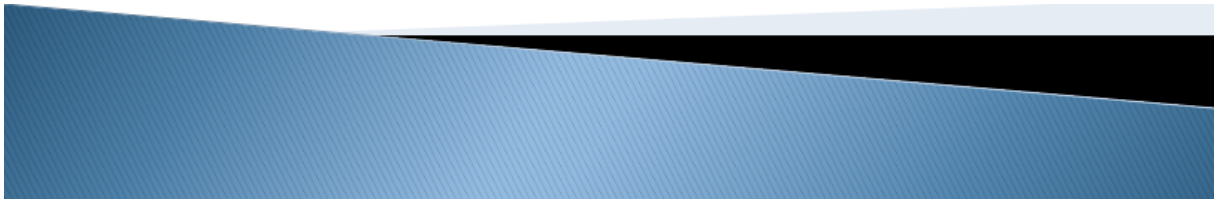


Summative
Report

March
2015

NEW VISIONS HUNTER COLLEGE URBAN TEACHER RESIDENCY

MEASURES OF SUCCESS



Kay Sloan
Juliane Blazeovski
Rockman et al

3925 Hagan St., Suite 301
Bloomington, IN 47401
812.333.8883

595 Market St., Suite 2570
San Francisco, CA 94105
415.544.0788

Table of Contents

Introduction	5
Methods	6
Data Sources and Samples	6
Data Analysis	10
Supplemental Study	11
SUMMARY OF FINDINGS	12
1. Synopsis of the UTR Model	15
2. Recruitment and Hiring	17
Who are the UTR residents?	18
Ethnic Diversity	18
Gender Representation	19
Credentials	19
Mentors' Observations on Recruitment and Selection	21
How many residents complete program requirements and obtain jobs?	21
Overall Completion Rates	21
Variations by Subject	23
Hiring in High-Need Schools	23
Hiring for Hard-to-Staff Subjects	24
Who are the UTR Mentors?	25
Mentors' Roles and Selection	25
Mentors' Background and Experience	25
Support, Satisfaction, and Return Rates among Mentors	27
3. Teacher Quality and Confidence	28
How does UTR define and gauge teacher quality?	28
How do residents perform on the teacher quality measures	31
Overall Ratings	31
Trends in Danielson Ratings	31
Differences by Content Area	31
Special Education Addendum	32
How do residents rate their own self-efficacy and confidence?	33
Overall Confidence, Efficacy, and Practice Ratings	33
Scale Averages and Differences by Subject and Setting	34
Self-Assessment and Danielson Ratings	35
4. Student Achievement: UTR vs. Non-UTR	37
What longitudinal trends emerged in the achievement data, Years 1–5?	38
Trends across Years, by Outcome Measure	38
Trends by Subject Area	38
Trends by Cohort	39
Did the Year 5, SY 2013–2014 results follow previous trends?	40
Course Performance Results, by Cohort	40
Regents Performance Results, by Cohort	44
UTR Graduates, Cohorts 1–4, Combined Results	46
Additional Special Education Analyses, Cohorts 1–4	47

5: Teacher Performance, Student Performance	49
Do the assessment tools form a coherent set of measures?	50
Links between Measures, by Cohort	50
Can residents' scores predict students' performance?	52
Positive Correlations, Strongest Links	47
Cohort 5	47
Can teachers' performance in their residency year predict their students' performance in their induction years?	55
Residency vs. Induction Year Performance	55
Composite Ratings	56
5. Retention and School Impact	58
Are the residents staying?	58
Comparisons	58
Retention Rates by Subject	60
Why do teachers stay, or leave?	60
Participation in Collaborative Activities and Leadership Opportunities	61
School Environment and Job Satisfaction	62
Changes by Project Year	63
Teaching as the Chosen Profession	64
Are schools with more UTR teachers different?	66
Job Satisfaction	66
Student Performance	66
CONCLUSIONS	67
References	69
Appendices	69

LIST OF TABLES

Table 1. Mentor and Resident Survey Responses by Cohort	9
Table 2. Sample Sizes, SY2013–2014, Course and Regents, Cohorts 1–4)	10
Table 3. Sample Sizes, SY2013–2014, Course and Regents, Cohort 5	10
Table 4. UTR Residents from Underrepresented Groups, by Cohort	19
Table 5. UTR Gender Representation, by Cohort	20
Table 6. Program Completion Figures, by Cohort	23
Table 7. Returning Mentors by Cohort	23
Table 8. Trends by Subject, Residents and Graduates Combined, Cohorts 1–5	40
Table 9. Course Performances, Students Taught by Cohort 5 Residents	42
Table 10. Course Performance, Cohort 4	42
Table 11. Course Performance, Cohort 3	44
Table 12. Course Performance, Cohort 2	44
Table 13. Course Performance, Cohort 1	45
Table 14. Regents Performance, Cohort 5	45
Table 15. Regents Performance, Cohort 4	45
Table 16. Regents Performance, Cohort 3	46
Table 17. Regents Performance, Cohort 2	47
Table 18. Course Performances, All Graduates (Cohorts 1–4)	47
Table 19. Regents Performances, All Graduates (Cohorts 1–4)	48

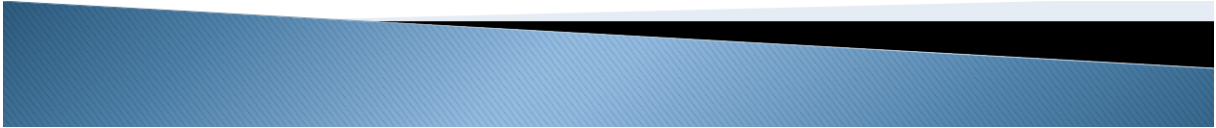
Table 20. Comparisons of Credits Earned and Regents Passed, Sped	49
Table 21. Correlations between Cohort 3 Resident Ratings	51
Table 22. Correlations between Cohort 4 Resident Ratings	51
Table 23. Correlations between Cohort 5 Resident Ratings	52
Table 24. Correlations between Cohort 5 Resident Ratings, Special Education	52
Table 25. Correlations between MASTER Resident Ratings	52

LIST OF FIGURES

Figure 1. Degrees Held by UTR Residents	20
Figure 2. Percentages of Career Changers & Recent Graduates	21
Figure 3. Career Changers' Years of Experience	21
Figure 4. Residency Completion and Attrition	23
Figure 5. Program Completion by Subject	24
Figure 6. Hiring Status—All Cohorts	25
Figure 7. Total Number of Teachers Hired in High-Need Schools	25
Figure 8. Induction Year Hiring by Subject	25
Figure 9. UTR Mentors' Years of Previous Teaching Experience	27
Figure 10. % of Mentors Entering Profession through Alt. Cert. Programs	28
Figure 11. Danielson Combined Ratings, Cohort 5	33
Figure 12. Sped Resident Ratings, Cohorts 4 and 5	33
Figure 13. Overall Scale Averages	36
Figure 14. Scale Averages on Danielson Items	36
Figure 15. Scale Averages by Teaching Hospital vs. Regular Host Schools	30
Figure 16. Scale Averages on Danielson Items	31
Figure 17. Residents' Preparation and Practice Scale Responses, by Cohort	37
Figure 18. UTR vs. Non-UTR, Residents and Graduates	40
Figure 19. UTR vs. Non-UTR, residents vs. Graduates	40
Figure 20. UTR higher: Regents Scores	40
Figure 21. UTR Higher: Course Grades	40
Figure 22. IEP vs. Non-IEP, Course Grades, Cohort 4	43
Figure 23. IEP vs. Non-IEP, Cohort 3	44
Figure 24. IEP vs. Non-IEP, Cohort 2	44
Figure 25. IEP vs. Non-IEP, Regents, Cohort 4	45
Figure 26. IEP vs. Non-IEP, Regents, Cohorts 1–4	49
Figure 27. Cohort 5 Correlations	53
Figure 28. Cohort 5 Correlations, Resident Ratings and Student Outcomes	55
Figure 29. Cohort 3 Residents: Course Grades and Regents across Years	56
Figure 30. Cohort 4 Residents: Course Grades and Regents across Years	56
Figure 31. Cohort 3: Regents Avg. Score by Resident Composite Rating	57
Figure 32. Cohort 4: Regents Avg. Score by Resident Composite Rating	57
Figure 33. Cohort 5: Course Grade by Resident Composite Rating	58
Figure 34. Retention Rates among UTR Teachers, 2010–2014	53
Figure 35. UTR vs. City-Wide Retention	60
Figure 36. UTR Retention Rates by Subject	60
Figure 37. Residents' Participation in Collaborative Activities	62
Figure 38. Mentors' Participation in Collaborative Activities	63
Figure 39. School Environment Scale Responses	64

Figure 40. School Environment Scale Responses by Year (Residents)	65
Figure 41. School Environment Scale responses by Year (Mentors)	65
Figure 42. UTR Teachers Thoughts about Their Chosen Profession	66
Figure 43. % of Teachers “Satisfied” on School Quality Survey	67
Figure 44. Average 2014 “Achievement” and “Progress” Scores	68
Figure 45. 4-Year Graduate Rate for 2013	68
Figure 46. Average 2014 SAT Scores	68

NEW VISIONS HUNTER COLLEGE URBAN TEACHER RESIDENCY MEASURES OF SUCCESS



Introduction

In 2009, New Visions for Public Schools, Hunter College, and the New York City Department of Education (NYC DOE) launched the Urban Teacher Residency (UTR), a U.S. Department of Education (USDE) funded effort with two ambitious goals for the city's high-need secondary schools: raising achievement and lowering attrition. Like other current reform efforts, UTR focused on teacher quality—on not just slowing the stream of departing teachers and getting more in the pipeline, but on making sure that those new teachers were well prepared. Recent studies, and a certain amount of buzz in the education media, had identified teacher quality as the single most important factor in a student's academic success, at least during the school day.¹ Educators and reformers couldn't easily take on all the social inequities and out-of-school issues that affect students' success, but they could rethink how teachers were recruited and prepared, and how they were supported, especially during their critical first years in the classroom.

Efforts to address teacher shortages in urban schools by refocusing recruitment were not entirely new. For two decades, Teach For America had focused on attracting the best and the brightest to a profession they might not otherwise consider. To attract a broader pool of candidates, Hunter College, like many other institutions, had added alternative paths to certification to its traditional teacher preparation programs. In 2000, the NYC DOE had introduced its Teaching Fellows program, which had recruited hundreds of aspiring teachers to meet the city's growing needs in hard-to-staff subjects like English, mathematics, science, and special education.

What was relatively new, and a key part of UTR, was the focus on recruiting and training candidates for the complex urban environments where they would work. The Teaching Fellows program did this as well, but UTR, like other programs in the national Urban Teacher Residency United (UTRU) network, extended the training period to a full year. That year in the classroom is preceded by summer preparation to familiarize UTR residents with school curriculum, culture, and policies, which residents subsequently learn about first-hand as teachers responsible for a single "focus" class in a host NYC DOE school, under the guidance of a trained mentor, with ongoing coaching and support from New Visions program staff. As they gain clinical experience they also take graduate coursework at Hunter College that combines content, pedagogy, and practice and culminates in a Master's degree and New York State certification for grades 7–12.

¹ See, for example, the RAND Corporation's "Teachers Matter: Understanding Teachers' Impact on Student Achievement," available at http://www.rand.org/pubs/corporate_pubs/CP693z1-2012-09.html.

Other features, described in more detail in Section 1 of this report, set UTR apart not just from other alternative certification programs but also other residency programs. From the start, UTR's focus is on improving student performance. That is the generally goal of all teaching and all teacher preparation, but it is woven into every aspect of the UTR model in very specific ways. Unit design, lesson planning, instruction—all are data-driven, evidence based. Residents are held accountable, based on their performance on a suite of assessment tools developed by New Visions—especially the Defense of Learning tool, which requires residents to demonstrate that they can improve student learning in order to become UTR graduates.

These tools more than met the call from the National Council for the Accreditation of Teacher Education (NCATE) for ensuring teacher quality by “continuously judging” the progress of teacher candidates.² They also anchored the coaching and monitoring process and turned the judging into an opportunity for shared learning and reflection. Supplying a different kind of readiness, they prepared residents for the teacher evaluations that would define the educational landscape they were about to enter, where quality and accountability were inextricably linked.

This report shares findings from five years of the UTR external evaluation, conducted by Rockman et al. The report focuses largely on summative findings, spanning 2009–2014, but also includes 2013–2014 results for Cohort 5, not previously shared. Section 1 describes the UTR model—its goals, vision of effective teaching, and underlying assumptions. Section 2 looks at who UTR residents are: their backgrounds and experience, and their success in completing program requirements and obtaining jobs in high-need schools. Section 2 also looks at UTR mentors, beginning a thread that runs throughout the report about their professional growth as mentors, teachers, and school leaders. Section 3 examines how the tools in the teacher quality assessment suite evolved, how residents performed, and how their performance lines up with their own reports of self-efficacy and confidence.

Section 4 turns to the project's broader goals: first, student achievement, or how UTR residents' and graduates' students perform compared to their peers, on course grades and annual New York State Regents exams. Section 5 circles back to the teacher quality tools, to explore whether there is a link between students' performance and residents' performance. Section 6 asks the longer-term questions: are UTR teachers staying, and what impact does the UTR project and its new corps of teachers have on factors that affect retention: collegiality, collaboration, support from school leaders, and the opportunity to grow? A conclusions section looks at UTR's accomplishments against the goals laid out in the logic model and the continued refinement and durability of the model.

Methods

Throughout the project we have used a mixed-methods design with both quantitative and qualitative components, and data collection tools that were refined, as needed, as the project evolved. For the summative study, we aggregated data across years to explore longitudinal impact, looking for trends and cohort or subject-area differences. Where possible, we compared quantitative results to feedback from residents, mentors, and school leaders and program partners. The summative study also provided an opportunity to explore the coherence and predictive value of internal suite of assessment tools. To further explore cumulative impact, we also chose a sample

² “Transforming Teacher Education through Clinical Practice: A National Strategy to Prepare Effective Teachers,” p. 5. The National Council for Accreditation of Teacher Education, 2010. Available at <http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OogPk%3D&tabid=715>. See also “Gathering Feedback for Teaching: Combining High-Quality Observations with Study Surveys and Achievement Gains.” Bill & Melinda Gates Foundation, January 2012. Available online at http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf.

of schools that hosted or hired multiple residents, and examined teacher confidence, practice, and satisfaction data, and at student achievement data in sites with a concentration of UTR-trained teachers.

DATA SOURCES AND SAMPLES

Section 1. The UTR Model

The description of the UTR model and short- and long-term goals is based on the logic model that guided implementation and multiple documents authored by partners over the project's five years, including New Visions' *Site Director Handbook* and *Vision for Effective Teaching*.

Section 2. Recruitment and Hiring

Data on recruitment, program completion, and hiring came from Hunter College's Office of Residency Programs and New Visions' Teacher Certification Data Team Warehouse. Additional feedback about recruitment activities, results, and changes over the course of the project came from interviews and focus groups with partners and participants. The sample included all residents, for all five cohorts, for whom we had data—generally as of September 2014, updated in March 2015.

Section 3. Teacher Quality and Confidence

Resident Performance Measures

Scores generated by New Visions' suite of teacher assessment tools, provided by the New Visions UTR team, or housed in their Process Monitoring System database or Rockman software database, have been the source of the residents' performance data. These data include scores on the Defense of Learning (DoL) and from the Danielson, lesson and unit design, professionalism, and using assessment rubrics. We looked at percentages of residents meeting project benchmarks for each assessment tool, differences by subject area, rater, and year, and at the performance arcs or successive scores. Numbers of residents in the samples varied based on when the tools were introduced (e.g., the DoL was used all five years; the Danielson framework, in Years 3–5) and how frequently residents were rated and scores recorded.

Teacher Surveys

Rockman surveyed UTR residents and mentors each year of the project, using the first year to pilot items and Year 2 and Year 5 responses to construct scales from factor analyses. The online surveys explored confidence or sense of preparation, practice, and school climate and support. Certain items were mapped to the competencies in the Danielson rubric—differentiating instruction, aligning objectives to standards, engaging students, handling classroom management issues, assessing learning—allowing us to explore how residents' self-reports compared to observers' ratings. The surveys also included items drawn from the New York City School Surveys, which allowed us to compare UTR and school-wide responses for the UTR concentration schools. Based on factor analyses and reviews of how survey items clustered or related and accounted for variance in the data, we created three sets of scales—overall scales for confidence, practice, and overall efficacy; a set of scales related to the Danielson domains; and a set for school climate. (See Appendix A, p. 62 for survey items included in each scale.)

A total of 114 residents and 88 mentors made up the full survey sample. Table 1 shows the responses and response rates, by cohort or year. Overall response rates were high—77% for mentors and 96% for residents. (During the project's first three years, we also invited graduates to complete surveys. Due to low response rates, we did not survey graduates during the last two years and have not included graduate responses in this report.)

Table 1. Mentor and Resident Survey Responses by Cohort

	MENTORS		RESIDENTS	
	Number	Percentage	Number	Percentage
Cohort 2 (2010–2011, N=35)	12	34%	34	97%
Cohort 3 (2011–2012, N=29)	25	86%	29	100%
Cohort 4 (2012–2013 (N=31)	27	87%	27	87%
Cohort 5 (2013–2014, N=24)	24	100%	24	100%
*Total (N=119)	88	77%	114	96%

*N's refer to numbers of program completers. Discussions do not include Cohort 1 because Year 1 survey items differed from those used in Years 2–5.

Focus Groups

An annual part of the UTR evaluation has been resident and mentor focus groups. Structured protocols included questions about training, use of Danielson, lesson and unit design, and professionalism rubrics; and the interplay of program components—the Hunter coursework, New Visions' support, and residents' classroom experiences. Returning mentors also discussed changes in training, support, and program implementation. Sample sizes or attendees varied, but averaged about 75% of the mentor and resident groups. For Cohort 5 (SY 2013–2014) we conducted a combined ELA and special education focus group with approximately 20 mentors.

Section 4. Student Achievement

For all five years of the project, we worked with New Visions to assemble achievement data and select matched comparisons schools. The process began with the identification of all early career teachers in the New Visions PSO; for residents, comparison teachers were those with 0 to 1 year of teaching experience; for graduates, comparison groups matched to graduates' experience. For each group, we selected all potential matches, checking course names and codes to make sure that teachers in our matched sets taught the same courses. In a parallel activity, we used a Mahalanobis distances (MD) measure, which identifies “nearest neighbors,” or schools with adjacent scores based on calculations of multiple covariates, which in this case included Regents percent passing for the designated courses (school-level aggregate for all tested students); aggregate incoming student 8th grade state test scores for mathematics and ELA; attendance rate; graduation rate; total enrollment; and percentages of students who are Black or Hispanic, female, eligible for free or reduced lunch, and have limited English proficiency or disabilities.

Our final matched sets, which excluded transfer high schools, charter schools, and schools with missing values, were created through an iterative process that included, for each subject, clustering schools by MD scores, confirming course equivalence, confirming names of teachers of record, matching teachers by years of experience, and determining that there were enough students with Regents scores and grades for reasonable comparisons.

Sample sizes varied from year to year, depending on how successful we were in finding a matching teacher and whether their students had Regents scores and grades. For example, despite Cohort 5 including ELA teachers, we had virtually no Regents scores associated with the English courses taught by the residents, due in part to the fact that the ELA Regents is not typically taken in students' freshman year. Also, for Cohort 5, we have subject areas other than ELA because some special education teachers in the cohort were associated with courses outside of ELA.

Tables 2 and 3 show the sample sizes (numbers of students and teachers) for our Regents and course grade analyses for SY 2013–2014, for Cohorts 1–4 graduates and Cohort 5 residents; as in previous years, sample sizes varied across analyses due to missing data. (Previous annual reports included yearly sample sizes.)

