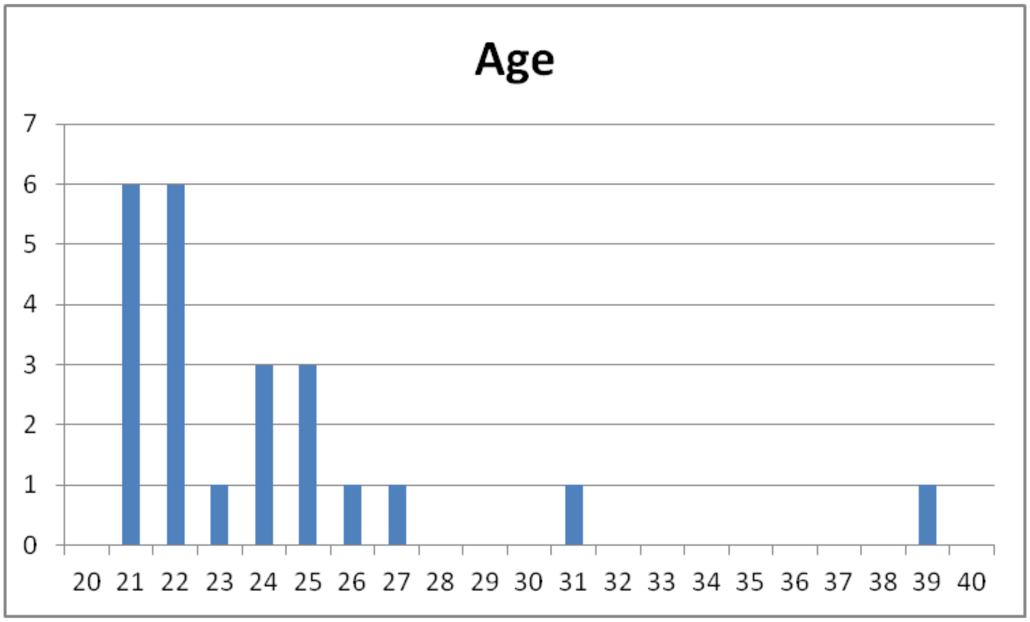

Figure A3. Cohort 1 math and science apprentice teachers' responses to: Please enter your current age?

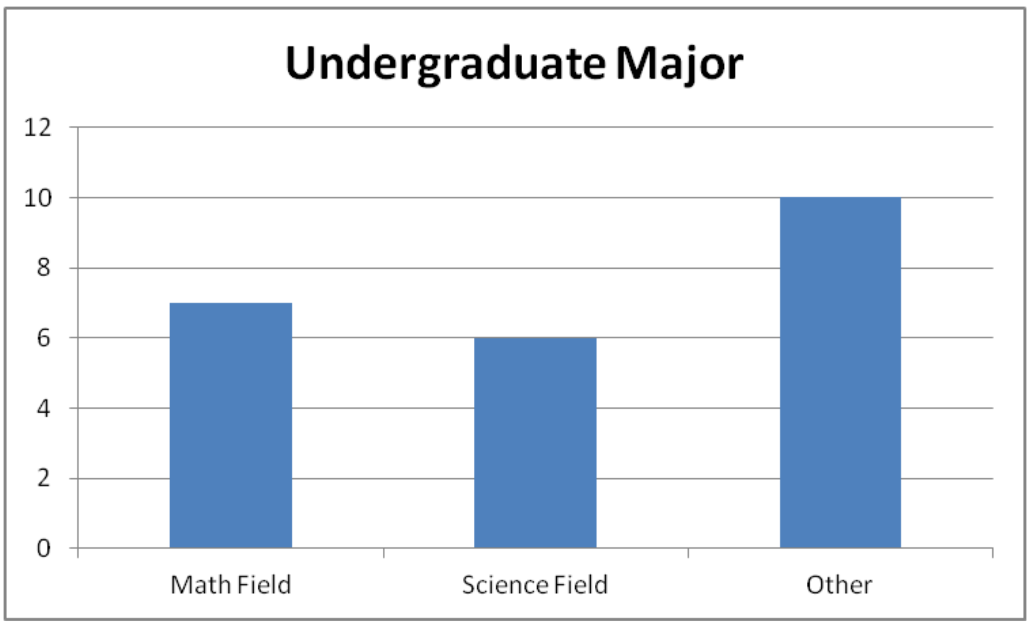

Figure A4. Cohort 1 math and science apprentice teachers' responses to the question: What was your major?

Table A1

Methods Coursework

Please rate your level of agreement with the following statements related to your IMPACT methods coursework.									
My IMPACT methods coursework (e.g. 320 & 405) helped me…									
	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
…learn to prepare lessons.	0%	5%	15%	60%	20%	0%	3.95	20	0.76
…learn to implement lesson plans.	0%	10%	20%	50%	20%	0%	3.8	20	0.89
…use assessment results formatively.	0%	15%	30%	40%	15%	0%	3.55	20	0.95
…build reflective teaching practices.	0%	0%	5%	40%	55%	0%	4.5	20	0.61
…develop strategies to improve my students' academic discourse.	0%	5%	20%	40%	35%	0%	4.05	20	0.89
…learn to provide equitable access to content.	0%	10%	30%	25%	30%	5%	3.79	20	1.03
…develop my identity as a social justice educator.	5%	0%	10%	30%	55%	0%	4.3	20	1.03
…reinforce what I learned in my fieldwork.	0%	0%	35%	40%	25%	0%	3.9	20	0.79

Table A2

Mentor

Please rate your level of agreement with the following statements about your mentorship experience:									
My mentor helped me…									
	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
…learn to prepare lessons.	0%	30%	10%	25%	35%	0%	3.65	20	1.23
…learn to implement lesson plans.	0%	30%	10%	20%	40%	0%	3.7	20	1.3
…use assessment results formatively.	10%	20%	25%	35%	10%	0%	3.15	20	1.18
…build reflective teaching practices.	10%	10%	25%	30%	25%	0%	3.5	20	1.23
…develop strategies to improve my students' academic discourse.	10%	20%	25%	35%	10%	0%	3.15	20	1.18
…learn to provide equitable access to content.	5%	10%	35%	30%	20%	0%	3.5	20	1.1
…develop my identity as a social justice educator.	25%	15%	15%	20%	25%	0%	3.05	20	1.57
… reinforce what I learned in my coursework.	5%	20%	25%	35%	15%	0%	3.35	20	1.14

Table A3

IMPACT Program

Please rate your level of agreement with the following statements about the effects of the IMPACT program. **The IMPACT program helped me…**	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
…develop strategies to maintain student engagement.	0%	5%	35%	30%	30%	0%	3.85	20	0.93
…develop practical classroom management strategies.	5%	0%	40%	25%	30%	0%	3.75	20	1.07
…feel prepared to work with students from diverse cultural backgrounds.	0%	5%	35%	25%	35%	0%	3.9	20	0.97
…engage parents in my students' education.	10%	20%	45%	15%	10%	0%	2.95	20	1.1
…communicate with parents from diverse linguistic and cultural backgrounds.	0%	35%	50%	10%	5%	0%	2.85	20	0.81
…collaborate productively with other teachers.	0%	20%	35%	30%	15%	0%	3.4	20	1
…build partnerships with multiple allies on and off campus.	5%	20%	40%	20%	10%	0%	3.11	20	1.05
…get to know the other students in my program well.	0%	5%	15%	35%	40%	0%	4.16	20	0.9
…learn to work with students with a diversity of prior skill levels.	0%	5%	25%	50%	20%	0%	3.85	20	0.81

Table A4

School Site Placement

Please rate your level of agreement with the following statements about your school site.	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
I feel connected to the school I teach in (OUTSIDE of my own classroom).	10%	10%	20%	50%	10%	0%	3.4	20	1.14
Observing teachers other than my mentor was useful.	0%	5%	5%	45%	45%	0%	4.3	20	0.801
Discussing instructional and classroom issues with other members of the school community was useful.	0%	5%	15%	45%	20%	15%	3.94	20	0.83
Working with members of the school community helped me learn to use assessment results to plan instruction.	10%	5%	35%	25%	0%	25%	3	20	1
Interacting with parent and/or community leaders supported my development as a teacher.	10%	10%	20%	40%	10%	10%	3.33	20	1.19
On-site professional development activities were useful.	25%	20%	20%	35%	0%	0%	2.65	20	1.23
Working with other members of the school community helped me to reflect on my teaching practices.	5%	20%	30%	40%	0%	5%	3.11	20	0.937
If given the opportunity, I would choose to teach at my current school placement in the future.	20%	0%	20%	30%	30%	0%	3.5	20	1.47
My field placement supported the development of my identity as a teacher.	0%	5%	20%	35%	40%	0%	4.1	20	0.91

Table A5

Instructional Time

Approximately what percent of instructional time with your mentor at the school site was spent on the following activities?	Fall Quarter					Winter Quarter					Spring Quarter					
	0%	1-25%	26-50%	51-75%	76-100%	0%	1-25%	26-50%	51-75%	76-100%	0%	1-25%	26-50%	51-75%	76-100%	
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	N
Watching your mentor lead instruction	5%	5%	35%	35%	20%	20%	40%	15%	25%	0%	35%	45%	15%	5%	0%	20
Co-leading instruction with your mentor	5%	50%	30%	10%	5%	25%	40%	30%	5%	0%	45%	35%	15%	5%	0%	20
Leading instruction independently (while your mentor observed you)	5%	55%	25%	0%	15%	0%	10%	30%	25%	35%	0%	0%	15%	5%	80%	20

Table A6

Instructional Activities with Mentor

How often did you engage in the following types of activities with your mentor outside of instructional time?

	Never 1	Rarely 2	Sometimes 3	Often 4	M	N	SD
Co-designing/planning lessons	0%	20%	40%	40%	3.2	20	0.77
Reviewing student work	5%	35%	35%	20%	2.74	19	0.87
Receiving feedback about your in-class instructional practices	0%	15%	45%	40%	3.25	20	0.716
Discussing educational research/pedagogy	15%	35%	35%	15%	2.5	20	0.95
Communicating with the parents of your students	40%	35%	25%	0%	1.85	20	0.813
Discussing how to leverage out-of-classroom resources (e.g. Parent Center,	40%	50%	10%	0%	1.7	20	0.66
Analyzing student or other school data	25%	35%	25%	15%	2.3	20	1.03
Meeting about other issues	5%	25%	50%	20%	2.85	20	0.81

Table A7

Communication with Mentor

Please rate your level of agreement with the following statements related to your mentoring experience.	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
Working alongside my mentor during instructional time in the classroom was helpful.	0%	5%	15%	35%	45%	0%	4.2	20	0.89
Using the dialog journal was a helpful way to communicate with my mentor.	60%	15%	20%	5%	0%	0%	1.7	20	0.98
The mentor-faculty advisor debrief was a useful way to communicate with my mentor.	10%	20%	10%	40%	10%	10%	3.22	20	1.26
The online journal was a useful way for me to communicate with my mentor.	70%	5%	10%	0%	5%	10%	1.5	20	1.1
Communicating with my mentor over email was helpful.	20%	5%	20%	20%	35%	0%	3.45	20	1.54
Outside of the dialog journal, the written feedback I received from my mentor was helpful.	15%	20%	20%	20%	20%	5%	3.11	20	1.41

Table A8

Community-Based Organizations

Please rate your level of agreement with the following statements about your experience with community-based organizations. **I feel the community-based organizations helped me…**	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
…by addressing the needs of teachers.	10%	35%	30%	15%	10%	0%	2.8	20	1.15
…build stronger ties to the community.	10%	30%	35%	10%	15%	0%	2.9	20	1.21
…build relationships with families.	15%	35%	35%	5%	10%	0%	2.6	20	1.14
…understand the strengths of the community I teach in.	5%	25%	20%	35%	15%	0%	3.3	20	1.17

Table A9

Community-Based Organizations Overall

Please rate your level of agreement with the following statement about your experience with community-based organizations.	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
Overall, I consider these community-based organizations to be valuable resources.	5%	5%	15%	40%	35%	0%	3.95	20	1.1

Table A10

Methods Coursework Overall

Please rate your level of agreement with the following statement related to your IMPACT methods coursework.	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
Overall, I feel satisfied with my methods coursework.	0%	10%	30%	40%	20%	0%	3.7	20	0.92

Table A11

Mentor Overall

Please rate your level of agreement with the following statement.	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
Overall, I feel satisfied with my relationship with my current mentor.	0%	20%	0%	35%	45%	0%	4.05	20	1.15

Table A12

School Placement Overall

Please rate your level of agreement with the following statement related to your School Site Placement.	Strongly	2	3	4	Strongly	NA	M	N	SD
Overall, I feel satisfied with my current school placement.	0%	15%	5%	40%	40%	0%	4.05	20	1.05

Table A13

IMPACT Program Overall

Please rate your level of agreement with the following statement about your experience with the IMPACT program.	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
Overall, I am satisfied with the IMPACT program.	0%	10%	30%	50%	10%	0%	3.6	20	0.82

Table A14

Weekly Seminar

Please rate your level of agreement with the following statements.