Table 2. Sample Sizes, SY 2013–2014, Course Performance and Regents, GRADUATES (Cohorts 1–4)

Course Grade Analyses			Regents Analyses		
Subject Area	UTR	Non-UTR	Subject Area	UTR	Non-UTR
	*N Students & Teachers	N Students & Teachers		N Students & Teachers	N Students & Teachers
English	740 (12)	992 (12)	English	126 (8)	146 (8)
Algebra 2 / Trig	111 (3)	42 (3)	English (CC)	87 (3)	55 (3)
Geometry	311 (7)	412 (8)	Integrated Algebra	184 (6)	243 (7)
Integrated Algebra	406 (10)	386 (11)	Integrated Algebra (CC)	172 (6)	187 (5)
Biology	77 (1)	24 (1)	Algebra 2 / Trig	29 (1)	11 (1)
Chemistry	598 (7)	395 (7)	Geometry	144 (5)	166 (3)
Earth / Environ. Science	235 (3)	247 (3)	Chemistry	814 (7)	499 (7)
Living Environment	647 (13)	722 (13)	Earth / Environ. Science	243 (1)	130 (1)
Physics	21 (1)	79 (1)	Living Environment	707 (12)	1031 (12)
Global Studies/ History	50 (2)	81 (2)	Global Studies / History	23 (2)	48 (2)

Table 3. Sample Sizes, SY 2013–2014, Course Performances and Regents, RESIDENTS (Cohort 5)

Course Grade Analyses			Regents Analyses		
Subject Area	UTR	Non-UTR	Subject Area	UTR	Non-UTR
	*N Students & Teachers	N Students & Teachers		N Students & Teachers	N Students & Teachers
English	283 (15)	828 (15)	Integrated Algebra	23 (2)	129 (2)
Geometry	24 (1)	44 (1)	Living Environment	13 (1)	39 (1)
Integrated Algebra	24 (2)	129 (2)			
Global Studies / History	12 (1)	58 (1)			

*In Tables 2 and 3, the first figure indicates the number of students included in the analyses; the figure in parentheses, the number of teachers.

We derived our sample for the examination of performance in UTR concentration schools by using the list of schools included in the Regents/course grade analyses, then merging that list with the SAT file and the UTR concentration index. The SAT data came from the file "2014SATWebsite10214 from the NYC DOE website; it includes SAT scores for "college bound" students at each school (as reported by The College Board/ SAT).

The UTR concentration index was based upon the "actively teaching" list from the Teacher Data Warehouse. Schools not on the list were assigned a concentration of "0"; schools with 1 to 2 UTR teachers, a concentration of 1=low; and schools with 3 or more UTR teachers, a concentration of 2=high.

Section 5. Teacher Quality and Student Performance

We used the teacher quality data generated by the assessment tools and student achievement data (Years 3–5) to

examine links between the two. We first examined correlations between measures, those between teacher quality measures and Regents scores and grades. To further explore the predictive value of the tools, we created, where possible, a composite score for residents based on the suite of tools—the Defense of Learning score, the Danielson individual and average score, and the Lesson and Unit Design, Professionalism, Using Assessment rubric scores.

The correlation analyses between measures generally included all those completing their residency year. The analyses examining links between resident and student performance included only those residents and graduates for whom we had Regents scores and/or grades, generally excluding, for example, special education residents. As possible, we used attendance data from either the New Visions data sources or the NYC DOE public data.

Section 6. Retention and School Change

Retention data for UTR teachers came from the New Visions Teacher Data Warehouse. We also drew on publicly available data from published articles or reports, noted in footnotes and references. In exploring support and satisfaction data between UTR concentration schools and comparable schools, we used the concentration figures from the Teacher Data Warehouse, and from the publicly available School Survey data from the NYC DOE website.

Interviews and with Hunter College, New Visions, and School Principals

We interviewed three principals in 2013-2014, discussing UTR’s implementation and evolution and differences between UTR-trained teachers and other candidates or hirees, and the impact of having a corps of UTR-trained teachers. During the planning and data assembling for the summative study, we talked with New Visions and Hunter partners, both formally and informally.

DATA ANALYSIS

Survey Data

We used basic descriptives and frequencies to analyze survey data, disaggregating data by subject (ELA, math, science, and special education) and by respondent group (residents and mentors), and comparing responses across years. Based on survey data and clusters of items related to classroom management, using assessment, and establishing a culture of learning, we compared residents’ self-reports of preparation and confidence to their Danielson scores. Using scales from the factor analyses, we examined data sets to see how responses varied by role, cohort, subject, and school type (e.g., teaching hospital vs. non-teaching hospital, UTR concentration level).

In analyzing answers to open-ended survey questions, we first reviewed responses to find emerging or recurrent themes, then coded responses as appropriate.

Achievement Data

We analyzed student performance data in multiple ways, examining:

- SY 2013-2014 Regents scores and grades, for all 5 cohorts (UTR residents and graduates), UTR vs. matched comparison groups;
- SY 2013-2014 school-level Regents performance between schools with a concentration of residents, graduates, and mentors, vs. a matched set of schools that have not taken part in UTR (or MASTER).
- Longitudinal review of all previous Regents data and grades, to explore trends by subject
- Other comparisons—e.g., student performance in UTR high vs. low concentration schools and comparison schools from the Regents/course grade analyses; UTR graduates’ students’ performance over time to see

how graduates' results changed in successive years.

We made statistical comparisons between course term grades and Regents scores from students taught by UTR-trained teachers and those from students taught by non-UTR teachers, for a set of similar courses. For example, for Science, we selected students taking the core Biology courses or other courses that are meant to prepare a student for the Regents Living Environment exam, avoiding the selection of honors, Advanced Placement, independent study, and credit recovery courses. We used t-tests to explore differences between the grades of the two student groups, and analysis of covariance (ANCOVA) to control for covariates of prior achievement (8th grade test scores in reading and math), gender, special education status, and ELL status. We also examined bivariate Pearson correlation coefficients between course grades and Regents exam scores.

Special Education

We continued to use the strategies similar to those from previous years to gauge the impact of special education residents, checking for significant interactions between IEP status and UTR-status to determine if the effect of having a UTR teacher was different for students with and without an IEP, based on course and Regents performance, looking at differences in grades between students with IEPs and their peers. We also looked at other available school data to see, for example, whether attendance rates vary.

Correlation Analyses

To examine links between the tools in the assessment suite, and between residents' and their students' performance, we continued the process used for Cohorts 3 and 4, creating a metric for residents' performance using practice ratings (based on the Danielson ratings, lesson and unit design, professionalism, and DoL), and running correlational analyses with Regents' scores and grades. We also formed composite scores by transforming individual scores into a percentage, based on the highest possible scores or points, then taking an average of those percentages to create the composite. Where possible, we also divided the composite into three relatively equal low/medium/high groups.

Supplemental Study

With a new cohort every year, five years is not enough time to fully explore the full impact of the UTR project. Over the next three years, we will continue to work with partners on a supplemental study to assess the longitudinal impact of the UTR project. This study will enable us to continue to see how the UTR model compares to other teacher preparation efforts; compare key metrics, such as retention figures; identify components that are most closely linked to student achievement; and explore other facets of the model, such as preparation and support for special education teachers, in more depth.

SUMMARY OF FINDINGS

Selectivity, diversity, hiring and retention rates over 90%, strong benchmark performance, solid student performance compared to peers—by multiple measures, UTR has been a success. The program has produced skilled teachers with a commitment to work in urban schools, and placed them in some of the neediest of those schools to teach the subjects hardest hit by attrition. With guidance and support from peers, mentors, school leaders, Hunter faculty members, and New Visions professional staff and coaches, cohort after cohort of UTR residents have—in 14 months—developed into strong teachers who have demonstrated that they can improve achievement. Mentors and the larger school communities have likewise strengthened their own skills through a shared commitment to accountable, reflective practice. The residents have completed program and graduate school requirements, gained confidence in their abilities, and obtained jobs, often in the schools in which they were residents or where other graduates of the program also teach. Most important, they have stayed.

Five years of external evaluation results and summative, longitudinal analyses show that:

- **UTR's selective recruitment and screening process yielded five diverse cohorts of residents committed to working in urban schools.** Admissions data show that the UTR project is highly selective—with an acceptance rate of 10%—and its cohorts of recruits highly diverse. Residents bring a range of backgrounds, credentials, and experiences: Over a fifth of the UTR enrollees hold advanced degrees, and over two-thirds are career changers. Close to half are people of color, and 42% of the program completers are from populations who are traditionally underrepresented in the teaching profession but make up approximately two-thirds of the NYC DOE student populations.
- **Strong program completion rates provide additional confirmation of an effective recruitment process and effective coaching, instruction, and support.** The overall figures of those completing their residency who also completed all program requirements—including the Master's degrees and certification requirements—currently stand at 89%, but will likely tick up as some of the remaining Cohort 5 residents earn their degrees. As a result of UTR's intervention plans for residents who need additional support or for whom UTR was not the best fit, overall, relatively few residents—17%—withdrew or were counseled out of the program.
- **Overall placement rates of 96% and rates of 86% in high-need NYC DOE schools confirm that UTR is meeting critical staffing needs.** UTR recruits candidates with a commitment to teaching in high-need urban schools, and during a full year of clinical residency grooms them to work in those complex environments with a diverse population of students. Overall, 86% of the 135 UTR-trained teachers found work in the kinds of high-need schools they apprenticed in and were trained for: 67% in a New Visions network school; 19%, in other NYC DOE high-need schools; and 2%, in high-needs schools outside the city. *(NOTE: The figures reported here are based on hiring data from Fall 2014, through the fifth cohort of residents.)*
- **UTR's clinically rich practice has given residents the skills and knowledge to improve teaching and learning and a sense of confidence and self-efficacy.** To graduate from the UTR program means to have demonstrated the ability to design, deliver, assess—and improve student learning. Residents have consistently met program benchmarks. Even as novice teachers, they have also reported high levels of confidence in their ability to engage students, teach their subject matter, and use data to inform their instruction.

- **Performance results, especially high-stakes test scores, show that the focus on student achievement and assessment has paid off.** UTR participants have been successful in improving academic achievement, often to a greater degree than other early career teachers. A longitudinal look at student achievement data shows UTR residents' and graduates' students performing as well as or better than their peers, especially on Regents exams. In almost three-fourths of the comparisons (16 out of 22) with statistically significant differences, students taught by UTR-trained teachers posted higher Regents scores and grades, most often in Living Environment and ELA, and also in Integrated Algebra and Chemistry. Achievement data also suggest that having a UTR-trained teacher can narrow the performance gap between students with IEPs and their peers, especially in Living Environment and English classes.
- **Links between the measures on which residents are assessed and the higher academic performance of their students provide additional evidence that UTR's focus on assessment works.** Correlation analyses indicate that a number of the teacher quality measures—especially those related to classroom culture—are linked to student performance; the Defense of Learning also appears to be a fairly good predictor of students' performance. There are also some indications that higher-performing residents are more effective teachers. While it is not clear that higher performing residents get progressively more effective, there are hints that teachers who did not have a strong effect on student performance during their residency year have a greater impact as they mature.
- **The teacher quality assessment tools have additional benefits for the project and the field.** As the UTR project got underway, NCATE urged continuous evaluation of teacher candidates, and the Gates MET project urged multiple measures. The UTR tools do both, and our analyses of correlations between them suggest not only links between resident and student performance but also a coherent set of measures. Perhaps more important, mentors and residents say that using the tools made coaching conversations productive rather than punitive, helped identify areas for improvement, and prompted self-reflection. The experience helped prepare residents—and often mentors—not just for future classrooms and employment but also for future evaluations.
- **As it trains residents, UTR also builds capacity in mentors and in schools.** UTR has always been as much about developing mentors—as tutors for novice teachers and school leaders—and creating a teaching corps and constellation of schools committed to accountable practice. Of the 81 schools that have hired UTR residents, 21% have previously hosted residents and 25% have hired three or more. Survey feedback from mentors also suggests positive changes over five years—a better school climate, more involvement in collaborative school decisions, and more professional engagement.
- **We are beginning to see signs of higher job satisfaction in schools with higher UTR concentrations.** Publicly available school environment and quality data and teachers' ratings for instructional core items, systems for improvement, and school culture, suggest that schools with the highest UTR concentrations have the most satisfied teachers. Our survey data also indicates that while apprenticing in a teaching hospital, with an onsite coordinator and more resident-mentor pairs did not seem to make a difference in residents' confidence, it did result in higher mentor confidence. The differences are small, but the trends promising.
- **Retention among UTR-trained teachers is impressively high, exceeding city rates after four years by substantial margins.** One of the main goals of the UTR project was to stem the flow of teachers from high-need secondary schools. The 135 teachers trained thus far may not have reversed the current, but with a

retention rate of 93% after four years—almost 20 percentage points higher than overall NYC DOE rates—UTR appears to be a highly effective model for engineering change, not just in retention but the in factors that support it. This is in the face of challenges that only seem to be growing, and the next five years will confirm whether the model is as durable as the issues it was designed to address.

Synopsis of the UTR Model

1

Among the multiple, non-traditional pathways to certification, many districts—especially urban districts faced with declines in both achievement and retention—have increasingly turned to residency programs that give teachers more clinical practice, more time to develop their skills alongside effective teachers, and more time to prepare for the demands of urban classrooms. What sets the New Visions–Hunter UTR model apart from other efforts is how it defines “effective” and “well-prepared.”

One of the first inputs listed in the UTR logic model is the creation of the project partnership, which included the NYC DOE, the country’s largest school district; Hunter College, one of the city’s most highly regarded institutions, which has long supplied many of its teachers; and New Visions for Public Schools, one of several support organizations established by the NYC DOE to provide leadership and guidance for schools, around 80 of which, serving close to 50,000 students, are in the New Visions network. Together, they developed the project activities laid out in the logic model—recruitment plans, academic and clinical training; ongoing coaching, support, and monitoring—that would lead to shorter-term goals of graduating confident, well-prepared teachers and placing them in high-need schools, and the longer-term goals of raising achievement and retention.

Key Features of the Model

- Strong partnerships
- Selective recruitment
- Accountable practice linking resident performance to student achievement
- Ongoing coaching and monitoring to prepare residents—and mentors—for data-driven practice, collaborative inquiry, and leadership roles
- Induction year support

The UTR model rests on the assumption that these classrooms are complex working environments that require teachers to continually make—and just as often rethink—decisions about classroom management, curriculum, instruction, assessment, and ways to ensure that diverse populations of students with manifold needs achieve academic success.

From the summer preparation that familiarizes residents with students’ needs, curriculum, and assessment, through the induction support provided 14 months later when residents enter the classroom as teachers of record, partners agreed that each decision the teacher makes should be informed by ongoing feedback from students. A New Visions template for effective teaching portrays the process as an ongoing inquiry cycle through which teachers—as designers, instructors, and assessors—are constantly assessing needs, framing and delivering lessons, and checking to confirm that students are learning.

This is not only how UTR defines effective teaching but also how it ensures that residents achieve it. In the UTR model, it is not only students but also residents whose needs, performance, and progress are steadily assessed. Project partners and other stakeholders—mentors, New Visions program officers and coaches, Hunter faculty, school administrators—all monitor and guide residents, through coaching, reflection, classroom visits, and use of the suite of assessment tools.

The goal is for mentors and schools themselves, as well as residents, to develop capacity for improving teaching and learning. As a reference point for coaching conversations that prompt self-assessment and reflection, the assessment tools built mentors' skills and capacity to assume leadership roles in their schools. As UTR matured, some host sites became teaching hospitals, where a Site Director took over support for multiple mentor-resident pairs and the school assumed responsibilities for teacher development.

A UTR matured, UTR set benchmarks, and fairly high bars, along the way, for resident, mentor, student, and program performance. This report charts the progress and overall success of the effort.

Recruitment and Hiring

2

Although the hallmark of the UTR model is the careful preparation and placement of teachers with a commitment to urban schools and accountable practice, it all starts with the equally careful recruitment and selection of candidates who can take on that role.

For aspiring candidates, the first step is admission to the NYC Teaching Fellows program, a not insignificant hurdle given that only about a fifth of the applicants to the Teaching Fellows' program are accepted.³ Over the project's five years, and through a steadily refined process, UTR enrolled 165 residents, for an overall acceptance rate 10%, which edged above the UTRU rate of 11%.⁴

In the various recruitment activities, screenings, and interviews, UTR partners were looking for candidates with a strong academic background to help close achievement gaps—and prepare students for new standards and more rigorous assessments—but also the flexibility and cultural sensitivity to serve diverse student populations with wide-ranging learning needs. Research on the dynamics between urban adolescents and teachers also suggests that role models are an important consideration in staffing urban schools, and engaging students who may not only lack role models but also face stereotype threat.⁵ It goes without saying that aspiring teachers also needed a solid commitment and a certain amount of tenacity, commonly referred to as the “grit” to face the demands of urban schools.

- UTR is a highly selective program that successfully attracts, trains, supports, and places a diverse pool of teachers with a commitment to urban schools.
- Admissions figures show an acceptance rate of 10%.
- 22% of the residents hold advanced degrees; 69% are career changers.
- Half of the residents enrolling in the program are people of color.
- 42% of program completers are from under-represented groups.
- 83% of the enrollees completed their residency; 89% of those have completed all program requirements.
- 96% of the 135 new teachers trained over 5 years obtained jobs, 86% in NYC DOE high-need schools.
- Of the 81 hiring schools, 25% have hosted residents; 21% have hired 3 or more.

³ See NYC Teaching Fellows' statistics at https://www.nycteachingfellows.org/about/program_statistics.asp. The screening process for the program was created by The New Teacher Project.

⁴ See http://www.utrunitied.org/EE_assets/docs/UTRU_Network_Trends_Report_2014.pdf. UTR rates based on reports to UTRU.

⁵ 2014, New York City Independent Budget Office, “A Statistical Portrait of New York City’s Public School Teachers.” Available at: <http://www.ibo.nyc.ny.us/iboreports/2014teacherdemographics.pdf>.

Program completion and placement rates discussed in this section indicate that UTR was both selective and effective in choosing residents who would successfully complete the program and secure positions in these high-need schools. The section also discusses mentors' selection and background, and the importance of developing teacher leaders and building capacity school-wide. New Visions professional staff members were looking for mentors who would excel at coaching and providing feedback, but also teachers who would reflect on their own practice, model collegiality and leadership, and help residents grow as part of a school community.

Who are the UTR residents?

Ethnic Diversity

Demographic make-up varied by cohort, but figures show that, overall, just under half of the UTR residents were people of color. This percentage is slightly lower than the 2014 NYC Teaching Fellows rate of 55%, but exceeds the UTRU average by 15 percentage points (49% vs. 34%), and, based on 2014 NYC Independent Budget Office data, overall city-wide percentages.⁶

Federal grant requirements also stress the importance of including populations traditionally underrepresented in recruitment or applicant pools and in the profession as a whole. UTR program data indicates that, overall, approximately a third (34%) of the recruits who enrolled in the UTR program were part of underrepresented groups (Black or Hispanic). (See Table 4.) According to 2011–2012 IBO figures, UTR percentages are slightly higher than city-wide percentages of 31.2%, though a little lower than representation in the city's high-poverty high schools, where figures climb to 39.1%, and to the overall citywide rate of 48.9% (pp. 4 and 6). Forty-five percent of the 2014 Teaching Fellows numbers are from traditionally underrepresented groups.

In addition to comparing percentages of recruits or enrollees across cohorts to city averages, we also looked at the percentages of underrepresented groups among the UTR program completers. Percentages began to edge up with Cohort 3, to levels equal to the percentages of underrepresented groups in the city's high-poverty schools (42%) and representation among the 2014 cohort of Teaching Fellows.

Table 4. UTR Residents from Underrepresented Groups, by Cohort

	Number Enrolled	Percentage of Enrollees	Number of Program Completers	Percentage of Program Completers
Cohort 1	5	21%	5	31%
Cohort 2	9	23%	9	26%
Cohort 3	13	33%	13	48%
Cohort 4	17	43%	17	55%
Cohort 5	14	40%	12	50%
TOTAL/Average	58	32%	56	42%

Source: New Visions Data Warehouse

*Data for Cohorts 4 and 5 include program completers and those continuing to actively work toward degree requirements.

⁶ See 2014, IBO, "A Statistical Portrait of New York City's Public School Teachers." Available at: <http://www.ibo.nyc.ny.us/iboreports/2014teacherdemographics.pdf>. For NYC Teaching Fellows' statistics see https://www.nycteachingfellows.org/about/program_statistics.asp. for UTRU figures see: http://www.utrunited.org/EE_assets/docs/UTRU_Network_Trends_Report_2014.pdf

Gender Representation

Breakdowns by gender show that among UTR residents, females outnumber males, as they do in the city's and in most teaching corps. Margins were about ten percentage points wider for UTR, at 68% female and 32%, male, compared to overall citywide percentages, though the numbers of female teachers tick up in higher-poverty schools, according to the IBO stats. The UTR figures varied over cohorts, with higher percentages of females in Cohorts 1 and 5, and lower figures, by around 15 percentage points, for Cohorts 2–4, possibly due to the fact that both math and science were included in those years. (See Table 5.)