The Weekly Seminar Course (EDUC 481) I attended at UCLA helped me…

	Strongly Disagree 1	2	3	4	Strongly Agree 5	M	N	SD
…further develop my own personal teaching philosophy.	5%	15%	35%	40%	5%	3.25	20	0.968
…think about the relationships among theory, research, and practice.	15%	15%	25%	35%	10%	3.1	20	1.25
…use assessment results formatively.	10%	45%	40%	0%	5%	2.45	20	0.89
…build reflective teaching practices.	0%	10%	15%	55%	20%	3.85	20	0.88
…develop strategies to improve my students' academic discourse.	10%	35%	35%	20%	0%	2.65	20	0.93
…learn to provide equitable access to content.	15%	40%	40%	5%	0%	2.35	20	0.81
…develop my identity as a social justice educator.	0%	15%	40%	30%	15%	3.45	20	0.95
…reflect on my experiences in urban schools.	0%	0%	15%	50%	35%	4.2	20	0.7
…build strong relationships with other IMPACT teachers in my cohort.	0%	15%	15%	25%	45%	4	20	1.12

Table A15

Master’s Portfolio

Please rate your level of agreement with the following statements.								
The master's portfolio helped me…								
	Strongly Disagree 1	2	3	4	Strongly Agree 5	M	N	SD
…develop classroom management strategies.	5%	35%	30%	20%	10%	2.95	20	1.1
…develop strategies for working with families.	15%	35%	30%	20%	0%	2.55	20	1
…better understand the Performance Assessment for California Teachers.	15%	10%	20%	40%	15%	3.3	20	1.3
…further develop my own personal teaching philosophy.	0%	0%	30%	40%	30%	4	20	0.8
…think about the relationships among theory, research, and practice.	0%	5%	20%	50%	25%	3.95	20	0.83
…use assessment results formatively.	5%	5%	40%	40%	10%	3.45	20	0.95
…build reflective teaching practices.	0%	0%	5%	55%	40%	4.35	20	0.59
…develop strategies to improve my students’ academic discourse.	5%	15%	10%	60%	10%	3.55	20	1.05
…learn to provide equitable access to content.	5%	15%	20%	50%	10%	3.45	20	1.05
…develop my personal identity as a social justice educator.	0%	0%	40%	30%	30%	3.9	20	0.85
…reflect on my experiences in urban schools.	0%	0%	10%	50%	40%	4.3	20	0.66

Table A16

Cohort 1 science apprentice information about their current teaching position

Cohort I Science Teachers	1	2	3	4	5	6	7	8	9	10	Totals:
Grade(s)											
6th	X										1
7th	X								X		2
8th	X										1
9th			X	X	X		X			X	5
10th			X	X		X	X			X	5
11th		X	X	X		X	X	X		X	7
12th		X	X				X	X			4
Course(s)											
STEM Lab	X										1
Advisory	X			X							2
Chemistry		X					X			X	3
Physics			X					X			2
Earth Science			X								1
Biology				X		X				X	3
Intercoordinated Science					X		X				2
Anatomy and Physiology						X					1
Life Science									X		1
Number of Class Periods											
3						X				X	2
4		X			X		X				3
5	X			X				X			3
6			X						X		2
Average Class Size											
20-25		X		X	X	X		X			5
26-30									X		1
31-35	X		X							X	3
36-40							X				1
Special Education Students											
Yes	X	X		X	X	X	X		X	X	8
No			X					X			2

Table A17

Cohort 1 math apprentice information about their current teaching position

Cohort I MathTeachers	1	2	3	4	5	6	7	8	9	10	Totals:
Grade(s)											
6th			X								1
7th											0
8th											0
9th	X	X		X				X	X	X	6
10th	X	X		X	X	X		X	X		7
11th	X	X		X	X	X	X	X	X		8
12th	X	X		X	X	X	X		X		7
Course(s)											
6th Grade Math			X								1
Algebra I	X	X		X				X		X	5
Algebra II		X		X	X	X	X	X	X		7
Geometry					X				X		2
Trigonometry	X						X				2
Calculus				X		X					2
Math Analysis						X	X				2
CAHSEE Math						X					1
Finite Math									X		1
Number of Class Periods											
3	X						X	X		X	4
4				X	X				X		3
5		X	X			X					3
Average Class Size											
16-20		X	X		X						3
21-25						X					1
26-30	X			X							2
31-35							X		X	X	3
36-40								X			1
Special Education Students											
Yes	X	X	X			X	X	X	X	X	8
No				X	X						2

Table A18

Cohort 1 math and science apprentice information about their current teaching position

Cohort I -Math and Science	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Totals:
Interact with Parents																					
Never																					0
Rarely	X	X		X		X			X	X		X			X	X	X	X			11
Sometimes			X		X		X	X			X		X	X					X		8
Often																				X	1
Use Parent Center?																					
Yes				X															X	X	3
No	X	X	X		X			X	X	X	X	X		X	X		X	X			13
No parent center						X	X						X			X					4
UCLA Teachers @ School? IMPACT and TEP																					
Yes	X		X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	18
No		X														X					2
How Often Do You Meet?																					
Never			X															X			2
Rarely						X														X	2
Sometimes	X				X		X	X		X			X				X				7
Often				X					X		X	X		X	X				X		7
N/A		X														X					2

Table A19

The School You Currently Teach In

Please rate your level of agreement with the following statements related to **the school you currently teach in.**	Strongly Disagree 1	2	3	4	Strongly Agree 5	M	N	SD
The mission/vision of the school is compatible with my own.	10%	15%	30%	30%	15%	3.25	20	1.2
This school supports my development as a social justice educator.	5%	25%	20%	40%	10%	3.25	20	1.19
I feel I am part of a collegial network of teachers.	5%	10%	35%	15%	35%	3.65	20	1.23
I feel supported by the school leadership.	10%	10%	20%	25%	35%	3.65	20	1.35
This school offers me opportunities for professional growth and development.	10%	10%	15%	45%	20%	3.55	20	1.23
Overall, I am satisfied with the school I currently teach in.	0%	10%	30%	40%	20%	3.7	20	0.92

Table A20

Compared to My IMPACT Placement Last Year (a)

Please compare **the school you currently teach in** with your **IMPACT placement last year.** Compared to my IMPACT placement last year…	Less 1	2	The Same 3	4	More 5	M	N	SD
I observe other teachers teach…	36.80%	15.80%	21.10%	10.50%	15.80%	2.53	19	1.5
I discuss instructional and classroom issues with other members of the school community…	10%	10%	15%	20%	45%	3.8	20	1.4
I work with members of the school community to learn to use assessment results…	10%	5%	35%	25%	25%	3.5	20	1.23
I interact with parents and/or community members…	10%	5%	35%	35%	15%	3.4	20	1.14
I participate in on-site professional development…	0%	0%	5%	20%	75%	4.7	20	0.57

Table A21

Compared to My IMPACT Placement Last Year (b)

Please compare **the school you currently teach in** with your **IMPACT placement last year.** Compared to my IMPACT placement last year…	Strongly Disagree 1	2	3	4	Strongly Agree 5	M	N	SD
I feel more connected to the school I currently teach in.	5%	0%	20%	10%	65%	4.3	20	1.13
my current school better supports my development as a social justice educator.	5%	25%	30%	20%	20%	3.25	20	1.21
I am more satisfied with the school I currently teach in.	0%	15%	35%	25%	25%	3.6	20	1.05
I would rather teach at this school in the future.	0%	20%	40%	10%	30%	3.5	20	1.15

Table A22

Training Prior to June

The following questions refer to how your training prior to June in the IMPACT program prepared you to work in your current teaching assignment. Please rate your level of agreement with the following items.