Table 5. UTR Gender Representation, by Cohort

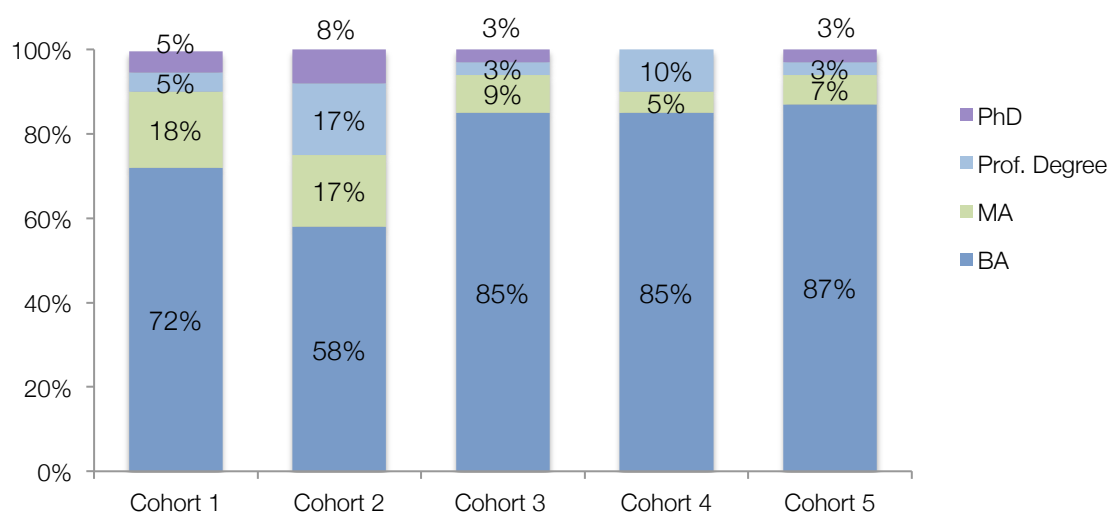
Gender	Cohort 1		Cohort 2		Cohort 3		Cohort 4		Cohort 5		Total (N=179)	
	N	%	N	%	N	%	N	%	N	%	N	%
Female	19	79%	25	63%	26	65%	26	65%	25	71%	121	68%
Male	5	21%	15	38%	14	35%	14	35%	10	29%	58	32%

Source: New Visions Data Warehouse

Credentials

Efforts to attract more teachers to high-need urban schools came about not just because of shortages and the assignment of too many secondary teachers to courses outside their subject areas, but also because schools needed to boost achievement in core academic subjects to ensure that students were sufficiently prepared for college or other post-secondary pursuits. Like other alternative certification programs, UTR looked for candidates with strong academic backgrounds in high-need academic subjects. The majority of residents selected for each UTR cohort had bachelor's degrees, but between 42% (Cohort 2) and 13% (Cohort 5) also had advanced degrees (see Figure 1). Some of the variations across years may be due to the subjects included—for example, Year 2 added mathematics, a subject where both subjects where undergraduate degree holders may pursue advanced degrees.

Figure 1. Degrees Held by UTR Residents

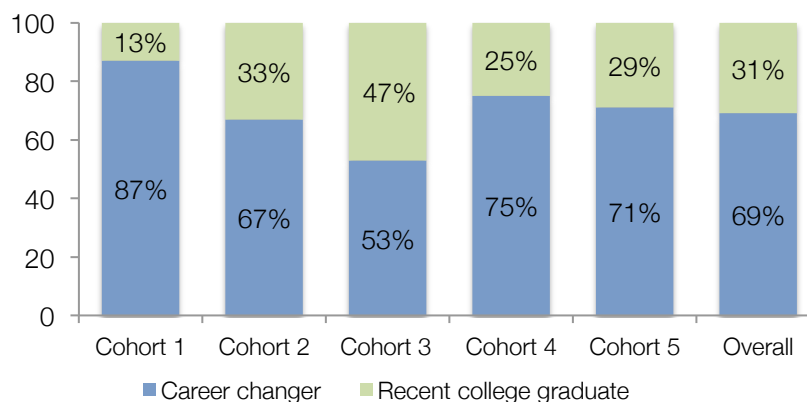


Source: New Visions Data Warehouse

Some of the residents with advanced degrees were also career changers, a group actively recruited by UTR, the NYC Teaching Fellows, and urban teacher residency programs—to bring a different kind of diversity to the pool.

Recruitment and enrollment figures also show that, for UTR, the balance tipped fairly dramatically toward the latter. Well over half of all cohorts, except Cohort 3, considered themselves career changers. (See Figure 2.)

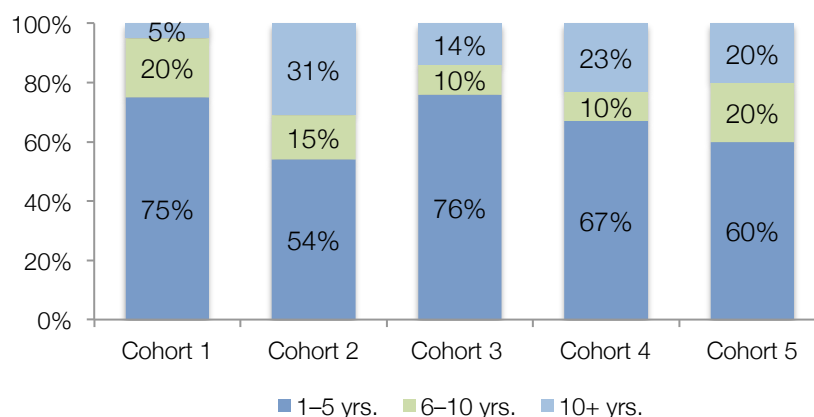
Figure 2. Percentages of Career Changers & Recent Graduates



Source: New Visions Data Warehouse

An average of two-thirds of the residents, across cohorts, had one to five years work experience, but sizeable percentages also had anywhere from six to over 10 years' experience. Cohort 2, almost a third of whom had more than a decade's experience prior to becoming a resident—again, possibly due to the addition of math—included the most experienced group of career changers; for Cohorts 4 and 5, the percentages dropped to around a fifth of the residents. (See Figure 3.)

Figure 3. Career Changers' Years of Experience



Source: New Visions Data Warehouse

Mentors' Observations on Recruitment and Selection

UTR mentors often aid in evaluating candidates during the resident selection process, and, during focus groups with the mentors, we discussed residents' backgrounds and experiences and the mix of characteristics of a successful

recruit and teacher. There was general agreement, across cohorts of mentors, that there was no sure-fire formula, and mentors pointed out—with some consistency—the benefits, and downside, of seeking out recent, top-notch graduates and career changers.

A benefit of both groups, they agreed, is their idealism and commitment to urban teaching. Mentors also observed that the desire “to do something more meaningful” could lead to unrealistically high expectations, especially among career changers. Although mentors found that more experienced residents were generally more mature, some were also more impatient, especially about how organizations work and what should be done to make an impact. Mentors also observed that the “personality and grit needed to succeed in the school system” does not always go hand in hand with idealism. Nor is idealism always durable. For a few residents, noted mentors, UTR was “like the Peace Corps,” just a “stint of something different” that lacked “a sense of urgency.” This, they also noted, “affected professionalism, [the] need to understand the school district, standards, bureaucracy. What it means to work in an urban environment.”

Mentors, along with some principals, observed that residents with strong academic backgrounds, one of the selection criteria, tended to be “high achievers,” sometimes “highly intellectual,” and accustomed to success, often “a little bit type A, which is a great attribute in terms of organization, but less so in terms of the ever-changing school environment where you have to be able to turn on a dime.” One of the responsibilities of mentors is to help residents keep things in balance, and some mentors found that high-achievers could also be high-maintenance, struggling at times to juggle classroom responsibilities and commitment to students with trying to be “at the top of the class” in their Hunter coursework.

Mentors’ recommendations about how to select, orient, or acclimate residents generally fell into two categories: Some suggested more visits, more exposure to actual schools and classrooms, even an analysis of the demographics, attendance data, and achievement numbers—“and what that might mean for what you’re trying to teach.” Some felt that efforts such as these, and summer orientation and professional development around cultural competency, could only go so far. For them, “urban prep...wasn’t something that could be grasped or taught beforehand,” but only gained through experience.

How many residents complete program requirements and obtain jobs?

Overall Completion Rates

Of the 165 residents who enrolled in the UTR program over the five years, 137, or 83% completed their residency year; 126, or 76% completed degree requirements, and 75% completed *all* program requirements, which include completing their residency year and Master’s degree and passing the required teacher certification exams. (See Figure 4.)

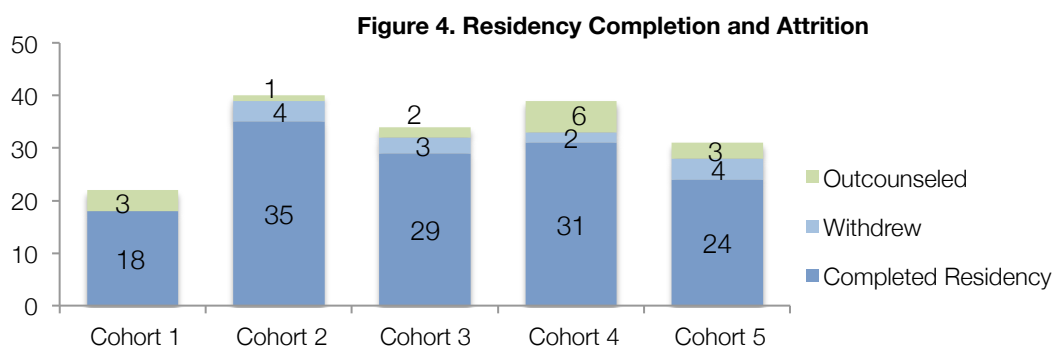
Percentages of UTR enrollees completing the residency year varied some by cohort; lower rates for Cohorts 4 and 5 likely reflect the fact that some are still in the process of completing requirements. Overall rates of 89% will likely edge up once the final cohorts of residents fulfill requirements. (See Table 6.)

Table 6. Program Completion Figures, by Cohort

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Total
Number of participants originally enrolled	21	40	34	39	31	165
Number/percentage completing residency year	18 (86%)	35 (88%)	29 (85%)	31 (80%)	24 (77%)	137 (83%)
Number completing degree requirements, 14 mos.	12	29	23	23	21	108
Number completing degree requirements, 14+ mos.	4	6	4	4	n/a	18
Total number completing degree requirements	16	35	27	27	21	126
*Total # still working toward degree	0	0	1	4	3	8
*#/% enrollees completing all requirements	16 (76%)	35 (88%)	27 (79%)	27 (69%)	19 (61%)	124 (75%)
*% completing residency year & all requirements	88%	100%	93%	87%	79%	89%

Source: Hunter College Office of Residency Programs (as of Sept. 2014)

*Includes degree requirements plus requirements for state certification



Source: Hunter College Office of Residency Programs

Just under half of those who did not complete their residency year chose to leave the program for personal reasons or because it was not the right fit (13 out of 28). The other 15 were counseled out, after consultation with program staff. Over UTR's five years, partners developed a multi-step intervention plan, initiated if the resident was not meeting expectations in either the residency or academic coursework or not adhering to professional standards—and not responding to feedback—or if their students were falling behind. When this happened, either the program officer or site director engaged multiple stakeholders—mentors, faculty, and administrators, as well as the resident—to determine what should change, what support the resident needed, and what timeframe should be set, typically a 6-week period. At this point the stakeholders determined if the resident should regain their good standing, required some additional time to do so, or should be counseled to withdraw from the program.⁷

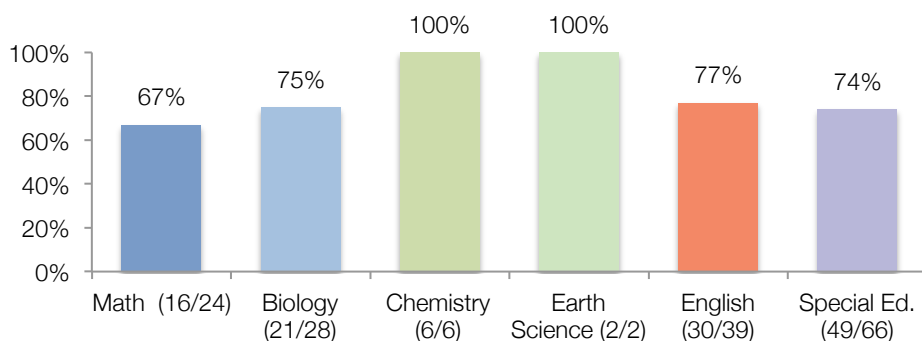
Variations by Subject

Completion rates for enrollees vary by subject. Their numbers were small (see Figure 5 below), but 100% of the chemistry and earth science residents completed all program requirements, as did three-fourths of the biology (75%),

⁷ Information from UTR Site Director's Handbook.

English (77%), and Special Education (74%) residents, whose numbers were higher. The non-completers in English and Sped may include Cohort 5 residents still finishing requirements. Math residents had the lowest completion rates, at 67%. (See Figure 5.) The lower rates may be due to the fact that recent graduates or career changers devoted to the study and application of mathematics, somewhat more so than those teaching other subjects, found themselves unsuited for the classroom. Interestingly, the better-suited math residents who stayed and obtained jobs had marked success with students (see Section 4) and the highest retention rates (see p. 62).

Figure 5. Program Completion by Subject



Source: Hunter College Office of Residency Programs

Hiring in High-Need Schools

New Visions program staff offers multiple means of support to residents in their search for full-time teaching positions at high-need schools. These include conducting job fair registrations, alerting residents to vacancies, helping them navigate the hiring process, and recommending them to principals—who have repeatedly expressed their preference for UTR-trained teachers, because of their skill and interest in collaborative, data-driven decision-making.

Of the UTR 137 enrollees completing their residency year, almost all obtained jobs, for an overall hiring rate of 96%. 88 or 65% obtained teaching positions in a New Visions PSO network high-need school. Another 26 or 19% were hired in one of the city's non-network high-need schools, for a total of 114 or 84%. An additional three graduates were hired in a New Visions charter school, also serving largely high-need areas, and three in high-need schools outside New York City. Three UTR graduates were hired in non-high-need schools. Four are still seeking jobs, and eight are not currently teaching. Overall, UTR graduates were hired in 81 different schools, 21 of which (26%) had previously hosted residents. Seventeen (21%) of the hiring schools hired three or more residents, and all but two of those schools had previously served as hosts.

Figure 6 shows the overall hiring percentages; Figure 7, shows numbers hired in high-need schools by cohort.

Figure 6. Hiring Status—All Cohorts

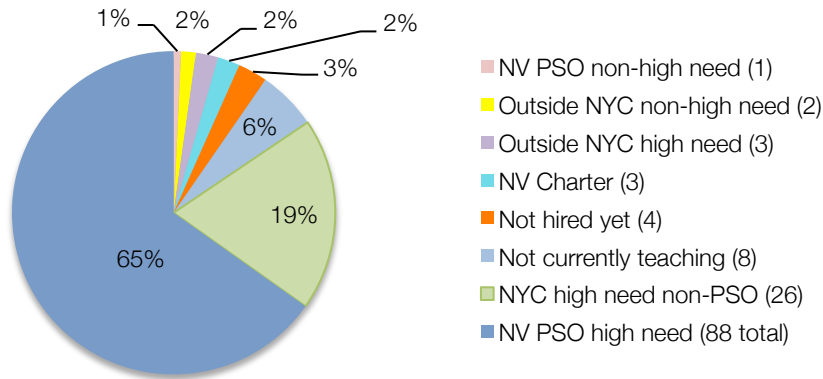
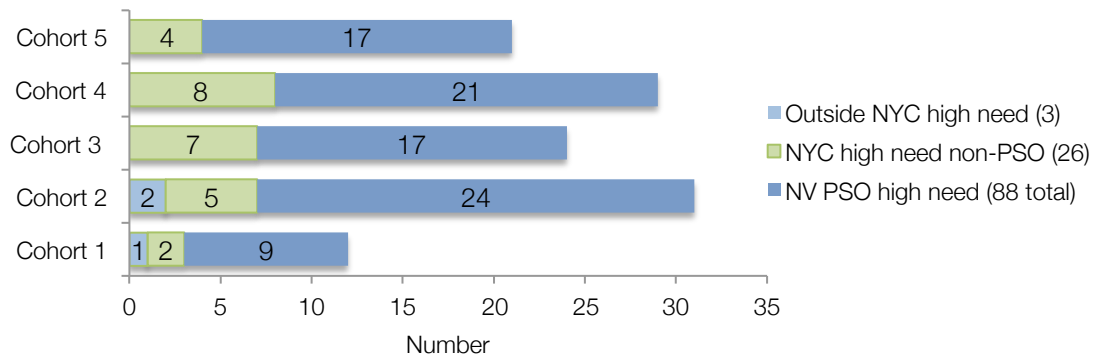


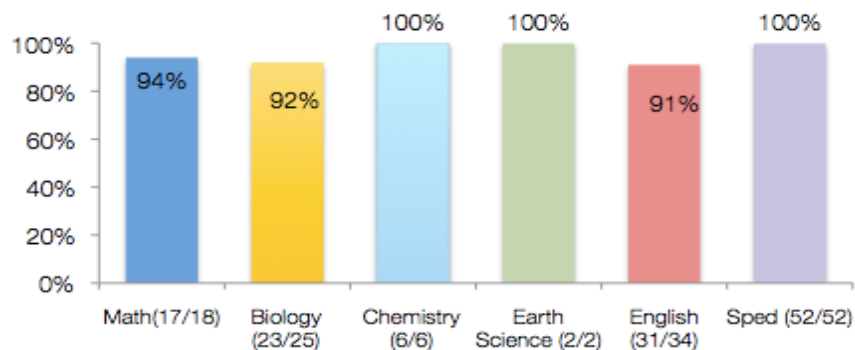
Figure 7. Total Number of Teachers Hired in High-Need Schools



Hiring for Hard-to-Staff Subjects

A grant requirement—and UTR project goal—was to train residents for the hardest-to-staff subjects where teachers were in short supply. The program focused on English language arts and science in Year 1, and, in Years 2–4, on science, mathematics, and special education. In Year 5, as the UTR model was replicated in the MASTER project, a federally-funded initiative focused on math and science, UTR focused on ELA and special education. (See Figure 8.)

Figure 8. Induction Year Hiring by Subject



Source: Hunter College Office of Residency Programs

For the supplemental analysis, we expect to have data from Hunter and the NYC DOE to see how completion and placement rates compare. In 2014, the NYC DOE began issuing teacher preparation program reports showing data for the city’s major teacher preparation institutions, including Hunter College. Data from the Hunter report includes traditional programs only; it also breaks schools down by “highest need” schools, a classification that may be narrower than our “high-need school” designation. Figures from the NYC DOE’s Teacher Preparation Program Reports, based on 2008–2009 through 2011–2012 hires, show that, overall, of the 464 teachers prepared at Hunter College through traditional programs—excluding the NYC Teaching Fellows—28% were hired into the city’s highest-need schools. Of the 10,135 teachers hired overall in that period, 30% were hired in the highest-need school.⁸

Who are the UTR Mentors?

Mentors’ Roles and Selection

Residency programs give novice teachers more practice alongside experienced teachers—more than traditional programs provide during student teaching, and more than provided in alternative certification programs in which new teachers hit the ground running much earlier. This apprenticeship gives residents time to hone instructional plans and strategies—and engages mentors in a multi-faceted role. During the summer preparation, they help orient residents to their schools and school and district policies, and help them plan units and assessment activities for the fall semester. They guide residents through the myriad classroom tasks, challenges, and responsibilities related to designing, instructing, and assessing. Using the teacher quality tools, they coach and monitor residents, providing feedback and helping them set goals. According to the Site Director Handbook, mentors ultimately “embody the UTR Vision of Effective Teaching.” A residency is a demanding year, so mentors also help residents maintain balance as they juggle teaching, preparation, and Hunter coursework. The goal is for mentors themselves to grow professionally, assuming not only more and more responsibility in preparing residents but assuming leadership roles in schools.

The selection and development of mentors was clearly a key part of the UTR program, something that could not be left to chance or to principals’ discretion. In Year 3, as New Visions program staff developed teacher quality tools for residents, they also revised mentor recruitment and selection activities and established a set of mentor competencies to ensure more consistency.

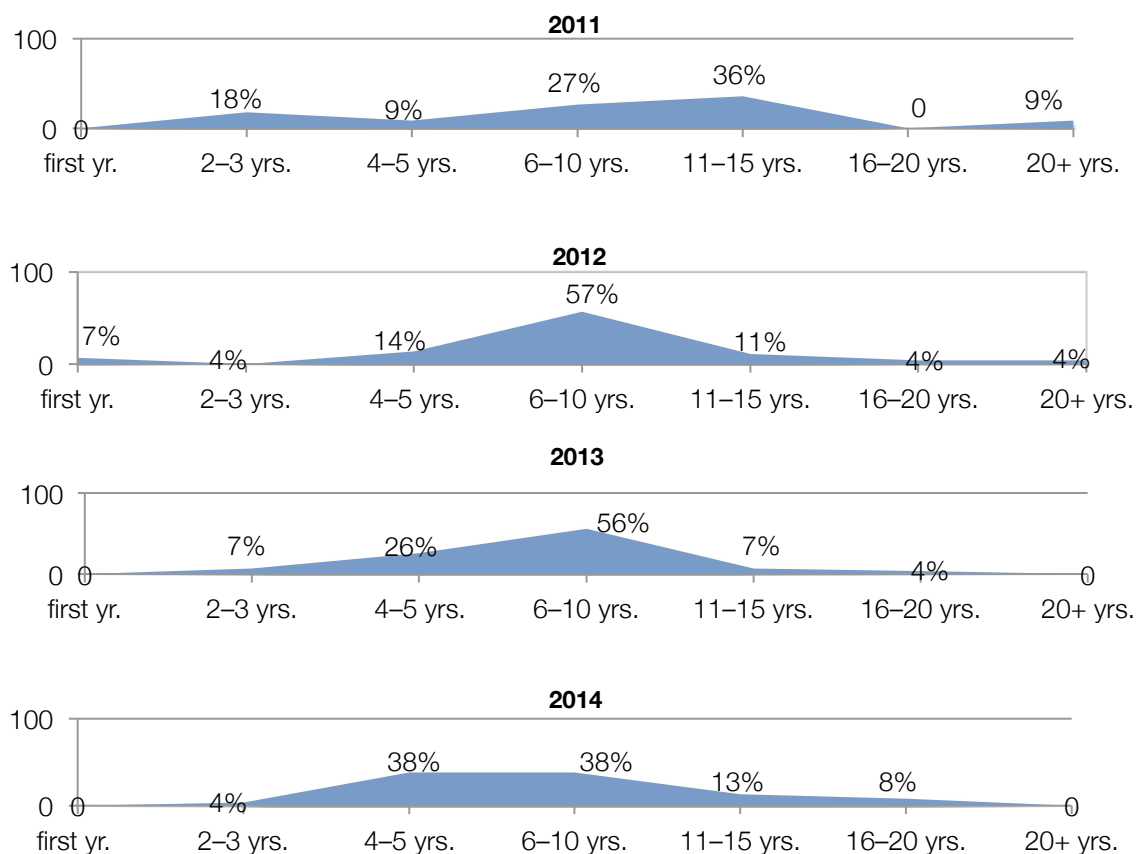
Mentors’ Background and Experience

Maximizing project benefits for mentors and schools as well as residents was an overarching goal of the UTR project, key to its success and sustainability. Mentors’ background and experience also indicated room for professional growth because many were themselves relatively new to the profession. Mentors’ years of experience varied by cohort, but self-reported figures indicate prior experience in the 6–10-year range; the mid-point in Figure 9 shifts to the right, indicating more than 6–10 years, for Cohort 1, and, for Cohort 5, to the left, or fewer than 6–10 years (see Figure 9). These figures generally reflect a city-wide teaching corps reduced by the attrition that in part inspired initiatives like UTR. Based on IBO figures, the average number of years of teaching experience for an NYC DOE teacher is around 10.6 years; for all high schools, that figure edges down to 9.7 years, for high-poverty high schools, to 8.5 years. Other counts and reviews of attrition also suggest that half of all NYC teachers don’t reach 10 years of service. Hiring and retention figures show some changes in trends—more teachers are being hired, and attrition

⁸ See schools.nyc.gov/NR/ronlyres/7691067F-B33E-4849-BC50-9C77AEF01F74/0/NYCDOETeacherPreparationProgramReportHunterCollege.pdf See also http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2013-2014/08142013_prepprograms.htm.

among those in their third to fifth year ticked down—data still show that teachers with 6–15 years of experience, in their “prime teaching years,” are leaving at troubling rates.⁹

Figure 9. UTR Mentors’ Years of Previous Teaching Experience

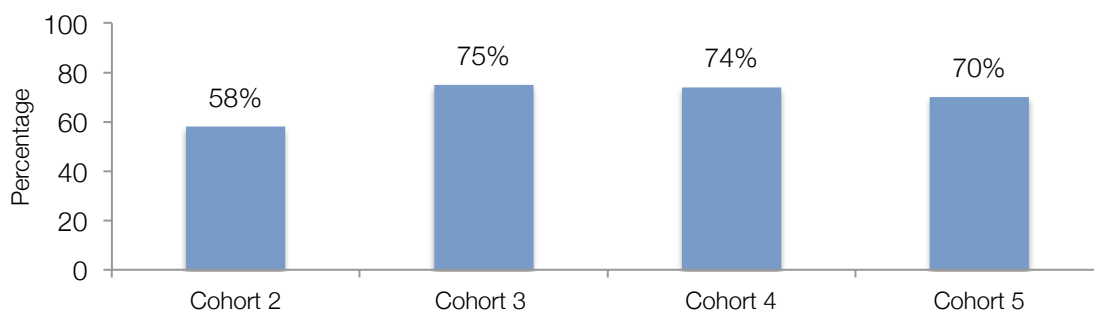


Source: REA Annual Teacher Surveys

Survey responses also show that the majority of UTR mentors—an average of 70%—entered teaching through an alternative certification program (see Figure 10), which suggested a commitment to urban education, valuable, recent first-hand experience with juggling the challenges of being a first-year teaching and meeting the demands of graduate coursework, but not necessarily a toolkit stocked by years of experience with urban populations.

⁹ See IBO “Statistical Portrait,” p. 6.

Figure 10. Percentages of Mentors Entering Profession through Alternative Certification Programs



Source: REA Annual Teacher Surveys

Support, Satisfaction, and Return Rates among Mentors

In addition to revising selection criteria and setting competencies for mentors, New Visions also provided ongoing professional development, seminars, and other activities designed to support mentors, especially in coaching. Comments from mentors, in focus groups and on surveys, suggest that they welcomed the support, especially activities that gave them more ownership for assessing and coaching residents and insights into their own teaching. They also indicated that defining their role took time, and that there was not always a shared understanding of what that role entailed. As one Year 5 mentor explained, the role may have been “defined by New Visions,” but “for the school it was just another student-teacher program.” From the New Visions perspective, the needs of the project and the growth of the mentor required an apprenticeship model with a fairly complex mentor role that enabled two adult learners to develop skills at different rates but in tandem, a process made possible by the reflection, self-assessment, continuous learning, and collaboration built into the original residency model.

Supporting and enriching this role was sometimes a challenge. Acknowledging that logistics and busy schedules complicate cross-partner sharing, throughout UTR’s five years mentors have consistently expressed a desire for more communication with Hunter faculty about the ways theory intersects with practice.

The subject-area focus of UTR changed from year to year, so there was not always the opportunity to serve as a mentor in successive years; in addition, being a mentor required devoting a lot of time to the resident and UTR program, so the expectation was not that mentors would return each year. Still, percentages of returning mentors were relatively high (see Table 7), and returning mentors pointed to positive changes from year to year: One Year 5 returning mentor singled out positive changes in the professional development provided by New Visions: “PDs have gotten better in addressing how to support the mentors.” Another praised the evolving process: “it’s only gotten more supportive and more positive.”

Table 7. Returning Mentors by Cohort

Cohort	Year 2 (N=40)		Year 3 (N=33)		Year 4 (N=39)		Year 5 (N=31)	
Returning Mentors	Number	Percent	Number	Percent	Number	Percent	Number	Percent
	4	10%	13	33%	16	41%	7	23%

Source: New Visions Data Warehouse

Teacher Quality and Confidence

3

As the UTR project progressed, developing measures or proxies of effectiveness became an important focus, for the project and the broader education community. As noted in the introduction, NCATE had called for “continuously judging” novice teachers enrolled in preparation programs to monitor progress and ensure teacher quality. As part of its Measures of Effective Teaching (MET) initiative, the Gates foundation explored and tested multiple measures.¹⁰

UTR residents received ongoing feedback from their Hunter coursework and faculty, in their practicum and other classes and as part of the assessments required for certification. UTR partners also discussed structures for monitoring residents’ growth in key instructional practices and aligning fieldwork and coursework.

In Year 3, New Visions debuted a suite of assessment tools specifically designed to guide and evaluate residents, ready them for changes in teacher evaluations, and, more broadly, define effectiveness and clarify expectations for good practice. Mapping to other project criteria, the suite generally assessed teachers as designers, instructors, and assessors.

The tools included the Danielson framework, adapted by states and districts across the country—including the NYC DOE—for their teacher evaluations; the Defense of Learning, a presentation that gives residents the opportunity to demonstrate their ability to diagnose and address learning needs; and rubrics for lesson and unit planning, professionalism, and using assessment.

To make the tools more inclusive, New Visions also created a special education addendum and a set of mentor competencies.

- Consistently high percentages of residents meet benchmarks for each teacher quality measure.
- Residents consistently score between 2 and 3, or “basic” and “proficient,” on Danielson ratings.
- Use of assessment tools started conversations, targeted areas for improvement, and prompted self-reflection.
- 89% of the residents feel prepared to teach their subject matter;
- 87%, to use data to inform instruction;
- 82%, to handle classroom management;
- 78%, to motivate and engage students;
- 70%, to improve student achievement.
- Survey scales show high, consistent, and similar levels of confidence and best practices among both residents and mentors.
- Some of the highest levels of use and confidence are for formative assessment practices and instruction informed by data.

¹⁰ See Gates MET Project, “Gathering Feedback for Teaching” at <http://www.metproject.org>.

The tools, a central and unique feature of the UTR model and reflection of its unswerving focus on the inquiry process and data-driven instructional decisions, easily met the MET recommendations for using multiple measures to gauge effectiveness. The tools gave residents with a broader view of the characteristics that a teacher in an urban school must achieve to be successful. As noted before, there were benefits not just for residents but also for mentors. Using the Danielson tool as the evaluator rather than the one being evaluated gave mentors valuable experience and a different perspective. As one mentor said in reference to the tool, “We have to use it/be evaluated using it—so let’s become expert on it! Interact with it!”

This section looks at the development and use of the assessment tools, and at residents’ own self-assessments, or confidence ratings, reported at the end of a year of clinical experience when they were not just brimming idealism, but fully aware of the challenges of urban classrooms. Also discussed here is the equally important role of the teacher quality tools in strengthening relationships between residents and mentors, providing a starting point for conversations about practice and making the process “more transparent” and “less personal,” and developing a shared definition of “well prepared” and “effective.”

How does UTR define and gauge teacher quality?

Development and Refinement of Tools

- *Defense of Learning.* The Defense of Learning (DoL), introduced in Year 1, gave residents an opportunity to show how they had used the inquiry cycle to assess and address their classes’ learning needs—in a single, end-of-year presentation to an audience of New Visions program officers, mentors, administrators, and Hunter faculty. Stakes were high for this one-time demonstration, and so were anxiety levels. In Year 2, a mid-year DoL gave residents a practice run, based on fewer students and before a smaller audience. The DoL is the tool that most clearly reflects UTR’s emphasis on accountability: residents’ graduation from the program depends on their being able to demonstrate that they are meeting learning goals and raising achievement.
- *Danielson Observation Rubric.* Introduced in Year 3, the rubric covers six domains of practice: Designing Instruction, Establishing a Culture of Learning, Managing Behavior, Using Questioning, Engaging Students, and Using Assessment. Raters use a four-point scale that ranges from unsatisfactory to basic to proficient to distinguished. Mentors expressed concerns that the rubric was not well-suited to Special Education, so, In Years 4 and 5, added a special education addendum.
- *Lesson and Unit Design and Using Assessment Rubrics.* The suite of assessment tools introduced in Year 3 also included Lesson and Unit Design rubrics. Residents had prepared lessons and units from the beginning, some as part of their summer experience or Hunter coursework, but the tools added structure and clearer expectations to the process.

The coaching around these rubrics also gave residents an opportunity to gradually improve their instructional strategies as they gained more experience in trying them out in the classroom, but, as in other UTR activities, benefits accrued to mentors as well. The lesson design reviews formed the basis of coaching conversations between mentors and program officers, themselves a model for collaborative conversations around practice.

- *Professionalism Rubric.* Professionalism had been an implicit expectation from the start, but it became clear that residents needed clearer direction about school responsibilities such as participating in meetings, responding to feedback, respecting school hierarchies and norms. New Visions introduced the Professionalism rubric with Cohort 4.

As UTR moved forward, each tool underwent revisions, based on norming activities and program officers', site coordinators', and mentors' insights about where expectations should be set, but also about how the tools could assist equally well in coaching and evaluating residents and in identifying specific areas for improvement. Benchmarks that began as best guesses came to reflect actual experience and "reference scores" for mentors and residents as they devised or followed up on intervention plans. The intervals and duties of observers or raters changed somewhat to give residents feedback from different perspectives and periodic or progressive feedback, and gradually transition the ownership of the assessment of residents to mentors, in effect shifting some of the responsibility for preparing effective teachers and modeling effective teaching to the host school.

The Danielson rubric was perhaps the most involved of the tools, due to its multiple domains and the knowledge that it would be used as part of teachers' own evaluations. Both mentors and residents agreed that the four-point Danielson rubric offered too few scoring options and not enough "room for differentiation and growth." Even residents, they noted, should not be earning 1's, and 4's seemed to be reserved for that rare teacher. As part of the Year 3 pilot, Rockman helped develop an online tool that gave observers more latitude by expanding the scale to 11 more granular options, from a 1 to a high 4, which aided in evaluations of practice and conversations with residents. For the purposes of overall scores, and to align with citywide evaluations, the high's and low's, or pluses and minuses added to the scale, could be collapsed.

A second important change, for the Danielson tool, was breaking the rubric down to look at benchmarks or focus areas for each month, which helped address mentors' concerns that it was a challenge to both focus with residents and take a comprehensive look at all the competencies. Program officers or site directors observed residents' classes, and engaged in three-way coaching conversations with the residents—another way of providing feedback to residents but recalibrating expectations and steering away from punitive critiques.

Starting in Year 3, and continuing through Years 4 and 5, New Visions also allowed—and encouraged—residents to use feedback to revise their lesson and unit plans. Allowing revisions meant that more residents met requirements, as adding a second Defense of Learning had given residents time to practice and revise. This opportunity meant that the assessments also became self-assessments, arguably a benefit on par with striving to meet benchmark scores. This self-assessment allowed residents, individually and with their mentors, to take ownership of the process. New Visions now asks residents to self-assess and engage in a "mid-year goal-setting narrative," as part of their progress on the Danielson competencies and other measures, including the Defense of Learning. The overall goal is to encourage residents to "question themselves," and create a corps of more self-reflective, self-critical practitioners.

How do residents perform on the teacher quality measures?

Overall Ratings

A review of residents' scores and percentages meeting benchmarks indicate that they generally met expectations, if not on the first try, then with revisions. For the DoL, for example, all 17 residents met requirements for their Defense of Learning In Year 1. Of the 33 Year 2 residents, 95% met or exceeded DoL expectations; in Year 3, 86% passed the DoL without revisions; 100%, after changes. In Year 4, two-thirds or 23 of the residents met the mid-year DoL benchmark, on the first try; with revisions, three-fourths or 26 of the 34 residents met the mid-year goal. In June, for their final DoL, 71% met the benchmark the first time around; with revisions, another 10%, or all but six residents met the goal—confirming that they had acquired the tools and strategies to have a positive impact on student achievement.

On Unit Design rubrics, initial numbers were a little lower, but around 50% or higher, on the first round, and, on average, near 100% with revisions. There was more variation in Lesson Design, lower ratings in the fall, but, by the spring, percentages edged close to 100%. Percentages of residents meeting the professionalism benchmarks fluctuated some, but were consistently high for the two fall, mid-year, and spring ratings, with an average of 93% of the residents meeting benchmarks, based on mentors' assessments.

In Year 5, 89% of the Special Education residents met benchmarks for the first DoL, and 100% did so for their final DoL presentations. ELA residents posted similar scores, or 87% for time one and 100% for time two. Unit and lesson design scores showed more range, as they had been previously; the average ratings by mentors were somewhat lower than program officers'. All of the special education residents met the professionalism benchmarks, compared to 93% of the ELA group. 100% of both groups met the benchmarks on all six Danielson domains by the end of the year.

Trends in Danielson Ratings

Benchmarks across the six Danielson domains varied, due to the fact that some domains—e.g., Using Assessment—required more experience and time in the classroom, and some adjustments over time. Some consistent trends emerged over time in Danielson ratings:

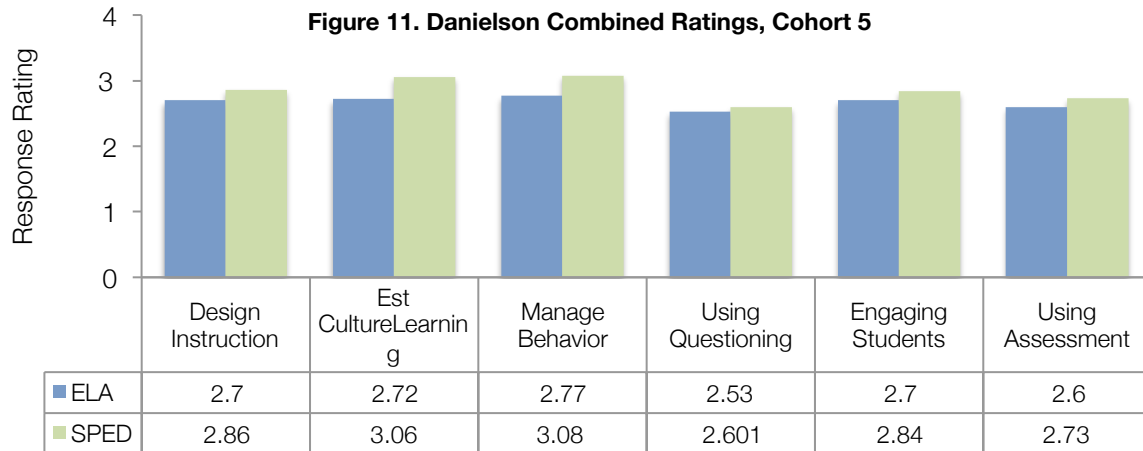
- Average ratings, across domains, were about where novice teachers should be, between 2 and 3—or developing and effective.
- Mentors consistently rated residents slightly higher, though margins narrowed from Year 3 to Year 5, and ratings in most cases converged by the year's end. Variations were not likely due to different standards or beliefs about teaching. In some cases, the mentors may have had additional contextual knowledge—they, for example, may have seen prior lessons or discussed strategies with the resident; in others, the two observers may have evaluated different lessons. There seemed to be general agreement that variations were not only tolerable but gave residents a wider range of feedback on practice.
- Residents typically scored a little higher on affective or behavioral elements of practice—engaging students, creating a culture of learning; they scored a little lower on domains related to more academically focused elements of practice—using assessment, questioning—or, again skills that required more time in the classroom.

Differences by Content Area

When the Danielson framework and rubric were introduced, mentors were worried that it might not work, or work equally well, for all subjects, especially for math or special education classes. Some data did confirm their concerns: math residents did seem to get slightly lower ratings, by a few tenths of a point. A review of scores across observers, observations, and competencies showed that both mentors and program officers tended to score special education

residents a little higher. Figure 11 shows the average ratings, across observers and ratings, for the Cohort 5 ELA and special education residents.