The training I received prior to June through the IMPACT program prepared me to…

	Strongly Disagree 1	2	3	4	Strongly Agree 5	NA	M	N	SD
…prepare lessons.	0%	10%	10%	45%	35%	0%	4.05	20	0.95
…implement lesson plans.	0%	0%	25%	35%	40%	0%	4.15	20	0.81
…use assessment results formatively.	0%	15%	35%	35%	15%	0%	3.5	20	0.95
…build reflective teaching practices.	0%	0%	5%	40%	55%	0%	4.5	20	0.61
…develop strategies to improve my students' academic discourse.	0%	5%	25%	35%	35%	0%	4	20	0.92
…provide equitable access to content.	0%	5%	35%	40%	20%	0%	3.75	20	0.85
…work as a social justice educator.	0%	0%	30%	20%	50%	0%	4.2	20	0.9
…have strategies to maintain student engagement.	0%	5%	40%	30%	25%	0%	3.75	20	0.91
…have practical classroom management strategies.	20%	30%	30%	5%	15%	0%	2.65	20	1.31
…work with students from diverse cultural backgrounds.	5%	15%	35%	10%	35%	0%	3.55	20	1.28
…work with special education students.	30%	10%	30%	25%	5%	0%	2.65	20	1.31
…engage parents in my students' education.	0%	36.80%	42.10%	10.50%	10.50%	0%	2.95	19	0.98
…communicate with parents from diverse linguistic and cultural backgrounds.	10%	25%	40%	15%	10%	0%	2.9	20	1.12
…collaborate productively with other teachers.	0%	15%	35%	40%	10%	0%	3.45	20	0.89
…work with students with a diversity of prior skill levels.	0%	15%	35%	30%	20%	0%	3.55	20	1
…effectively manage teaching multiple class periods.	30%	5%	30%	5%	25%	5%	2.89	20	1.6

Table A23

Weekly Seminar Overall

Please rate your level of agreement with the following statement.	Strongly Disagree 1	2	3	4	Strongly Agree 5	M	N	SD
Overall, I feel satisfied with the weekly seminar (EDUC 481) course I attended at UCLA.	0%	25%	30%	25%	20%	3%	20	1.1

Table A24

Master's Portfolio Overall

Please rate your level of agreement with the following statement related to the **master's portfolio**.	Strongly Disagree 1	2	3	4	Strongly Agree 5	M	N	SD
Overall, I feel satisfied with the master's portfolio.	5%	20%	25%	50%	0%	3.2	20	0.95

Table A25

Feedback Item Descriptives

Item	Math			Science		
	Occasion 1	Occasion 2	Combined	Occasion 1	Occasion 2	Combined
I provided oral feedback to students about their work one on one	70.00%	61.80%	66.20%	83.70%	70.80%	79.10%
I provided oral feedback to students about their work in a whole-class setting.	80.00%	79.40%	79.70%	69.80%	75.00%	71.60%
I provided grades on student output (e.g. grades on homework assignments)	43.60%	35.30%	39.70%	31.00%	29.20%	30.30%
I provided other written feedback on student output, including written comments on homework assignments.	17.90%	14.70%	16.40%	31.00%	33.30%	31.80%
Students assessed their own work	43.60%	58.80%	50.70%	38.10%	29.20%	34.80%
Students assessed the work of others	35.90%	44.10%	39.70%	24.40%	25.00%	24.60%

Table A26

Questioning Type Item Descriptives - Math

Item	Math		
	Occasion 1	Occasion 2	Combined
IRE (Initiate-Respond-Evaluate): (Yes/No questions, questions that ask students to recall facts or definitions)	85.00%	91.20%	87.80%
Self-Reflection: (Why do you think that? Why is that true? How did you reach that conclusion?)	70.00%	85.30%	77.00%
Reflection-on-Others: (Do you agree? Does anyone have a different way to explain it? Would you ask the rest of the class that question?)	64.10%	64.70%	64.40%
Pattern-Finding and Conjecturing: (How did you predict the next case? What is similar and what is different about your solution and his/hers? Do you see a pattern?)	40.00%	35.30%	37.80%
Mathematical Reasoning: (Does that always work? Is that true for all cases? Can you think of a counter example?)	37.50%	17.60%	28.40%
Mathematical Connections: (Have we ever solved a problem like this before? What have we learned before that were useful in solving this problem?)	67.50%	52.90%	60.80%

Table A27

Questioning Type Item Descriptives - Science

Item	Science		
	Occasion 1	Occasion 2	Combined
IRE (Initiate-Respond-Evaluate): (Yes/No questions, questions that ask students to recall facts or definitions)	70.50%	75.00%	72.10%
Understanding: (Explain the concepts of..., Give me an example of...,Illustrate how ____ works)	73.30%	87.50%	78.30%
Analyze: (compare and contrast questions, How do your conclusions support your hypothesis? Which errors most affected your results?)	32.50%	45.80%	37.50%
Evaluate: (Is that true for all cases? Can you think of a counter example? How could you prove that?)	19.50%	33.30%	24.60%
Create: (Design a lab to show..., Predict what will happen to ____ as _____ is changed, Using a principle of science, how can we find...?	27.50%	16.70%	23.40%

Table A28

Instructional Strategy Item Descriptives - Math

Item	Math		
	Occasion 1	Occasion 2	Combined
Sharing thinking with a partner (Dyad, Think-Pair-Share)	20.50%	47.10%	32.90%
Jigsaw	2.60%	0.00%	1.40%
Save the Last Word for Me	2.60%	0.00%	1.40%
Poster/Problem Presentation	28.20%	38.20%	32.90%
Socratic Seminar (Fish Bowl)	0.00%	0.00%	0.00%
Whip Around, Repeating/Re-voicing	55.30%	23.50%	40.30%
Homework was assigned to students	80.00%	58.80%	70.30%
Teacher-made test or quiz	35.90%	32.40%	34.20%
Commercial test or quiz	5.10%	14.70%	9.60%
Open ended tests or assessments (performance tasks, portfolios)	12.80%	29.40%	20.50%
Problem of the Week Protocol	2.60%	0.00%	1.40%
Five(four) fold way (multiple representations or solutions)	12.80%	0.00%	6.80%
Short written summary or reflection	33.30%	23.50%	28.80%
Say-Mean-Matter	2.60%	0.00%	1.40%
Write-Pair-Share	7.70%	14.70%	11.00%
Concept map or Graphic Organizer	30.80%	26.50%	28.80%
KWL Chart	7.70%	0.00%	4.10%
Boards up	15.40%	11.80%	13.70%
Interactive technologies (clickers, calculators, SmartBoard, etc.)	30.80%	23.50%	27.40%
Notetaking: Interactive Notebooks, Journaling, Cornell Notes	56.40%	17.60%	38.40%
Role Play and Acting it Out	0.00%	0.00%	0.00%
Gestures (Thumbs Up/Down, Head nodding, etc.)	70.00%	64.70%	67.60%