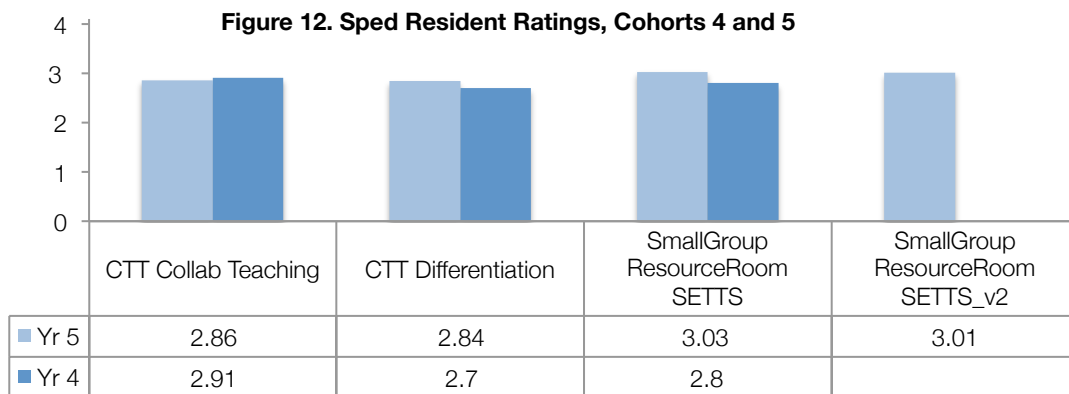
These differences are, however, small, and the takeaway may be that, across focus areas, residents' ratings are similar, suggesting that the tool is agnostic when it comes to subject area.



Source: New Visions Process Monitoring System, REA database

Special Education Addendum

Though the Danielson tool seemed generally versatile, in focus groups and on surveys special education mentors still expressed an important concern: that, for Collaborative Team Teaching (CTT) inclusion classes, there was “No mechanism where she [the resident] can be assessed as the other teacher in the room,” and that they were “lacking that concrete thing to point to.” To capture that team teaching and other special education teaching configurations, the project team created an addendum to the Danielson competency set for Years 4 and 5. Figure 5 shows the ratings for both years—with only slight variations across years, and residents achieving ratings right around 3 on the 4-point scale by the end of the year; or, in the case of small-group instruction, maintaining close to a 3 rating throughout. (See Figure 12.)



Source: New Visions Process Monitoring System

How do residents rate their own self-efficacy and confidence?

Overall Ratings

End-of-year survey responses from residents and mentors have provided teachers' own self-assessments of confidence and practice, and, over all five years of the project, responses have been both consistent and positive: Residents' self-ratings, not surprisingly, have been 10–15 percentage points lower than mentors. The Cohort 5 results and Years 1–5 averages for the confidence and sense of preparedness items show that:

- 90% of the Cohort 5 residents, compared to 100% of the mentors, feel prepared to teach their subject matter. The average over five years is 89% for residents (due largely to a dip in confidence among Cohort 3 residents) and 99% for mentors.
- 85% of the Cohort 5 residents (vs. 100% of the mentors) feel prepared to handle a range of classroom management issues. The average for residents is 82% vs. 95% for mentors.
- 80% of the Cohort 5 residents, vs. 100% of the mentors, feel confident in their ability to motivate students. The average across years is 78% for residents vs. 98% for mentors.
- In a slight departure from other items, 85% of the Cohort 5 residents vs. 79% of the mentors are confident using data to inform instruction. The five-year average for residents is 87% average, vs. 89% for mentors. Confidence among mentors declined from Years 2 through 5.
- 85% of the Year 5 residents, vs. 92% of the mentors are confident differentiating instruction. This figure is higher than the 73% average for residents, and a little higher for mentors (92% vs. an average of 89%).
- 65% of the Year 5 residents are confident in their ability to improve achievement, vs. 96% mentors. Year 5 levels are a little lower than the overall average of 70%. Averages for mentors have remained steady and high—overall, at 94%.
- Reports of the frequencies with which residents and mentors engage in various practices have been high, consistent, and similar. Cohort 5's survey responses show that they, like mentors, often or regularly:
 - have students process information by talking or writing (89% for residents vs. 96% for mentors);
 - encourage active participation (90% vs. 100%)
 - create opportunities for students to work both independently and collaboratively (92% vs. 100%).

Responses to other items related to practice indicated that both residents' and mentors' lessons had concrete, measureable objectives, and that their lessons and strategies allowed students to build on what they know, aligned with standards, to a "moderate" or "great" extent. Again, residents' ratings trailed mentors' ratings by a few percentage points. Their confidence likely reflects the benefits of clinical practice characterized by routine monitoring. The small but consistent margins between residents and mentors may also suggest the steady, comprehensive gains across areas by residents and by early-career, some barely mid-career mentors.

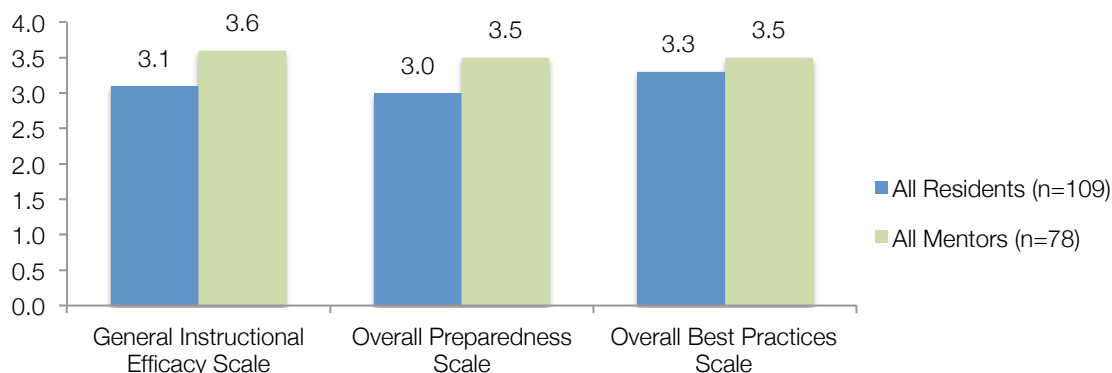
Overall Confidence, Efficacy, and Practice Scales

To get a different angle on residents' confidence and practice across years and subject areas, compared to mentors', we created three sets of overall scales, all based on four-point response options (1=not at all confident; 4=very confident; and 1=almost never; 4=almost always), for:

- confidence, which included items about teachers' confidence in handling a range of classroom management challenges, differentiating instruction, and using data to inform instruction;
- practice, with items such as how often they checked for understanding, asked questions that pushed students toward higher-level thinking, or created opportunities for students to work independently and collaboratively; and
- overall efficacy, which included items related to teaching subject matter and improving student achievement.

Overall averages, across years, showed that mentors felt a little more prepared than residents, by half and two-tenths of a rubric point. Mentors' self-reports were the same for both scales—M=3.5 on the 4-point scales; residents' frequency ratings for the best practices items were three-tenths of a point higher than their ratings for overall preparedness (M=3.3 vs. M=3.0).

Figure 13. Overall Scale Averages

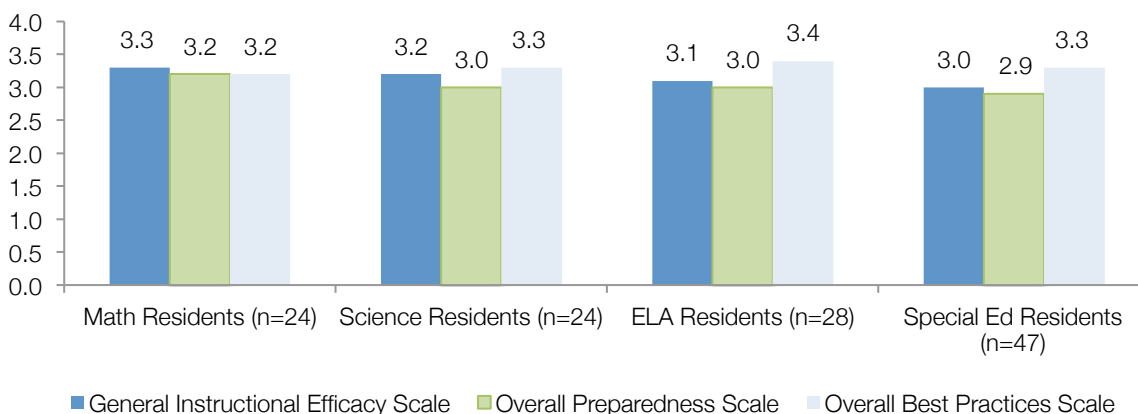


Source: REA Annual Teacher Surveys

Scale Averages and Differences by Subject and Setting

Differences were small, but subject-area ratings showed most residents assigning slightly higher ratings for what they do in the classroom, on the best practices scale, than for how prepared or confident they feel doing them. Except in the case of the math residents, residents' overall best practices ratings edged higher by three to four-tenths of a point. Special education residents' sense of preparedness was lowest, and math residents' highest, but differences were again small.

Figure 14. Scale Averages by Subject

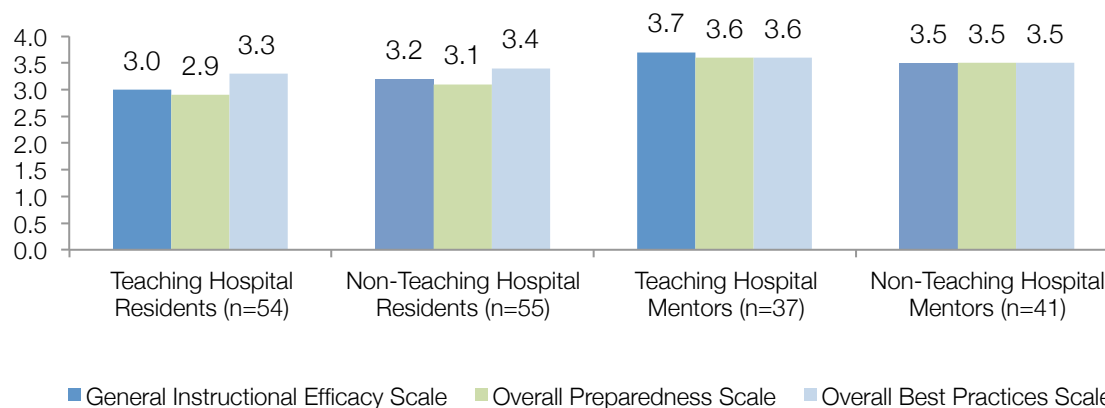


Source: REA Annual Teacher Surveys

Mentors' ratings were consistent across subjects, with a slight dip for Cohorts 3–5, possibly due to new standards or new teacher evaluations. Ratings by cohort for residents also ticked down slightly for Cohort 3 but otherwise remained fairly consistent, again with practice ratings a tenth or two of a point higher than confidence ratings.

As part of this analyses, we also looked at the effect of “teaching hospitals,” mature host sites with a concentration of several resident-mentor pairs, where administrators, faculty, and a site-based director assume more responsibility for the residents’ clinical preparation. Disaggregating ratings for residents and mentors by teaching hospital and regular host schools, we found that ratings were slightly lower for the residents in teaching hospitals, and slightly higher for mentors, and, as in other ratings, that mentors’ ratings topped residents’ by a few tenths of a point. All ratings were around a three on the 4-point rubric scales, indicating that regardless of setting mentors and residents felt confident or often engaged in certain practices. (See Figure 15.)

Figure 15. Scale Averages by Teaching Hospital vs. Regular Host Schools

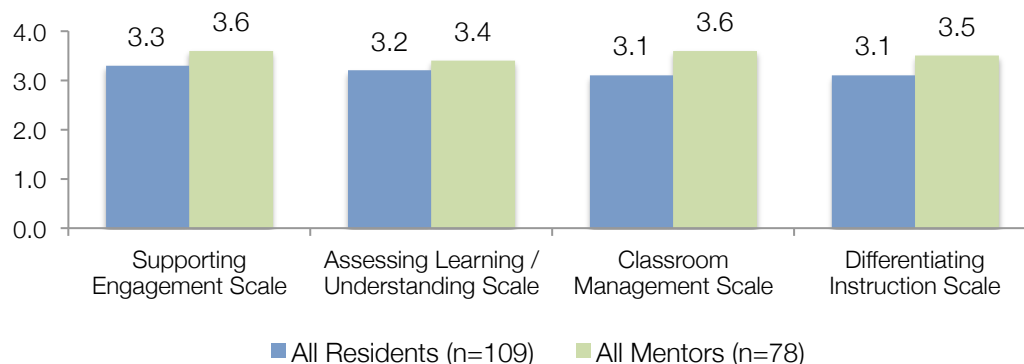


Source: REA Annual Teacher Surveys

Self-Assessment and Danielson Ratings

We created four additional scales mapped to the Danielson items, again based on how the items clustered, or loaded. We looked at how ratings varied by Cohort, and at how residents’ and mentors self-assessments compared. On the latter, we found that all ratings were between the “well prepared” or “very well prepared” levels, or between the “most of the time/often” and “regularly” levels. Residents’ self-reported ratings were around 3.2; mentors,’ a few tenths of a point higher, or around 3.5. (See Figure 16.)

Figure 16. Scale Averages on Danielson Items



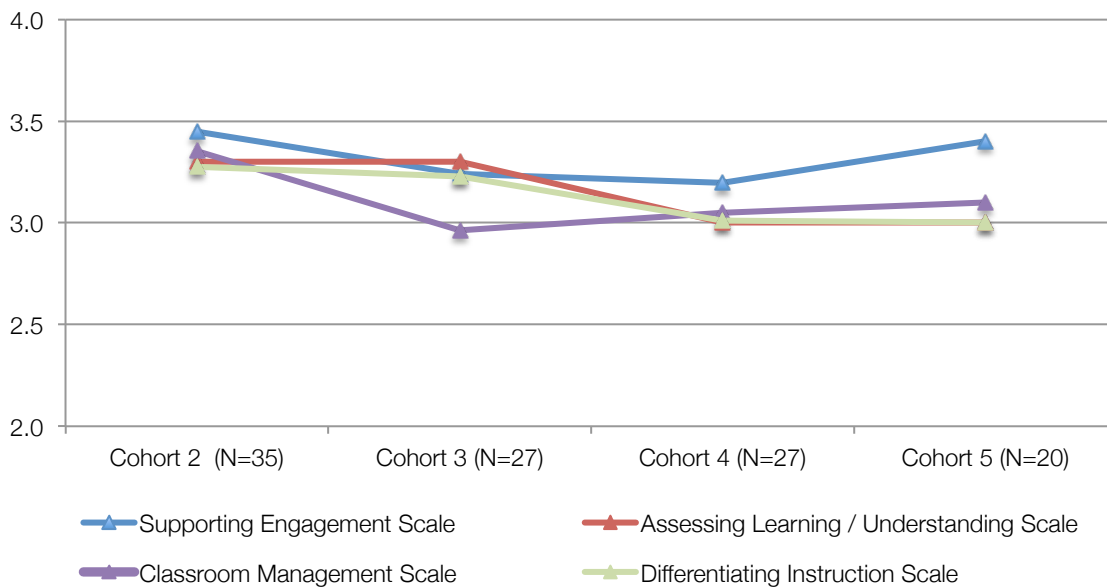
Source: REA Annual Teacher Surveys

One of the reasons we created the survey scales was that, in previous years, there were some discrepancies in how residents’ performed on the Danielson ratings and how they assessed their own skills on surveys. For example, though they scored relatively high on Managing Student Behavior and Establishing a Culture of Learning, residents

confessed some doubts about their ability to manage the classroom. Conversely, they scored a little lower on Using Assessment, but expressed high levels of confidence in these skills—likely, we assumed, because a good portion of their professional development and their performance at the Defense of Learning focuses on data-driven instruction. Year 3 residents’ survey responses suggested that, in practice, they were carrying out activities associated with classroom management: most (80%), for example, said they frequently “maintained class routines” and “ensured that discipline problems did not interfere with lessons”—but their confidence ratings still showed some doubts. In Year 4, responses were reversed: residents were a little more confident, but engaged in classroom management practices a little less often. In Year 5, confidence and practice ratings lined up.

The scaled responses indicated that, from Cohort 2 through Cohort 5, residents’ confidence and use of practices associated with management, engagement, differentiation, and assessment started at approximately the same point, then, with Cohort 3, the first two ticked down, while the differentiation and assessment items held steady. With Cohort 4, these two also dropped, but management ticked back up. For Cohort 5, these three remained unchanged, but supporting engagement rose to its higher, Cohort 2 levels. (See Figure 17.) The use of the assessment suite starting in Year 3, and fully articulated in Years 4 and 5, may explain the drop in management ratings in Year 3 and the slight declines for three of the four scales—not necessarily because residents were less confident but because successive Cohorts engaged in more self-scrutiny. The trajectory of the engagement scale results may simply reflect the subject focus of Cohort 5, and the fact that ELA residents feel more confident in engaging students.

Figure 17. Residents' Preparation and Practice Scale Responses, by Year



Source: REA Annual Teacher Surveys

Student Achievement

UTR vs. Non-UTR

4

The teacher quality ratings all involve using formative assessment data to determine learning needs, set instructional goals, and design lessons. All five years of the project, we have analyzed achievement data to see how residents' students fared on summative assessments—Regents and course grades—compared to matched groups of peers taught by other early career teachers. This was an important metric, but not a true gauge of teacher effectiveness and impact because mentors may be present in the classroom. Starting in Year 2, we also examined the performance of UTR graduates' students, again compared to peers.

Findings have consistently indicated that students taught by UTR-trained teachers are doing as well as or better than their peers. Annual analyses designed to see if UTR was narrowing the gap between students with and without IEPs have also yielded positive results.

This section begins with findings from a set of longitudinal questions posed as part of the summative study: Were there discernible differences across cohorts? Were there clear trends or consistent performance in particular subject areas? Did residents who appeared to be effective teachers become more effective in successive years?

Clear affirmatives in the overall results indicate that the work with mentors, coaches, and faculty to focus classroom efforts on student achievement and assessment paid off in higher performance. The results for Year 5 (SY 2013–2014, based on achievement data from Cohort 5 residents and graduates from Cohorts 1–4) were not quite as conclusive as earlier or overall results. This is likely due to commonplace fluctuations in annual test scores, perhaps a reflection of the instability in math scores stemming from new standards and assessments. There were, however, still instances where students taught by UTR-trained residents and graduates were performing on par with or better than their peers.

- Reviews of 5 years of student achievement data show that overall trends favor UTR over matched comparison groups, especially for Regents scores.
- In 16 out of 22—73% vs. 27%—comparisons with statistically significant differences, students taught by UTR residents and graduates outperformed peers.
- The UTR group had higher *Regents* scores most often for Living Environment; higher *grades* for ELA.
- UTR *residents'* students scored higher than peers on Regents and grades; UTR *graduates'* students, on Regents but not grades.
- Achievement trends were most positive for Cohorts 2 and 4.
- UTR also appears to help narrow achievement gaps—on Regents and grades—for students with IEPs.

What longitudinal trends emerged in the achievement data, Years 1–5?

Trends across Years, by Outcome Measure

Five years of achievement data (Cohorts 1–5 for all available assessment years), show a positive edge for UTR, especially on Regents exams. In the 22 comparisons where differences were statistically significant, students taught by UTR teachers—residents and graduates combined—outperformed those taught by non-UTR early career teachers 16 times on Regents scores, or in 73% vs. 27% of the comparisons. Results were closer for course grades: across all comparisons, UTR teachers' students outperformed the non-UTR group 12 out of 22 times, or in 55% vs. 45% of the comparisons. (See Figure 18. All achievement data retrieved by New Visions data team from Datacat and student performance databases.)