Table A29

Instructional Strategy Item Descriptives - Science

Item	Science		
	Occasion 1	Occasion 2	Combined
Sharing thinking with a partner (Dyad, Think-Pair-Share)	57.10%	43.50%	52.30%
Think Aloud	4.90%	37.50%	16.90%
Student Linking Ideas	56.10%	62.50%	58.50%
Teacher Linking Ideas	57.50%	62.50%	59.40%
Reciprocal teaching	26.80%	56.50%	37.50%
Jigsaw	0.00%	8.30%	3.10%
Describe, not Show	30.00%	16.70%	25.00%
Save the Last Word for Me	0.00%	0.00%	0.00%
Poster/Problem Presentation	22.00%	8.30%	16.90%
Socratic Seminar (Fish Bowl)	0.00%	8.30%	3.10%
Whip Around, Repeating/Re-voicing	2.40%	4.20%	3.10%
Sentence Starters	4.90%	4.30%	4.70%
Read Around (Read out Loud)	16.70%	12.50%	15.20%
Repeating/revoicing	28.60%	12.50%	22.70%
Cold Call	17.50%	16.70%	17.20%
Popcorn response	2.40%	4.20%	3.10%
Predict, Observe, Explain (POE)	12.20%	16.70%	13.80%
Homework was assigned to students	72.10%	33.30%	58.20%
Teacher-made test or quiz	19.00%	4.20%	13.60%
Commercial test or quiz	2.40%	4.20%	3.10%
Open ended tests or assessments (performance tasks, portfolios, Lab report)	41.50%	25.00%	35.40%
Short written summary or reflection	34.10%	29.20%	32.30%
Say-Mean-Matter	0.00%	4.20%	1.50%
Write-Pair-Share	12.20%	8.30%	10.80%
Concept map or Graphic Organizer	39.00%	12.50%	29.20%
KWL Chart	7.30%	0.00%	4.60%
Boards up	0.00%	0.00%	0.00%
Draw Around	0.00%	0.00%	0.00%
Interactive technologies (clickers, calculators, SmartBoard, etc.)	28.60%	12.50%	22.70%
Notetaking: Interactive Notebooks, Journaling, Cornell Notes	52.40%	50.00%	51.50%
Role Play and Acting it Out	2.40%	0.00%	1.50%
Gestures (Thumbs Up/Down, Head nodding, etc.)	29.30%	29.20%	29.20%

Table A30

Instructional Goals Item Descriptives

Math		Science	
Goal	Percent	Goal	Percent
Procedural Fluency	77.60%	Recognizing and recalling scientific facts	68.60%
Mathematical Conceptual Understanding	61.80%	Scientific conceptual understanding	79.40%
Problem Solving, Application, Mathematical Reasoning	55.30%	Developing testable scientific questions and hypotheses	2.90%
Academic Language	42.10%	Academic Language	70.60%
Positive Mathematical Disposition	36.80%	Designing or conducting investigations to test scientific questions/hypotheses	8.80%
		Analyzing data and drawing conclusions from scientific investigations	11.80%
		Using scientific concepts and evidence to support hypotheses, and explain and justify claims	8.80%
		Communicating results of investigations to others (writing, oral and visual)	25.00%
		Mastery of laboratory skills, procedures and techniques	10.30%
		Positive science disposition	33.80%

Table A31

Cross-Tabulation of Instructional Goal and Questioning Strategy for Mathematics Teachers

Goal		IRE (Initiate-Respond-Evaluate)		Self-Reflection		Reflection-on-Others		Pattern-Finding and Conjecturing		Mathematical Reasoning		Mathematical Connections	
		No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Procedural Fluency	Yes	8% (5)	92% (54)	24% (14)	76% (45)	34% (20)	66% (38)	63% (37)	37% (22)	76% (45)	24% (14)	37% (22)	63% (37)
	No	27% (4)	73% (11)	20% (3)	80% (12)	40% (6)	60% (9)	60% (9)	40% (6)	53% (8)	47% (7)	47% (7)	53% (8)
Mathematical Conceptual Understanding	Yes	9% (4)	91% (43)	19% (9)	81% (38)	37% (17)	63% (29)	64% (30)	36% (17)	64% (30)	36% (17)	34% (16)	66% (31)
	No	19% (5)	81% (22)	30% (8)	70% (19)	33% (9)	67% (18)	59% (16)	41% (11)	85% (23)	15% (4)	48% (13)	52% (14)
Problem Solving, Application, Mathematical Reasoning	Yes	17% (7)	83% (35)	26% (11)	74% (31)	33% (14)	67% (28)	62% (26)	38% (16)	67% (28)	33% (14)	31% (13)	69% (29)
	No	6% (2)	94% (30)	19% (6)	81% (26)	39% (12)	61% (19)	63% (20)	38% (12)	78% (25)	22% (7)	50% (16)	50% (16)
Academic Language	Yes	0% (0)	100% (32)	13% (4)	88% (28)	28% (9)	72% (23)	59% (19)	41% (13)	63% (20)	38% (12)	34% (11)	66% (21)
	No	21% (9)	79% (33)	31% (13)	69% (29)	41% (17)	59% (24)	64% (27)	36% (15)	79% (33)	21% (9)	43% (18)	57% (24)
Positive Mathematical Disposition	Yes	11% (3)	89% (25)	14% (4)	86% (24)	7% (2)	93% (26)	54% (15)	46% (13)	75% (21)	25% (7)	32% (9)	68% (19)
	No	13% (6)	87% (40)	28% (13)	72% (33)	53% (24)	47% (21)	67% (31)	33% (15)	70% (32)	30% (14)	43% (20)	57% (26)

Table A32

Cross-Tabulation of Instructional Goal and Questioning Strategy for Science Teachers