Figure 18. UTR vs non-UTR, Residents and Graduates

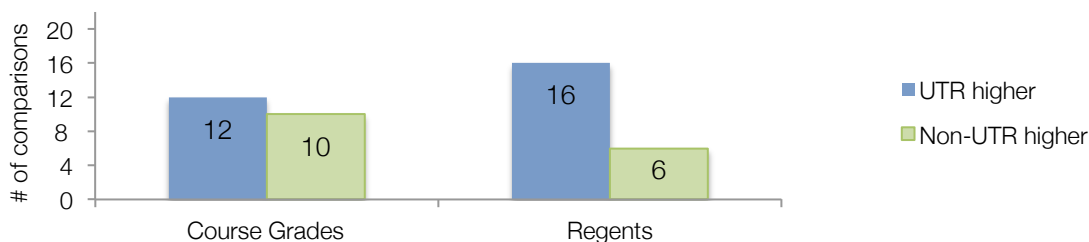
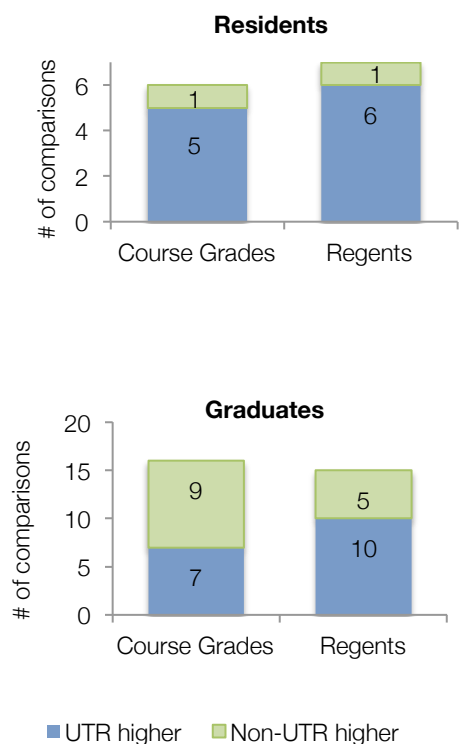


Figure 19. UTR vs. Non-UTR, Residents vs. Graduates



Breakdowns by resident and graduate groups show that there were more instances where the UTR-taught students had higher grades and Regents scores. **Residents'** students earned higher grades than peers in five instances, and outperformed peers six times on Regents. There were only two instances where peers had higher grades.

The only set of comparisons where the balance shifted to the non-UTR group was on course grades for **graduates'** students. Non-UTR students earned higher grades nine times, vs. seven times for the UTR group. (The shift could indicate that UTR-trained teachers score harder, though we have no other evidence of this.) UTR **graduates'** students, however, outperformed peers 10 times on Regents: the non-UTR group had higher scores five times. (See Figure 19.)

Trends by Subject Area

The shading in Table 8 shows, by subject, whether UTR-taught (blue) or non-UTR taught (gray) students earned higher Regents scores and grades; statistically significant differences are in bold. Figures 20 and 21 illustrate the numbers of times the UTR group had higher scores and grades in individual subjects. UTR-taught students earned higher Regents scores most often on the Living Environment exam (N=5),

followed by English, Integrated Algebra, and Chemistry (N=3 each); they posted higher grades most often in English (N=6) and Living Environment (N=3).

Table 8. Trends across Years, by Subject, Residents and Graduates Combined, Cohorts 1–5

	Regents (N=22)		Course Grades (N=22)	
	UTR higher (N=16)	Non-UTR higher (N=5)	UTR higher (N=12)	Non-UTR higher (N=9)
English	3	0	6	3
Integrated Algebra	3	1	0	2
Algebra 2/Trig	1	0	1	0
Geometry	0	1	0	1
Living Environment	5	3	3	1
Earth Science	1	0	1	0
Chemistry	3	0	0	2
Physics	ND	ND	1	0

Figure 20. UTR Higher: Regents Scores (out of 16)

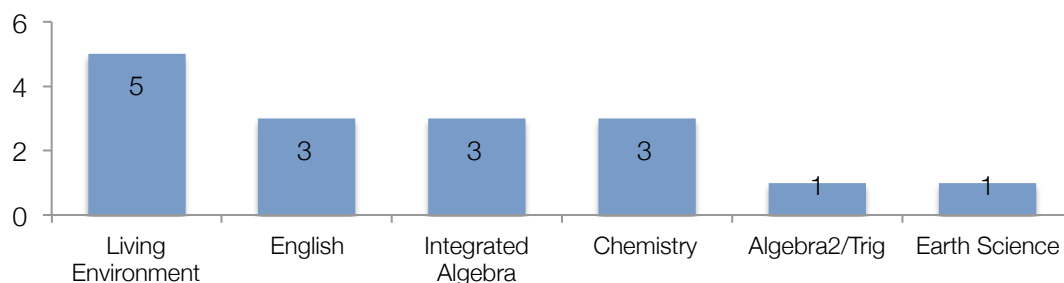
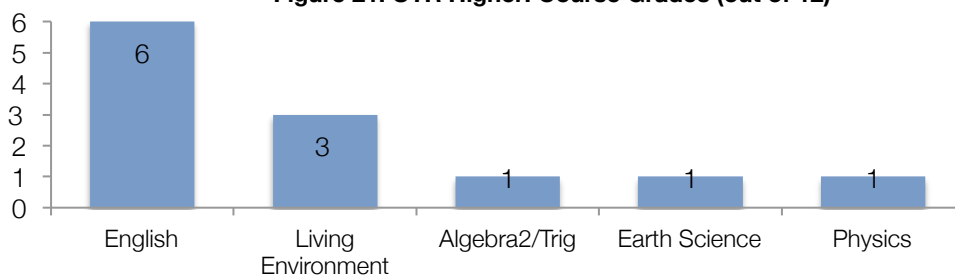


Figure 21. UTR Higher: Course Grades (out of 12)



Trends across Years, by Cohort

The trends by cohort are not entirely clear or consistent, likely due to the fact that even in the same school, student populations change from year to year. There are, however, instances where each UTR cohort outperformed the comparison group, noted in *italics* below. Students taught by Cohorts 2 and 4 teachers have consistently higher Regents scores, but not grades. Cohort 4 appears to have the most positive or consistent Regents results.

Cohort 1

- There is not a consistent pattern in Regents performance for Cohort 1. *Cohort 1 teachers' students have consistently earned higher grades than the comparison group in English (3 out of 5 years); one year they also had significantly higher grades in Living Environment. There was also a 2-year trend where Cohort 1 teachers' students had lower grades in Chemistry compared to peers.*

Cohort 2

- *Cohort 2 UTR teachers have consistently outperformed non-UTR teachers on Regents exams, particularly in the areas of English (3 out of 4 years) and Living Environment (3 out of 4 years); Cohort 2 UTR teachers also outperformed non-UTR teachers one year on the Integrated Algebra Regents exam; there were no instances in which non-UTR teachers had significantly higher Regents scores.*
- *There was comparably consistent positive pattern in course grades; in 8 instances of statistically significant differences, 5 favored the non-UTR group.*

Cohort 3

- There has not been a consistent trend in Regents exam performance or in course grades of students of Cohort 3 UTR teachers.

Cohort 4

- *Over the two years of available data, there has been a consistent trend for students of Cohort 4 teachers to outperform students of non-UTR teachers on the Integrated Algebra Regents exam and the Chemistry Regents exam (for both assessment years). Students of Cohort 4 UTR teachers also outperformed the comparison group on the Earth Science exam one year; there were only 2 instances where students of non-UTR teachers had significantly higher scores (Geometry and Living Environment, SY2013–2014).*
- There does not appear to be a comparably consistent trend for Cohort 4 in course grades.

Did the Year 5, SY 2013–2014 results follow previous trends?

In previous reports, we have noted where students taught by UTR-trained teachers (Cohorts 1–4, SYs 2010–2013) outperformed peers—findings that often favored UTR. For this report, we again looked at annual data (SY 2013–2014) which included all five cohorts: residents from Cohort 5, and graduates from Cohorts 1–4.

Below are SY 2013–2014 results by Cohort, starting with grades, then Regents scores. (See Appendix B, p. 64, for correlations between course grades and Regents scores, by cohort, for SY 2013–2014). As noted in the introduction to this section, it's not clear what account for the dip, though, for math subjects, it may be due to the new Common Core standards and assessments—in place for Algebra 1, on the horizon for Geometry and Algebra 2. Although there are no new exams in science, the introduction of the Next Generation Science standards may have resulted in shifts in practice but not in assessments, resulting in a misalignment that explains lower scores.

Course Performance Results, by Cohort, SY2013–2014

UTR Residents, Cohort 5

Using analysis of covariance (ANCOVA) and controlling for the covariates we used in previous years, we found that UTR was a significant positive predictor of course grades in English. Course grades for the UTR group were also higher, but differences were not statistically significant. (See Table 9 below for numbers of teachers and students, adjusted means, standard error, and p values. Blue shading indicates higher performance by the UTR group; gray, by the non-UTR group.

Statistically significant difference in bold.) (NOTE: Although Cohort 5 included only ELA and Sped residents, some of the latter are part of collaborative teams with content area teachers and listed as the primary teacher in the data file.)

Table 9. Course Performance, Students Taught by Cohort 5 Residents vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
English	283 (15)	73.8	0.7	828 (15)	71.6	0.4	p<.01
Geometry	24 (1)	72.4	5.8	44 (1)	69.2	3.5	ns
Integrated Algebra	24 (2)	72.3	2.5	129 (2)	74.3	1.0	ns
Global Studies / History	12 (1)	70.4	2.7	58 (1)	68.8	1.0	ns

UTR Graduates, Cohort 4

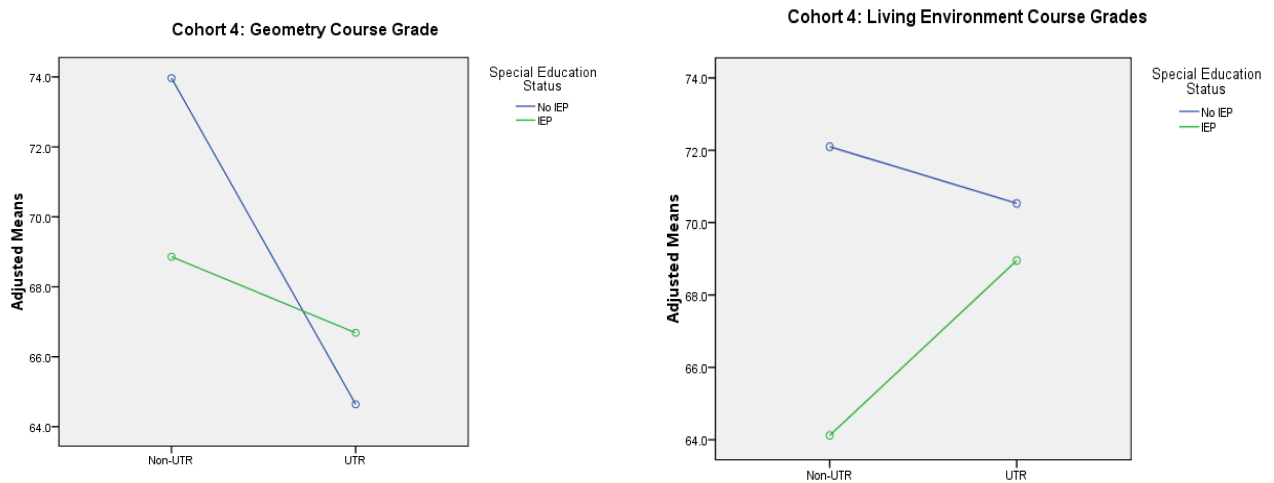
UTR participation was a significant negative predictor of course grades in Geometry (see Table 10). Students taught by UTR graduates earned higher grades in Biology, Earth Science, and Living Environment, but we did not find statistically significant differences.

Table 10. Course Performances of Cohort 4 UTR Students vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
Geometry	82 (3)	65.3	1.2	225 (4)	73.1	0.7	p<.001
Integrated Algebra	237 (4)	69.5	0.9	158 (4)	69.9	1.1	ns
Biology	77 (1)	82.3	1.4	24 (1)	77.2	3.4	ns
Chemistry	368 (3)	70.9	0.6	208 (3)	72.3	0.8	ns
Earth / Environ. Science	137 (1)	77.0	1.3	100 (2)	75.2	1.6	ns
Living Environment	165 (4)	70.2	1.0	220 (4)	70.0	0.8	ns

We also found a significant interaction between IEP status and UTR status for Integrated Geometry course grades. For students of UTR teachers, the course grades were more similar between students with and without an IEP; for students of non-UTR teachers, there was a larger discrepancy between students with an IEP and those without an IEP in terms of course grades ($F(1,296)=5.15, p=.023$). We also found an interaction between IEP status and UTR status for Living Environment course grades, but the gap between students with and without an IEP was smaller for students of UTR trained teachers ($F(1,374)=5.59, p=.019$). (See Figure 22.)

Figure 22. IEP vs. Non-IEP, Cohort 4



UTR Graduates, Cohort 3

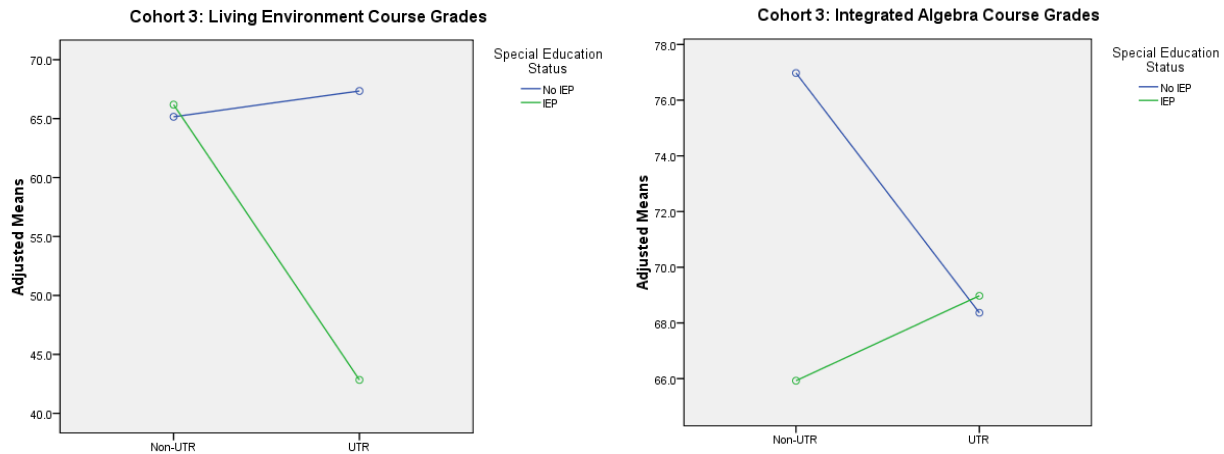
UTR participation was a significant positive predictor of course grades in Earth/Environmental Science, but a negative predictor of course grades in Integrated Algebra. The UTR group had higher course grades in Algebra 2/Trig and Geometry, but the differences were not statistically significant. (See Table 11.)

Table 11. Course Performances of Cohort 3 UTR Students vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
Algebra 2 / Trig	111 (3)	80.5	1.0	42 (3)	77.6	1.8	ns
Geometry	229 (3)	73.9	0.9	187 (3)	71.2	1.0	ns
Integrated Algebra	59 (2)	69.0	1.7	111 (2)	74.1	1.2	p<.05
Earth / Environ. Science	98 (2)	87.8	1.3	147 (2)	80.4	1.0	p<.001
Living Environment	59 (2)	65.3	2.3	130 (2)	65.6	1.5	ns

We found a significant interaction between IEP status and UTR status for Integrated Algebra course grades. For students of UTR teachers, the course grades were more similar between students with and without an IEP; for students of non-UTR teachers, there was a larger discrepancy between students with an IEP and those without an IEP in terms of course grades ($F(1,159)=6.49, p=.012$). A significant interaction was also found for Living Environment, however the pattern was in the opposite direction, with students taught by UTR teachers showing a greater discrepancy between students with and without an IEP ($F(1,178)=9.00, p=.003$). (See Figure 24.)

Figure 23. IEP vs. Non-IEP, Cohort 3



UTR Graduates, Cohort 2

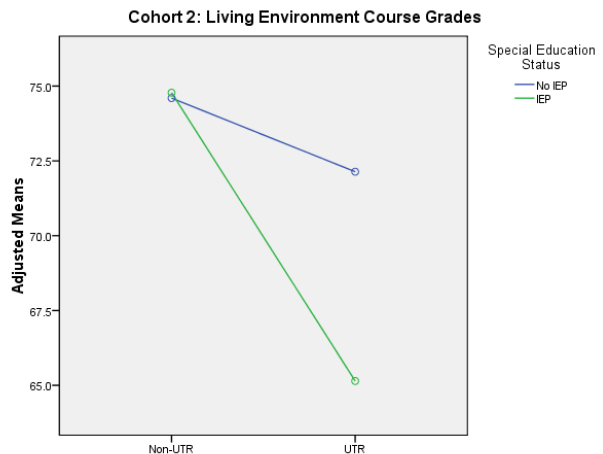
UTR participation was a significant positive predictor of course grades in English and in Physics, but a negative predictor of course grades in Living Environment and Global Studies/History (see Table 12).

Table 12. Course Performances of Cohort 2 UTR Students vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
English	562 (9)	77.3	0.5	795 (9)	74.7	0.4	p<.001
Integrated Algebra	82 (1)	68.4	1.6	72 (2)	73.5	1.8	ns
Living Environment	347 (6)	71.2	0.9	305 (6)	74.8	0.9	p<.01
Physics	21 (1)	85.6	6.6	79 (1)	65.5	2.8	p<.05
Global Studies / History	50 (2)	74.2	1.7	81 (2)	78.9	1.2	p<.05

Figure 24. IEP vs. non-IEP, Cohort 2

We found a significant interaction between IEP status and UTR status for Living Environment course grades. For students of UTR teachers, there was a greater discrepancy in the grades of students with an IEP and students without an IEP than there was for students of non-UTR teachers ($F(1,641)=4.90$, $p=.027$). (See Figure 24.)



UTR Graduates, Cohort 1

UTR participation was a significant positive predictor of course grades in English (UTR adjusted mean=79.2, Non-UTR adjusted mean=70.7, $F(1,365)=27.7$, $p<.001$). We did not find a significant interaction between IEP status and UTR status on course grades for any subject area.

Table 13. Course Performance of Cohort 1 UTR Students vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
English	178 (3)	79.2	1.1	197 (3)	70.7	1.0	$p<.001$
Integrated Algebra	28 (1)	69.4	2.4	44 (1)	68.7	1.9	ns
Chemistry	218 (3)	71.4	1.0	148 (3)	73.6	1.2	ns
Living Environment	76 (1)	78.3	1.5	67 (1)	74.9	1.6	ns

Regents Performance Results, SY2013–2014, by Cohort

UTR Residents, Cohort 5

UTR participation was a significant negative predictor of Regents scores in Living Environment. Again, there were no significant IEP interactions (see Table 14). (NOTE: Because students do not typically take the ELA Regents in ninth grade, we had no Regents scores of Cohort 5 ELA residents.)

Table 14. Regents Performance of Cohort 5 UTR Students vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
Integrated Algebra	23 (2)	61.4	2.8	129 (2)	62.1	1.0	ns
Living Environment	13 (1)	51.7	4.7	39 (1)	74.9	2.0	$p<.001$

UTR Graduates, Cohort 4

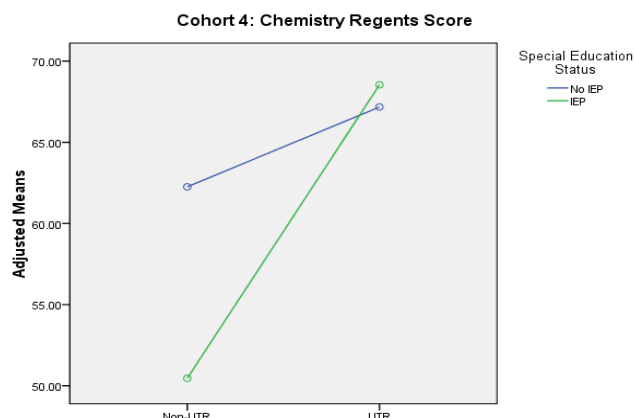
UTR participation was a significant positive predictor of Regents scores in Integrated Algebra (regular Regents), Integrated Algebra (Common Core), Earth/Environmental Science, and Chemistry, but a negative predictor of Regents scores in Geometry and Living Environment. (See Table 15.)

Table 15. Regents Performance of Cohort 4 UTR Students vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
Geometry	24 (2)	46.0	2.4	45 (2)	57.5	1.6	$p<.01$
Integrated Algebra	68 (2)	66.5	1.2	94 (2)	63.1	1.0	$p<.05$
Integrated Algebra (CC)	63 (1)	60.6	1.2	53 (1)	55.1	1.2	$p<.01$
Living Environment	216 (3)	67.7	0.7	207 (3)	69.8	0.7	$p<.05$
Earth / Environ. Science	243 (1)	62.2	0.9	130 (1)	56.0	1.4	$p<.01$
Chemistry	707 (3)	67.2	0.4	318 (3)	61.8	0.6	$p<.001$

Figure 25. IEP vs. Non-IEP, Cohort 4

We found a significant interaction between IEP status and UTR status for Chemistry Regents scores. For students of UTR teachers, the positive impact of having a UTR trained teacher was experienced by both groups of students (those with and without an IEP); however the positive benefits were greater for the IEP group resulting in a decreased gap between the groups in comparison to the gap seen for students who were taught by a Non-UTR teacher ($F(1,1014)=11.90$ $p=.001$). This represents an ideal interaction pattern. (See Figure 25.)