Goal		IRE (Yes/No questions)		Understanding (explain the concept of…)		Analyze (compare and contrast…)		Evaluate (Is that always true?)		Create (design a lab, predict what will happen…)	
		No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Recognizing and recalling scientific facts	Yes	23% (11)	77% (36)	25% (12)	75% (36)	60% (27)	40% (18)	75% (33)	25% (11)	84% (36)	16% (7)
	No	38% (8)	62% (13)	14% (3)	86% (18)	68% (13)	32% (6)	76% (16)	24% (5)	62% (13)	38% (8)
Scientific conceptual understanding	Yes	25% (13)	75% (40)	22% (12)	78% (42)	57% (28)	43% (21)	70% (35)	30% (15)	76% (37)	24% (12)
	No	43% (6)	57% (8)	21% (3)	79% (11)	79% (11)	21% (3)	93% (13)	7% (1)	79% (11)	21% (3)
Developing testable scientific questions and hypotheses	Yes	0% (0)	100% (2)	50% (1)	50% (1)	0% (0)	100% (2)	50% (1)	50% (1)	50% (1)	50% (1)
	No	29% (19)	71% (46)	21% (14)	79% (52)	64% (39)	36% (22)	76% (47)	24% (15)	77% (47)	23% (14)
Academic Language	Yes	23% (11)	77% (37)	21% (10)	79% (38)	62% (29)	38% (18)	81% (38)	19% (9)	78% (36)	22% (10)
	No	42% (8)	58% (11)	25% (5)	75% (15)	63% (10)	38% (6)	59% (10)	41% (7)	71% (12)	29% (5)
Designing or conducting investigations to test scientific questions/hypotheses	Yes	67% (4)	33% (2)	17% (1)	83% (5)	50% (3)	50% (3)	83% (5)	17% (1)	17% (1)	83% (5)
	No	25% (15)	75% (46)	23% (14)	77% (48)	63% (36)	37% (21)	74% (43)	26% (15)	82% (47)	18% (10)
Analyzing data and drawing conclusions from scientific investigations	Yes	25% (2)	75% (6)	25% (2)	75% (6)	50% (4)	50% (4)	88% (7)	13% (1)	88% (7)	13% (1)
	No	29% (17)	71% (42)	22% (13)	78% (47)	64% (35)	36% (20)	73% (41)	27% (15)	75% (41)	25% (14)
Using scientific concepts and evidence to support hypotheses, and explain and justify claims	Yes	17% (1)	83% (5)	17% (1)	83% (5)	50% (3)	50% (3)	100% (6)	0% (0)	100% (6)	0% (0)
	No	30% (18)	70% (43)	23% (14)	77% (48)	63% (36)	37% (21)	72% (42)	28% (16)	74% (42)	26% (15)
Communicating results of investigations to others (writing, oral and visual)	Yes	65% (11)	35% (6)	12% (2)	88% (15)	65% (11)	35% (6)	82% (14)	18% (3)	47% (8)	53% (9)
	No	16% (8)	84% (42)	25% (13)	75% (38)	61% (28)	39% (18)	72% (34)	28% (13)	87% (40)	13% (6)
Mastery of laboratory skills, procedures and techniques	Yes	86% (6)	14% (1)	29% (2)	71% (5)	71% (5)	29% (2)	71% (5)	29% (2)	14% (1)	86% (6)
	No	22% (13)	78% (47)	21% (13)	79% (48)	61% (34)	39% (22)	75% (43)	25% (14)	84% (47)	16% (9)
Positive science disposition	Yes	30% (7)	70% (16)	17% (4)	83% (19)	70% (16)	30% (7)	74% (17)	26% (6)	70% (16)	30% (7)
	No	27% (12)	73% (32)	24% (11)	76% (34)	58% (23)	43% (17)	76% (31)	24% (10)	80% (32)	20% (8)

Appendix B:
IMPACT Math/Science Mentor Survey Results

Demographics. Approximately half of mentors are female (55%). teacher mentors identified as Asian (50%), Hispanic (15%), and White (45%). Mentor responses to how long they have been working at their current school site ranged from 0 to 24 years, with 90% of mentors indicating 0 to 5 years. Half of mentors have been a supervisor of new teachers in the past. And when asked about their motivation for participating in the IMPACT program, 90% of mentors indicated additional training and professional development and the opportunity to work with new teachers.

Respondent perceptions of program participation. Overall, the math/science mentors believed their participation in the IMPACT program was a positive experience. On a scale from 1-5, the average rating of the overall experience was 4.17 (*SD* 1.043) The aspects of the IMPACT program overall that were the most meaningful to math/science mentors were working closely with their apprentice and having an impact on their teaching. Specifically, mentors voiced the strengths of their apprentices and the value of working with them over time. For example, a typical mentor response follows:

> Working with an apprentice who was a competent hard worker and collaborating to find strategies to engage students.

The aspects of the IMPACT program overall that were the least meaningful to math/science mentors were the lack of clear expectations from the program. Specifically, math and science mentors feel strongly that the program expectations regarding mentorship were unclear prior to program participation and constantly changed. For example, typical mentor responses follow:

> More upfront information and clearer definitions of the role of the mentor and mentee expectations.

> The roles and responsibilities were changed over the course of the year. It changed how often students saw their teacher.

Math/science mentors' suggestions for improving the IMPACT program overall focused primarily on creating a community where mentors can communicate with one another and learn strategies and discuss their experiences. Also, math and science mentors voiced a desire for more practice and education about how to mentor apprentices so they get the most out of the experience. For example a typical math/science apprentice response follows:

> Allow mentors more time to exchange ideas on mentor/mentee relationships and how to work with a mentee.

> Less cognitive coaching and more ways/strategies to help mentors help their apprentices and more time for reflection and best practices.

Opinions about the specific aspects of the development activities that mentors participated in ranged, and mentors found some dimensions more useful than others. In particular, mentors found the mentor development activities useful in developing an overall framework for mentorship—this includes developing strategies, reflecting on mentoring practice, and communicating with apprentices. Mentors found the development least helpful in four areas: reviewing student work, discussing research, classroom management, and assessing apprentice performance. The mean ratings of these items were nearly 2 scale points lower than the mean ratings for the highest rated items. The mentor's comments reflect these trends. Mentors often lauded the cognitive coaching development, and voiced the usefulness of practicing effective communication with their apprentices. For example, a typical mentor response follows:

> Cognitive coaching sessions were helpful in allowing me to communicate and guide my apprentice.

Other comments noted that the professional development sessions would have been more useful if they all took place at the IMPACT program, and if they had been scheduled closer together. For example, typical mentor responses follow:

> I wish activities were BEFORE the year, to be more <u>prepared</u>.

> A lot of them had a long time apart making it hard for us to remember what we did previously.

> I wanted more time to share out experiences with other mentors and to discuss and hone our roles as mentors.

Respondent perceptions of time. Overall, mentors felt that the compromises that they had to make to accommodate mentorship were worthwhile – a mean rating of 4.06 (*SD* .938). Mentors believe that they are spending the bulk of their program time with their mentees (mean of approximately 80%). However, this breakdown is very close to what mentors feel is an ideal breakdown between mentorship and other program elements (mentor development activities), as the average recommendation was for an 80/20 split among these activities.

While working with apprentices (outside of instructional time), mentors reported spending the most time analyzing student data, and the least amount of time planning lessons (almost a full scale point lower).

Coinciding with responses given by apprentices, over the course of the year mentor roles shifted in the classroom. Mentors spent less time leading instruction, and more time watching apprentices lead instruction, indicating that as the year progressed, apprentices were given more instructional responsibility in the average classroom.

Perceptions of apprentice preparedness. Mentors on average believed that their apprentices were well prepared to teach. There was strong agreement that apprentices began school year prepared to teach, and that they are prepared to implement lessons. In addition, there was strong agreement (*M*=4.35, *SD* = .875) that apprentices improved over the course of the school year.

Teaching philosophy. Two scales were administered as pre-post scales to mentors. These scales were broadly about attitudes concerning social justice and equity in schooling. There were several items where mentors exhibited change in levels of agreement. In the social justice scale, there were two items of note:

> Although teachers have to appreciate diversity, it's not their job to change society. (almost .9 scale points)
>
> Whether students succeed in school depends primarily on how hard they work. (slightly more than .75 scale points)

In both cases, mentors were, on average, less likely to agree with these statements on the post-survey than they were on the pre-survey.

On the Achievement Gap scale, mentors were less likely to believe that funding and student behavior are important factors (almost half a scale point), and more likely to believe that parent education is an important factor (.75 scale points).

Appendix C:
IQA Rubrics

Instructional Quality Assessment (IQA)

2012

National Center for Research on Evaluation, Standards, & Student Testing
UCLA | Graduate School of Education & Information Studies

TRANSFORMING PUBLIC SCHOOLS
UCLA CENTER X

Scale	Dimension 1: Academic Rigor	Rubric 1: Potential of the Task
4	**The task has the potential to engage students in exploring and understanding the nature of concepts, procedures, and/or relationships, such as:** • Requires complex thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example); OR • Procedures with connections: applying a broad general procedure that remains closely connected to concepts. **The task must explicitly prompt for evidence of students' reasoning and understanding. For example, the task MAY require students to:** • solve a genuine, challenging problem for which students' reasoning is evident in their work on the task; • develop an explanation for why formulas or procedures work; • identify patterns and form generalizations based on these patterns; • make conjectures and support conclusions with evidence; • make explicit connections between representations, strategies, or concepts and • procedures; • follow a prescribed procedure in order to explain/illustrate a concept, process, or relationship.	
3	**The task has the potential to engage students in complex thinking or in creating meaning for concepts, procedures, and/or relationships. However, the task does not warrant a "4" because:** • the task does not explicitly prompt for evidence of students' reasoning and understanding. • students may be asked to engage in academic concepts or procedures with connections, but the underlying content knowledge in the task is not appropriate for the specific group of students (i.e., too easy or too hard to promote engagement with high-level cognitive demands); • students may need to identify patterns but are not pressed for generalizations; • students may be asked to use multiple strategies or representations but the task does not explicitly prompt students to develop connections between them; • students may be asked to make conjectures but are not asked to provide evidence or explanations to support conclusions.	
2	The potential of the task is limited to engaging students in using a procedure that is either specifically called for or its use is evident based on prior instruction, experience, or placement of the task. • There is little ambiguity about what needs to be done and how to do it. • The task does not require students to make connections to the concepts or meaning underlying the procedure being used. Focus of the task appears to be on producing correct answers rather than developing understanding (e.g., applying a specific problem solving strategy, practicing a computational algorithm). OR The task does not require student to engage in cognitively challenging work; the task is easy to solve.	
1	The potential of the task is limited to engaging students in **memorizing or reproducing facts, rules, formulae, or definitions**. The task does not require students to make connections to the concepts or meaning that underlie the facts, rules, formulae, or definitions being memorized or reproduced. OR The task requires no academic activity. Representations include numbers and/or symbols, diagrams/pictures, use of written/verbal language, graphs, tables/charts, and concrete materials.	

<table>
<tr><th>Scale</th><th>Dimension 1: Academic Rigor</th><th>Rubric 2: Implementation of the Task</th></tr>
<tr><td>4</td><td colspan="2">Students engaged in exploring and understanding the nature of concepts, procedures, and/or relationships, such as:
• Using complex thinking (i.e., there is not a predictable, well rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example);
OR
• Procedures with connections: applying a broad general procedure that remains closely connected to concepts.
There is explicit evidence of students' reasoning and understanding in their written work. For example, students may have:
• solved a genuine, challenging problem for which students' reasoning is evident in their written work;
• provided an explanation for why formulas or procedures work;
• identified patterns and formed generalizations based on these patterns;
• made conjectures and supported conclusions with evidence;
• made explicit connections between representations, strategies, or concepts and procedures;
• followed a prescribed procedure in order to explain/illustrate a concept, process, or relationship.</td></tr>
<tr><td>3</td><td colspan="2">Students engaged in complex thinking or in creating meaning for concepts, procedures, and/or relationships. However, the implementation does not warrant a "4" because:
• there is no explicit evidence of students' reasoning and understanding;
• students engaged in academic concepts or procedures with connections, but the underlying content knowledge in the task was not appropriate for the specific group of students (i.e., too easy or too hard to sustain engagement with high-level cognitive demands);
• students identified patterns but did not make generalizations;
• students used multiple strategies or representations but connections between different strategies/representations were not explicitly evident;
• students made conjectures but did not provide evidence or explanations to support conclusions.</td></tr>
<tr><td>2</td><td colspan="2">Students engaged in using a procedure that was either specifically called for or its use was evident based on prior instruction, experience, or placement of the task.
• There was little ambiguity about what needed to be done and how to do it. Students did not make connections to the concepts or meaning underlying the procedure being used. Focus of the implementation appears to be on producing correct answers rather than developing understanding (e.g., applying a specific problem solving strategy, practicing a computational algorithm).
OR
• Student did not engage in cognitively challenging work; the task was easy to solve.</td></tr>
<tr><td>1</td><td colspan="2">Students engage in memorizing or reproducing facts, rules, formulae, or definitions. Students do not make connections to the concepts or meaning that underlies the facts, rules, formulae, or definitions being memorized or reproduced.
OR
Students did not engage in academic activity.</td></tr>
</table>

Scale	Dimension 1: Academic Rigor	Rubric 3: Rigor in Students' Responses to the Task
4	Students show written work and provide complete and thorough explanations of why their strategy, idea, or procedure is valid. Students explain why their strategy works and/or is appropriate for the problem by making connections to the underlying ideas (e.g., "I divided because we needed equal groups"). OR Student work displays use of more than one strategy or representation for solving the task, and provides a written explanation of how the different strategies/representations were used to solve the task.	
3	Students show written work and provide explanations BUT the explanations are incomplete or are procedural in nature. Students explain the steps of their work (e.g., what they did first, second, etc.) but do not explain why their strategy or procedure works and/or was appropriate for the problem; OR Student work displays use of more than one strategy or representation for solving the task.	
2	Students show written work for solving the task (e.g., the steps for a multiplication problem, finding an average, or solving an equation) with no written explanation; OR Student work displays use of only one strategy or representation for solving the task.	
1	Students provide brief or one-word answers (e.g., fill in blanks); OR Student's responses are not academic.	

Scale	Dimension 1: Academic Rigor	Rubric 4: Academic Rigor in Teacher's Expectations
4	**The majority of the teacher's expectations are for students to:** • use complex and non-algorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example); • explore and understand the nature of concepts, procedures, and/or relationships. [The expectations for academic content are stated explicitly in one of the sources indicated below.] **For example, the teacher may expect students to:** • solve a genuine, challenging problem; • develop an understanding for why formulas or procedures work; • identify patterns and form generalizations based on these patterns; • make conjectures and support conclusions with evidence; • make connections between representations, strategies, or concepts and procedures.	
3	At least some of the teacher's expectations are for students to engage in complex thinking or in understanding important concepts. However, the teacher's expectations do not warrant a "4" because: • the expectations are appropriate for a task that lacks the complexity to be a "4"; • the expectations do not reflect the potential of the task to elicit complex thinking (e.g., identifying patterns but not forming generalizations; using multiple strategies or representations without developing connections between them; providing shallow evidence or explanations to support conclusions); • the teacher expects complex thinking, but the expectations do not reflect the potential of the task.	
2	The teacher's expectations focus on skills that are germane to student learning, but these are not complex thinking skills (e.g., expecting use of a specific problem solving strategy, expecting short answers based on memorized facts, rules or formulas; expecting accuracy or correct application of procedures rather than on understanding concepts).	
1	The teacher's expectations do not focus on substantive content. The teacher's focus may be solely on activities or classroom procedures (e.g., following directions, producing neat work, or following norms for cooperative learning).	