UTR Graduates, Cohort 3

UTR participation was a significant positive predictor of regents scores in Algebra 2 / Trigonometry (UTR adjusted mean=55.6, Non-UTR adjusted mean=39.7, $F(1,32)=7.60$, $p=.010$) but a negative predictor of regents scores in Integrated Algebra (regular regents) (UTR adjusted mean=56.6, Non-UTR adjusted mean=78.3, $F(1,131)=131.50$, $p<.001$), Integrated Algebra (Common Core) (UTR adjusted mean=53.9, Non-UTR adjusted mean=67.7, $F(1,126)=49.90$, $p<.001$), and Living Environment (UTR adjusted mean=65.8, Non-UTR adjusted mean=72.7, $F(1,184)=14.75$, $p<.001$). (See Table 16.) We did not find any significant IEP interactions.

Table 16. Regents Performance of Cohort 3 UTR Students vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
Algebra 2 / Trig	29 (1)	55.6	2.8	11 (1)	39.7	4.8	$p<.05$
Geometry	120 (3)	68.5	1.3	120 (3)	66.9	1.2	ns
Integrated Algebra	43 (2)	56.6	1.5	98 (2)	78.3	0.9	$p<.001$
Integrated Algebra (CC)	40 (1)	53.9	1.6	96 (2)	67.7	1.0	$p<.001$
Living Environment	36 (2)	65.8	1.6	158 (2)	72.7	0.5	$p<.001$

UTR Graduates, Cohort 2

UTR participation was a significant positive predictor of Regents scores in English (regular Regents) (UTR adjusted mean=69.8, Non-UTR adjusted mean=63.9, $F(1,262)=12.2$, $p=.001$) and in English (Common Core) (UTR adjusted mean=74.8, Non-UTR adjusted mean=62.8, $F(1,132)=14.3$, $p<.001$). (See Table 17.) There were no significant IEP interactions.

Table 17. Regents Performance of Cohort 2 UTR Students vs. Non-UTR Students (ANCOVA)

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
English	126 (8)	69.8	1.2	146 (8)	63.9	1.1	p<.01
English (CC)	87 (3)	74.8	1.6	55 (3)	62.8	2.2	p<.001
Integrated Algebra	52 (1)	62.0	1.6	15 (1)	65.0	3.9	ns
Living Environment	385 (6)	73.6	0.6	532 (6)	74.8	0.5	ns
Global Studies / History	23 (2)	55.9	4.4	48 (2)	61.2	2.6	ns

UTR Graduates, Cohort 1

There were no significant differences or significant IEP interactions between UTR and non-UTR.

UTR Graduates, Cohorts 1–4 Combined Results, SY2013–2014*Course Performance Results*

For the combined graduate sample for SY 2013–2014, UTR participation was a significant positive predictor of course grades in English, but a negative predictor of course grades in Integrated Algebra (see Table 18). Students of UTR teachers earned higher grades in Algebra 2/Trig, Biology, and Physics, though there were not statistically significant differences.

Table 18. Course Performances, all GRADUATES (Cohorts 1–4) UTR Students vs. Non-UTR

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
English	740 (12)	77.5	0.5	992 (12)	74.1	0.4	p<.001
Algebra 2/Trig	111 (3)	80.5	1.0	42 (3)	77.6	1.8	ns
Geometry	311 (7)	70.9	0.7	412 (8)	72.8	0.6	ns
Integrated Algebra	406 (10)	69.2	0.6	386 (11)	71.6	0.7	p<.05
Biology	77 (1)	82.3	1.9	24 (1)	77.2	3.4	ns
Chemistry	598 (7)	70.8	0.5	395 (7)	72.7	0.7	p<.05
Earth / Environ. Science	235 (3)	73.1	0.8	247 (3)	78.4	0.8	p<.001
Living Environment	647 (13)	70.9	0.6	722 (13)	72.0	0.6	ns
Physics	21 (1)	85.6	6.6	79 (1)	65.5	2.8	p<.05
Global Studies/History	50 (2)	74.2	1.7	81 (2)	78.9	1.2	p<.05

Regents Results, SY 2013–2014

For the combined sample, UTR participation was a significant positive predictor of Regents scores in English (regular Regents, English (Common Core), Algebra 2/Trigonometry, Chemistry, and Earth Science, but a negative predictor of Regents scores in Integrated Algebra (regular Regents), Integrated Algebra (Common Core), and Living Environment. (See Table 19.) The 2013–2014 results are, again, not quite as favorable to UTR as other years' findings, or the

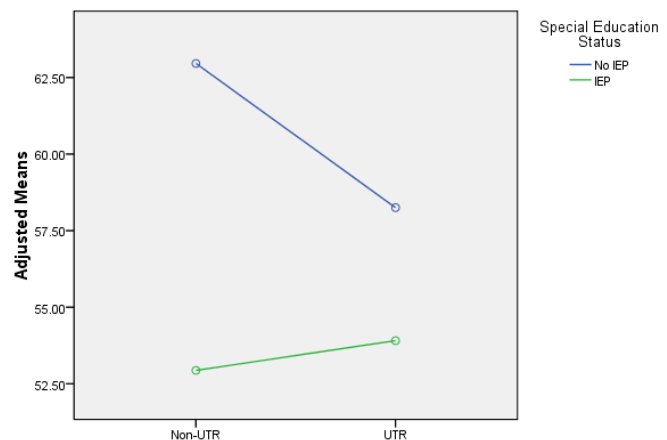
overall findings, but UTR trained graduates' students outperformed peers in five of the eight instances where differences were statistically significant.

Table 19. Regents Performance, all GRADUATES (Cohorts 1–4) UTR Students vs. Non-UTR

Subject Area	UTR			Non-UTR			Sig.
	N student (teacher)	Adjusted Mean	SE	N student (teacher)	Adjusted Mean	SE	
English	126 (8)	69.8	1.2	146 (8)	63.9	1.1	p<.01
English (CC)	87 (3)	74.8	1.6	55 (3)	62.8	2.2	p<.001
Integrated Algebra	184 (6)	61.6	0.9	243 (7)	67.8	0.8	p<.001
Integrated Algebra (CC)	172 (6)	57.3	0.8	187 (5)	60.8	0.7	p<.01
Algebra 2 / Trig	29 (1)	55.6	2.8	11 (1)	39.7	4.8	p<.05
Geometry	144 (5)	64.8	1.1	166 (3)	64.4	1.1	ns
Chemistry	814 (7)	64.8	0.4	499 (7)	59.7	0.5	p<.001
Earth / Environ. Science	243 (1)	62.2	0.9	130 (1)	56.0	1.4	p<.01
Living Environment	707 (12)	71.7	0.4	1031 (12)	73.4	0.3	p<.001
Global Studies / History	23 (2)	55.9	4.4	48 (2)	61.2	2.6	ns

Figure 26. IEP vs. Non-IEP, Cohort 1–4

Cohorts 1 - 4 (all graduates) Common Core Regents Score in Integrated Algebra



For the combined graduate sample (Cohorts 1– 4), there was a significant interaction between IEP status and UTR status for the Integrated Algebra Common Core Regents score. Students with an IEP did slightly better when taught by a UTR teacher; those who did not have an IEP did not fare better ($F(1,348)=4.58$, $p=.033$). (See Figure 26.) We did not find a significant interaction between IEP status and UTR status on course grades.

Additional Special Education Analyses, UTR Graduates, Cohorts 1–4

To examine the performance of students taught by special education graduates, Cohorts 1–4, we also looked at the numbers of total credits earned and required Regents passed, dividing students into three groups:

- special education students taught by a UTR-trained special education teacher
- special education students taught by a UTR-trained regular education teacher
- special education students taught by a non-UTR trained teacher

Results were either very similar or favored the non-UTR trained teachers for grades 9 and 10. For grades 11 and 12, the UTR-taught students in some cases earned more credits and passed more Regents. The most notable difference was for grade 12 students: those taught by a UTR-trained special education teacher earned approximately 2 more credits. Twelfth grader special education students taught by a UTR-trained regular ed teacher passed more Regents than those taught by the UTR-trained special education teacher, and more than those taught by a regular ed teacher. (See Table 20.)

Table 20. Comparison of Credits Earned and Regents Passed, Special Education Students

	Total Credits Earned			Required Regents Passed		
	Special Ed Students Taught by UTR-trained Special Ed Teacher	Special Ed Students Taught by UTR-trained REGULAR Ed Teacher	Special Ed Students Taught by Non-UTR teacher	Special Ed Students Taught by UTR-trained Special Ed Teacher	Special Ed Students Taught by UTR-trained REGULAR Ed Teacher	Special Ed Students Taught by Non-UTR teacher
Grade 9	11.3	11.2	13.5	0.8	0.9	1.1
Grade 10	22.6	24.1	25.1	1.4	2.3	2.7
Grade 11	36.5	38.3	37.6	2.2	3.9	3.6
Grade 12	51.6	49.8	49.5	2.4	4.1	3.9

Teacher Performance Student Outcomes

5

The student achievement data suggest that, for the most part, UTR-trained teachers' students perform on par with or better than their peers. The teacher quality assessment scores suggest that residents are meeting performance benchmarks. The question explored here is whether the two are linked.

To explore this, and test the predictive value of the teacher quality tools, we conducted a set of analyses with Cohort 3, 4, and 5 data. We first checked to see if the tools themselves formed a coherent set. Combining two metrics—student achievement data and teacher quality scores—we then explored links between residents' performance and their students' performance.

Using residents' scores on each tool (i.e., DoL, Danielson, etc.), we first looked at correlations between residents' performance and their students' performance Regents scores and grades. We then created a composite score to see if higher performing residents meant higher performing students. For the summative study, we also looked beyond the residency year to the induction year, to see if teachers' performance as residents was predictive of their effectiveness, or their students' performance, once they were teachers of record.

These analyses are to some degree exploratory, and we were sometimes limited in the conclusions we could draw because of small sample sizes, but we did find that the tools formed a coherent set—based on high r or correlation values—and indications that certain assessments were fairly strong predictors.

- Statistically significant correlations between the individual teacher quality measures suggest a coherent set of tools.
- The most tightly associated measures are scores on various Danielson domains and Lesson Design and Defense of Learning scores. In some cases, Professionalism and Using Assessment were also among the tightly linked measures.
- We found strong correlations between the Danielson domains related to classroom culture and both Regents and grades.
- Defense of Learning scores are also fairly strong predictors of student outcomes.
- Results were mixed, but there are some indications, based on composite scores and students' Regent scores and grades, that strong residents are more effective teachers.

Do the assessment tools form a coherent set of measures?

Links between Measures, by Cohort

Gauging how well the tools hang together is important for two reasons. First, all the tools generate quantitative, continuous variables that we can count or chart, and it's helpful to see how one might explain the other: for example, to what extent do Defense of Learning scores explain the variance in, say, the average Danielson ratings, or vice versa? Second, and perhaps more important, though they focus on individual practices, the assessment tools are interrelated and should work in concert; the statistical links should mirror the observed practice. What the correlation analyses do is confirm that data from one tool can shed light on data from another, and that the tools hang together statistically as well as operationally.

Analyses of the correlations between measures, based on scores for UTR Cohorts 3–5 (special education included) and MASTER residents, suggest that the measures themselves hold together reasonably well as a coherent set of tools. (In the tables below, shading indicates higher correlation values, and asterisks indicate statistically significant correlations. (See Appendix B, pp. 74–75, for scatterplots for Cohorts 3 and 4. Those for Cohort 5 are included in the text. See Appendix D, p. 76, for all correlation coefficient values, across cohorts, teacher quality measures, student outcome measures, and residency and induction years.) The source for all scores on teacher quality measures comes from the New Visions Process Monitoring System. We created average scores and composite scores from residency-year ratings on each measure.)

The notable finding from Cohort 3 is that the Danielson, Lesson Design, and Defense of Learning Scores were tightly associated ($r=.67$ or higher among the main scores). (See Table 21.)

Table 21. Correlations between Cohort 3 Resident Ratings (N=34)

	Danielson	Lesson Design	Defense of Learning
Danielson Average Score	----		
Lesson Design Score	.70*	----	
Defense of Learning Score	.75*	.67*	----

For Cohort 4, we again saw a statistically significant correlation between the DoL score and the Danielson average score and the Lesson design score ($r=0.35$ and 0.48).

Table 22. Correlations between Cohort 4 Resident Ratings (N=34)

	Danielson	Lesson Design	Defense of Learning
Danielson Average Score	----		
Lesson Design Score	0.31	----	
Defense of Learning Score	0.35*	.48*	----

For Cohort 5 ELA residents, there was a strong correlation between Lesson Design and Danielson ($r=0.59$), and between Lesson Design and Using Assessment ($r=0.56$) For the Special Ed group, there was, again, a strong link between Lesson Design and Danielson ($r=0.81$), Professionalism and Danielson ($r=0.83$), (See Tables 23 and 24.)

Table 23. Correlations between Cohort 5 Resident Ratings

	Danielson	Lesson Design	Defense of Learning	Professionalism	Using Assessment
Danielson Average Score	----				
Lesson Design Score	0.59*	----			
Defense of Learning Score	0.14	0.37	----		
Professionalism Score	0.34	0.24	0.29	----	
Using Assessment Score	0.26	0.56*	0.11	0.44	----
Unit Design	0.31	0.39	0.03	0.18	0.61*

Table 24. Correlations between Cohort 5 Resident Ratings, Special Ed

	Danielson	Lesson Design	Defense of Learning	Professionalism	Using Assessment
Danielson Average Score	----				
Lesson Design Score	0.81**	----			
Defense of Learning Score	0.37	0.63	----		
Professionalism Score	0.83**	0.71*	0.17	----	
Using Assessment Score	0.79	0.84*	0.31	0.64	----
Unit Design	0.51	0.59	0.49	0.78*	0.44

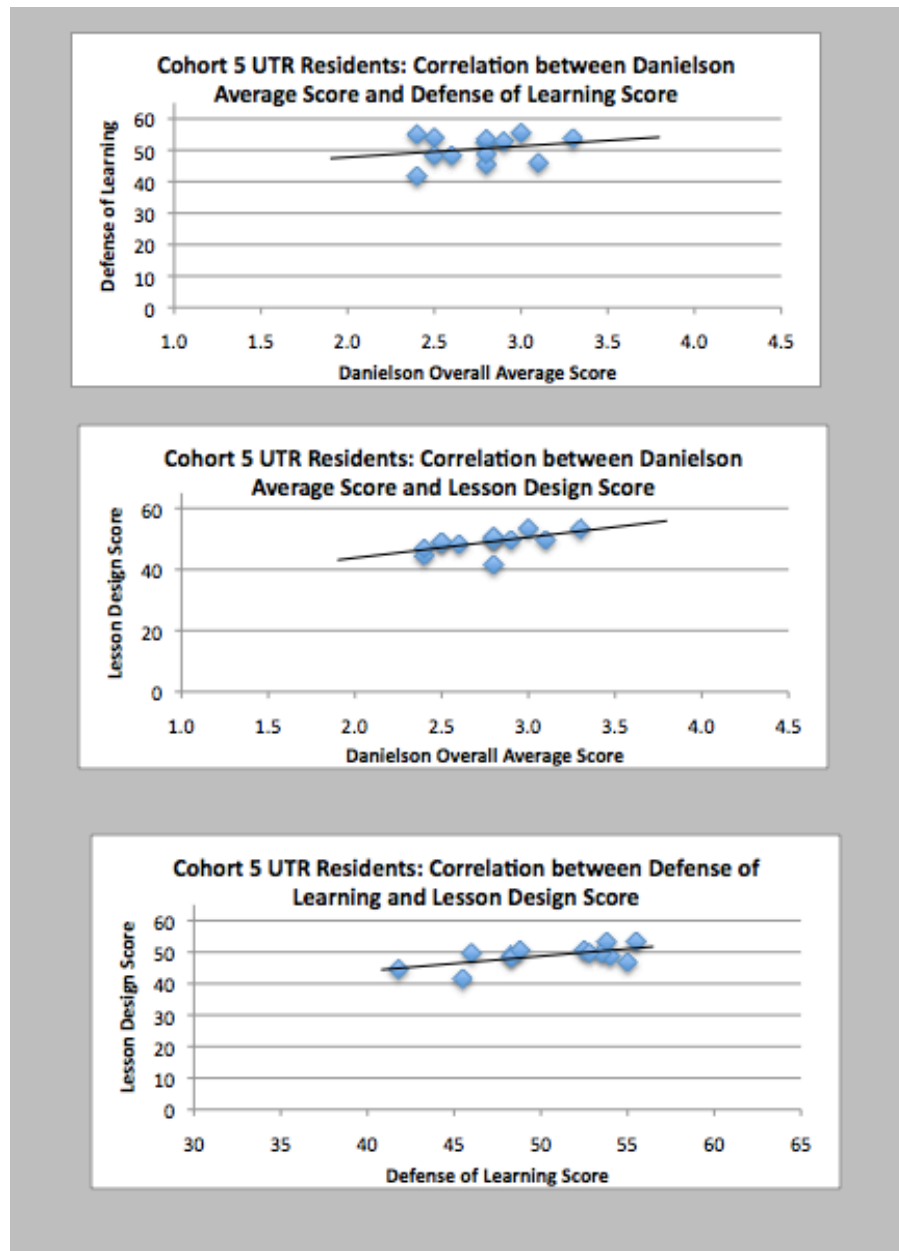
(As another confirmation of the consistency of the tools themselves, we examined the results for the MASTER project, and again found fairly strong correlations. See Table 25.)

Table 25. Correlations between MASTER Resident Ratings (Lesson Design N=10; others N=19)

Assessment Tool	Danielson	Lesson Design	Defense of Learning	Professionalism
Danielson Average Score	----			
Lesson Design Score	0.43	----		
Defense of Learning Score	0.49**	0.39	----	
Professionalism Score	0.50*	0.19	0.53*	----
Using Assessment Score	0.64**	0.59	0.68**	0.38

* p<.05; **p<.01

Figure 27. Cohort 5 Correlations



Can residents' scores predict their students' performance?

Positive Correlations, Strongest Links—by Rating and Cohort

Analyses conducted to explore links between residents' and students' performance, or gauge the predictive value of the tools, again revealed some statistically significant correlations and some consistent links. The discussion below includes results from Cohorts 3 and 4, using residency year teacher quality assessment data, and student data from

both residency and induction years, or SY 2013–2014. Cohort 5 results are for the residency year only (SY 2013–2014).

Overall, the most consistent or strong correlations appeared to be between student performance and the Danielson domains related to classroom culture:

- Residents' scores on Establishing a Culture of Learning and Managing Student Behavior were most often associated with students' performance on course grades and Regents.
 - For Cohort 4, we found 7 statistically significant links between residents' scores and student performance during their residency year (SY 2012-2013, Regents and grades).
 - For Cohort 5, residents' scores on these two domains were associated with grades.
- Residents' performance on the Defense of Learning was also a fairly strong predictor.

Cohort 5

Only 5 teachers had Regents data, therefore Regents Scores are not included in our Cohort 5 analysis. (As noted earlier, residents largely taught ninth grade English, which does not typically lead to a Regents exam; students usually take the ELA Regents in eleventh grade.)

We found significant positive correlations between Danielson Establishing Culture of Learning and course grades; Defense of Learning and grades and attendance; and Lesson Design and attendance. For Special Education, we found significant positive correlations between DoL and attendance, Unit Design and attendance, and Defense of Learning and Unit Design, shown in the scatterplots below, which appear to be the strongest positive predictors of outcomes for Special Education students, though they were not significant (low sample size). (See Figure 28.)

Figure 28. Cohort 5 Correlations: Resident Ratings and Student Outcomes



Can teachers' performance in their residency year predict their students' performance in their induction years?

Residency vs. Induction Year Performance

We explored this question in a couple of ways. We first looked at overall resident performance on the assessment suite and at their students' performance during the residency and induction years, for Cohorts 3 and 4.