Scale	Dimension 2: Clear Expectations	Rubric 1: Clarity and Detail of Expectations
4	The expectations for the quality of students' work are very clear and elaborated. Each dimension or criterion for the quality of students' work is clearly articulated. Additionally, varying degrees of success are clearly differentiated.	
3	The expectations for the quality of students' work are clear and somewhat elaborated. Levels of quality may be vaguely differentiated for each criterion (i.e., little information is provided for what distinguishes high, medium and low performance.)	
2	The expectations for the quality of student's work are broadly stated and unelaborated.	
1	The teacher's expectations for the quality of student's work are unclear OR the expectations for quality work are not shared with students	

Scale	Dimension 2: Clear Expectations	Rubric 2: Communications of Expectations
4	Teacher discusses the expectations or criteria for student work (e.g., scoring guide, rubric, etc.) with students in advance of their completing the assignment and models high-quality work.	
3	Teacher discusses the expectations or criteria for student work (e.g., scoring guide, rubric, etc.) with students in advance of their completing the assignment.	
2	Teacher provides a copy of the criteria for assessing student work (e.g., scoring guide, rubric, etc.) to students in advance of their completing the assignment.	
1	Teacher does not share the criteria for assessing students' work (e.g., scoring guide, rubric, etc.) with students in advance of their completing the assignment. (e.g., teacher may provide a copy of the scoring rubric to students when giving them their final grade.	

Scale	Dimension 3: Equitable teaching	Rubric 1: Participation Structures
4	**The teacher used this assignment as a vehicle to provide students with significant opportunities to participate in discussions.** • Collaboration* is a purposeful part of the assignment. • The collaborative work is well structured, and/or a strong routine is in place to support the collaboration. • There is accountability for the collaborative meaning making at the individual **and** group level.	
3	**The teacher used this assignment as a vehicle to provide students with adequate opportunities to participate in discussions.** • Collaboration* is a purposeful part of the assignment. • The collaborative work is adequately structured and/or an adequate routine is in place to support the collaboration. • There is accountability for the collaborative meaning making at the individual **or** group level.	
2	**The teacher provided students with minimal opportunities to participate in discussions.** • Collaboration* is not a purposeful part of the assignment.\ • The collaborative work has minimal or no structure. • There is minimal accountability for meaning making at the individual **or** group level.	
1	**The teacher provided students with no required opportunities to participate in discussions.** • Collaboration* is not a part of the assignment. • There is no accountability for meaning making at the individual **or** group level.	

*The act, process, or product of working together to create meaning.

<table>
<tr><th>Scale</th><th>Dimension 3: Equitable teaching</th><th>Rubric 2: Differentiation</th></tr>
<tr><td>4</td><td colspan="2">The teacher engages in equitable teaching by differentiating instruction to ensure the needs of all learners are met through multiple entry points, such as:
• The assignment is structured to allow for student choice, such as how they participate, present final work, and/or grouping.
• The assignment incorporates more than one type of thinking.
• The assignment incorporates varying levels of complexity to meet the needs of students at a range of proficiency levels.
• The instruction and assignment addresses learning through multiple modalities.</td></tr>
<tr><td>3</td><td colspan="2">The teacher engages in some equitable teaching by differentiating instruction to ensure the needs of all learners are met through multiple entry points, such as:
• The assignment may allow for student choice, such as how they participate, present final work, and/or grouping.
• The assignment incorporates more than one type of thinking.
• The assignment incorporates varying levels of complexity to meet the needs of students at a range of proficiency levels.
• The instruction and assignment addresses learning through at least two modalities.</td></tr>
<tr><td>2</td><td colspan="2">The teacher engages in little equitable teaching by differentiating instruction to ensure the needs of all learners are met through multiple entry points, such as:
• The assignment may not allow for student choice.
• The assignment is focused primarily on one type of thinking.
• The assignment incorporates more than one level of complexity to meet the needs of students at some proficiency levels.
• The instruction and assignment addresses learning through primarily one modality. (Other modalities may be minimal).</td></tr>
<tr><td>1</td><td colspan="2">The teacher engages in minimal or no equitable teaching by differentiating instruction to ensure the needs of all learners are met through multiple entry points, such as:
• The assignment is not structured to allow for student choice.
• The assignment is focused on one type of thinking.
• The assignment does not incorporate varying levels of complexity to meet the needs of students at different proficiency levels.
• The instruction and assignment addresses learning through one modality.</td></tr>
</table>

Scale	Dimension 3: Equitable teaching	Rubric 3: Academic Language
4	**The assignment well supports the development of academic language:** • The assignment includes ample academic language, incorporating common and content academic language *throughout* the assignment. • The assignment requires that students use academic language in written form and possibly verbally.	
3	**The assignment supports the development of academic language:** • The assignment includes some common and content academic language in the assignment. Academic language may not be well integrated across the assignment. • The assignment requires or encourages that students use academic language in written form and possibly verbally.	
2	**The assignment minimally supports academic language:** • The assignment includes minimal common and/or content academic language **and** relies heavily on one type of language in the assignment. Academic language may not be well integrated across the assignment. • The assignment does not encourage students to incorporate appropriate academic language in their responses.	
1	**The assignment minimally supports academic language:** • The assignment requires little/ no common or content academic language, relying on general language or numbers to engage students. Academic language is not integrated across the assignment. • The assignment does not encourage students to incorporate appropriate academic language in their responses.	

Scale	Dimension 3: Equitable teaching	Rubric 4: Relevance
4	**The assignment is relevant to students' lives** • The assignment exists within the context of an authentic, real-world problem that students are invited to relate to, **or** deals with content that is applicable to students' experiences or interest. • Relevance to students is an important facet of the assignment and integrated throughout the assignment. • The assignment encourages and builds on students' experiential knowledge.	
3	**Some aspects of the assignment are relevant to students' lives** • The assignment is introduced within the context of an authentic, real-world problem that students are invited to relate to **or** the introduction to the work deals with content that is applicable to students' experiences or interest. • Relevance to students is not integrated meaningfully throughout the lesson. • The assignment encourages students' experiential knowledge but does not build on it.	
2	**Some aspects of the assignment are somewhat relevant to students' lives** • The assignment is presented within the context of a problem that is minimally relevant to students' experiences or interest. • Relevance to students is not integrated throughout the lesson. • The assignment encourages students' experiential knowledge but does not build on it.	
1	**The assignment is not relevant to students' lives** • The assignment is not presented within the context of real-world problem that students are invited to relate to, **nor** does it deal with content that is applicable to students' experiences or interest. • Relevance to students is not an important facet of the assignment and is not embedded throughout the assignment. • The assignment does not encourage students' experiential knowledge.	