Residency year results for Cohort 3 did not suggest the same kinds of links that we found for other cohorts, in part because our sample size was small and Regents' scores limited. We found that Lesson Design and Unit Design scores were positively associated with Regents ($r = .36$ and $.57$), though they were not significant. Otherwise, several coefficients were negative, and the only statistically significant correlation between residents' performance score and student outcomes—for professionalism and course grades—was negative, suggesting an inverse relationship between residents' professionalism and their students' grades.

The SY 2013–2014 results, when Cohort 3 residents were in their second year as teachers of record, were, however, quite different. There was a statistically significant link between residents' Danielson Engaging Students domain score and their students' Regents' scores. There were fewer links to grades, but still a series of positive associations.

We found significant positive correlations during Cohort 4's residency year, but fewer when they were teachers of record, during their first induction year (SY 2013–2014).

We also found that average course grades and Regents scores for Cohort 4 residents—who, in our achievement data analysis, seemed to produce better student performance, compared to matched peer groups—went down slightly from their residency year to their induction year. (See Figures 29 and 30.)

Figure 29. Cohort 3 Residents: Comparison of Course Grades and Regents Scores across Years

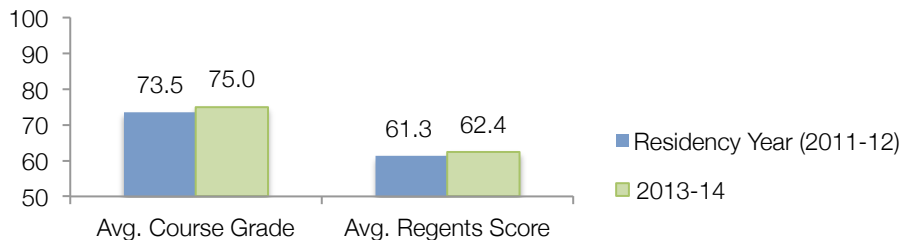
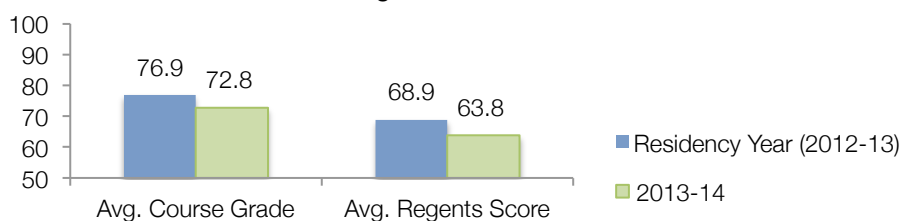


Figure 30. Cohort 4 Residents: Comparison of Course Grades and Regents Scores across Years

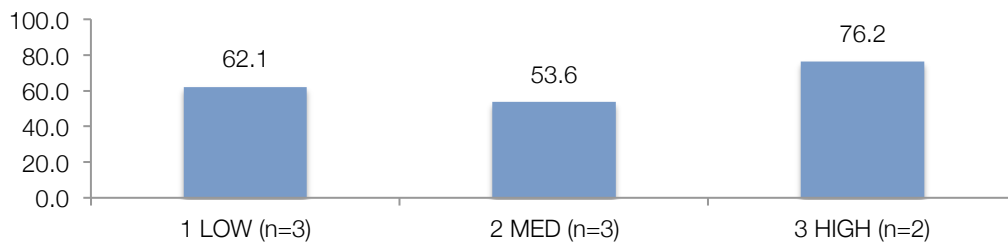


Composite Ratings

We also created three relatively equal groups to represent low/medium/high composite scores and looked at mean differences by composite ratings. Although the differences were not always statistically significant, some trends emerged for both grades and Regents, with students taught by residents in the medium and high categories scoring higher than those taught by lower-rated residents. (NOTE: the N's for residents are small in all the analyses, especially for Cohort 3.)

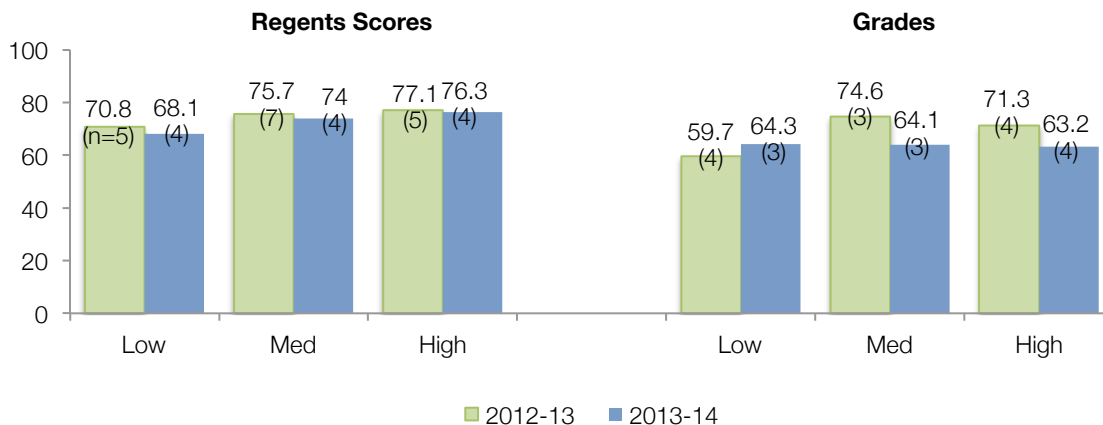
For SY 2013–2014, Cohort 3 residents' second year as teachers of record, we found that students taught by residents in the high group did perform better than those in the other two groups; students taught by residents in the middle group actually had the lowest Regents scores. (See Figure 31.)

Figure 31. Cohort 3: Regents Avg. Score by Resident Composite Rating, 2013–14



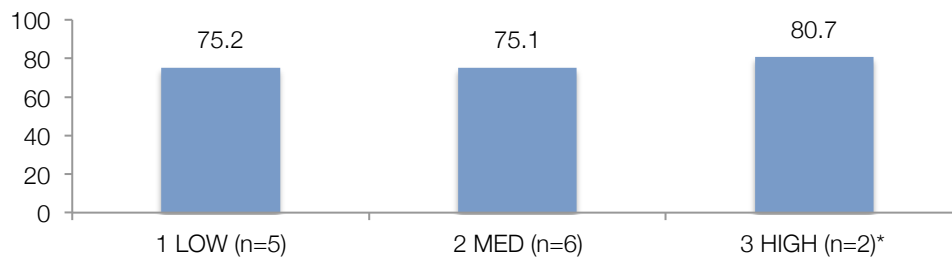
For Cohort 4, residency year results were somewhat different from induction year results. During the residency year 2012–2013, green bars below), students of higher performing residents did perform better on grades and Regents (see green bars below. In 2013–2014, the Cohort 4 residents' induction year (blue bars), Regents scores were a little higher, but grades were almost exactly the same. (See Figure 32.)

Figure 32. Cohort 4: Residents Composite Rating



For Cohort 5, we did not have enough teachers whose students had Regents scores to examine those, but, for their students' grades, we saw the pattern repeat: higher performing residents = higher performing students (see Figure 33).

Figure 33. Cohort 5: Course Grade by Resident Composite Score, 2013–2014



Retention and School Impact

6

The final measure examined as part of the summative evaluation is teacher retention, along with the factors that keep teachers in schools: administrative support, collaboration and collegiality among teachers, and opportunities for leadership. Over the last five years, there have been some improvements in retention rates among New York City teachers. Figures and reports vary, but data still indicate that attrition is still high where poverty is high, and that about a third of all new teachers still leave the district after three years.

The first cohort of residents has just now reached the five-year point, and retention rates look good. This section discusses how rates vary by subject area and cohort, and how they compare to retention rates for the city and other teacher preparation programs.

The section also shares feedback from residents and mentors about school climate and support factors linked to retention. UTR was designed not just to groom and brace teachers for urban schools, but to create the conditions in UTR host schools that support retention, to encourage mentors to become teacher leaders, and, in many ways, to make residents the ambassadors who would take skills and dispositions to other schools. These include a commitment to collaboration, open communication, accountable and reflective practice, and data-driven instruction and decision-making. This section also looks how school climate and levels of support in teaching hospitals and in UTR concentration schools that have hosted and hired several residents compare to other schools.

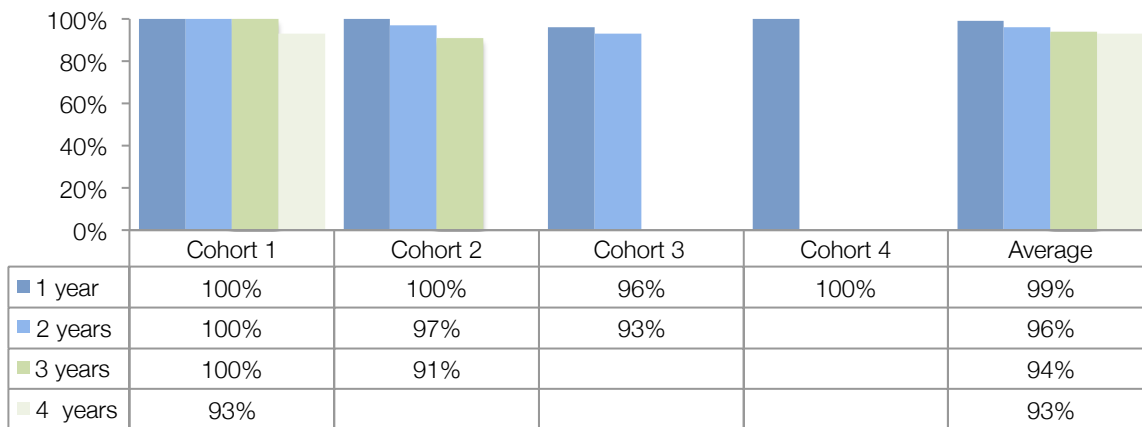
Are the residents staying?

Compared to the questions about the predictive value of assessment tools, this question is easy, and gets an unqualified “yes.” Retention rates for UTR teachers are high.

- UTR teachers are staying. Retention rates have held steady, above 90% for each cohort each successive year, at 93% overall after 5 years.
- Teacher retention after 5 years exceeds citywide rates by sizeable margins, and tops Teaching Fellows and UTRU rates.
- Retention is highest among math residents (100%); lowest, for special education residents (85%).
- Feedback from residents and mentors suggests positive changes over 5 years in factors that keep teachers in schools.
 - more positive school climate
 - more mentor involvement in collaborative decision-making
 - more professional engagement on the part of mentors
- Comparisons of teaching vs. non-teaching hospitals suggest some positive differences for mentors.
- Job satisfaction was a little lower for Cohort 5.
- Job satisfaction appears to be higher in schools with higher UTR concentrations; we found no marked differences in student performance.

Rates for the Cohort 1 residents hired in 2010 are a few percentage points lower than rates for those hired four cohorts later, but are still holding steady, at 100% for teachers' first three years in the classroom, dipping to 93% in their fourth year. For Cohort 2, retention rates started at 100%, and dropped a few percentage points each year, for a 91% rate after three years. Averages across cohorts show that, after a year in the classroom, 99% of the UTR-trained teachers stayed; with slight attrition, average percentages ticked down to 96% after two years, 94% after three years, and 93% after four years—still a high rate by any measure. (See Figure 34.)¹¹

Figure 34. Retention Rates among UTR Teachers, 2010–2014



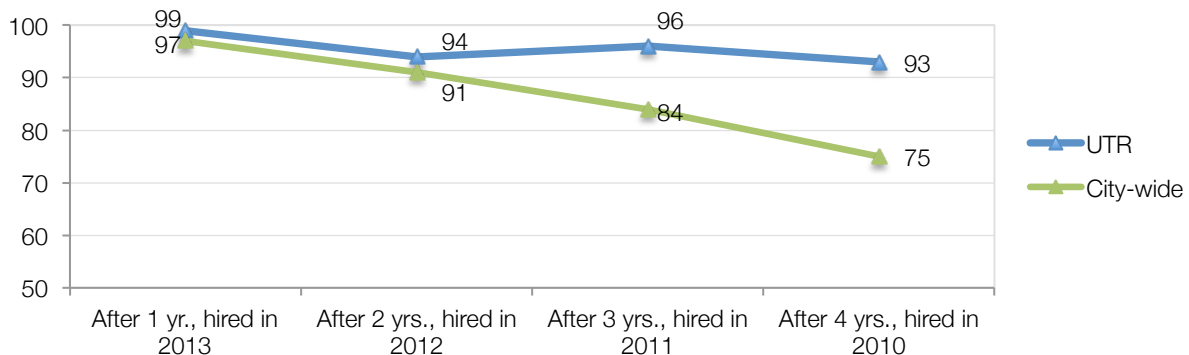
Source: Hunter College Office of Residency Programs

Comparisons

The UTR retention rates and city-wide rates are fairly comparable for teachers in their induction years—99% vs. 97%, 94% vs. 91%—but then diverge. By four years out, a fourth of the teachers have left schools city-wide, compared to only 7% of the UTR teachers. (See Figure 35.) UTR rates are also higher than rates among NYC Teaching Fellows: 87% begin a second year of teaching—just above the national average of 86%; 73% are still teaching after three years. The New Visions–Hunter UTR rates also exceed the overall UTRU rate of 89%.

10

Figure 35. UTR vs. City-Wide Retention



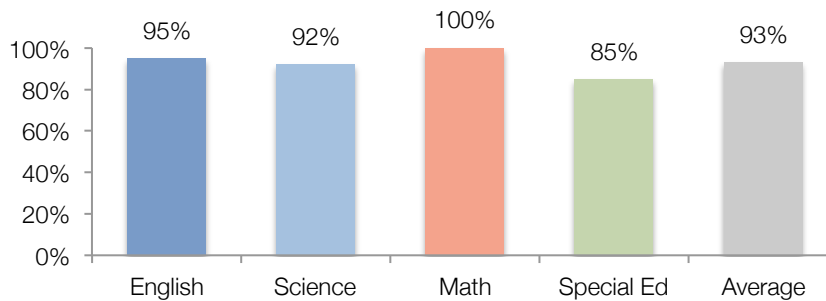
¹¹ Maisie Macadoo. "Teacher Attrition up after Recession-Driven Lull." From [UTF.org](http://www.uff.org), accessed 3/10/2014 from <http://www.uff.org>.

¹² See NYC Teaching Fellows' statistics at https://www.nycteachingfellows.org/about/program_statistics.asp.

Retention Rates by Subject

By subject, average UTR retention rates, after four years, show that rates are highest for math, followed closely by English and science. At 85%, rates are lowest for special education (see Figure 36). Special education is consistently identified as one of the hardest to staff subjects, and thus UTR included special education residents from Cohorts 2–5. Other analyses of retention also indicate that special education teachers leave the classroom at slightly higher rates than other teachers.¹³

Figure 36. UTR Retention Rates by Subject, After 4 Years



Source: Hunter College Office of Residency Programs

Why do teachers stay, or leave?

A growing body of research indicates that, though teachers are the most important factor in student achievement, it is not their students' achievement that keeps teachers in schools, or flagging achievement that drives them away. What's important to teachers is school environment: collaborative activities, leadership opportunities, support from administrators, and mechanisms such as common planning time that allow teachers to work together and grow. Absent these things, teachers, especially younger, talented teachers, may be inclined to leave, or seek out other schools with a more supportive environment.

All five years of the UTR project, through surveys and focus groups, we have asked residents and mentors about the environment or conditions in the host schools. Generally, residents' and mentors' assessments have remained positive, with residents' ratings, not surprisingly, a little lower than mentors'.

Participation in Collaborative Activities and Leadership Opportunities

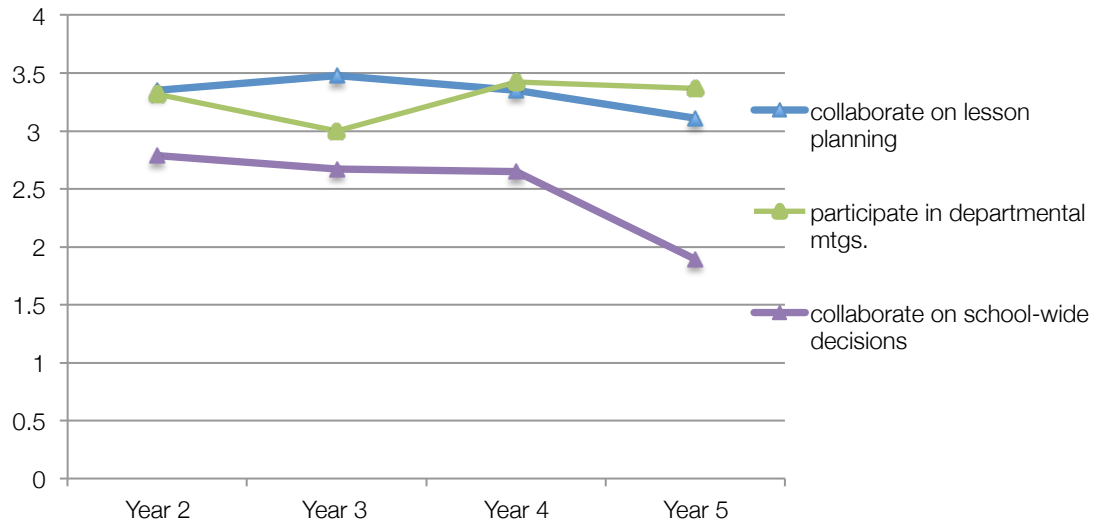
Residents

Residents' primary concerns, in some cases echoed by mentors, relate to how they are perceived and integrated into the school environment: each year, a few residents have noted that they are seen as student teachers, an extra staff person to proctor exams or chaperone field trips, but not a contributing member of the faculty. Survey responses have indicated three slightly different patterns in residents' participation in collaborative activities from year to year: collaboration on lesson planning ticked up based on responses from Year 2 residents, but then dipped slightly. Participation in departmental meetings followed the opposite pattern: levels dropped, then rose to a steady level. Collaboration of school-wide decisions held steady in Years 2, 3, and 4, then declined sharply in Year 5. (See Figure 37.) Our survey data does not reveal a lot about the drop for Cohort 5. The explanation may lie in the subject focus:

¹³ IBO, "A Statistical Portrait of New York City's Public School Teachers."

compared to the math residents who, in previous years, may have been included in decisions, or at least discussions about curricular and assessment changes required to meet the new Common Core math standards and ready students for Algebra 1 assessments, ELA teachers may simply not have had these discussions in 2013–2014 or ELA and special education residents may not have been included.

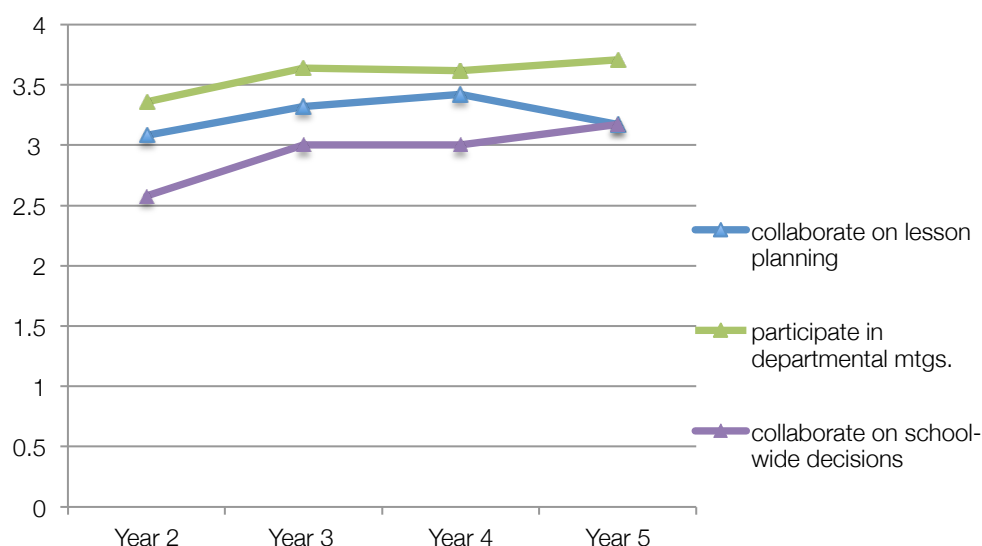
Figure 37. Residents' Participation in Collaborative Activities



Mentors

As noted in Section 1, UTR increasingly focused on selecting mentors interested in working collaboratively and willing to be held accountable for the learning of all students. Unlike residents' survey responses about the opportunities available to them, mentors' reflect more consistent and positive trends across years: collaboration on school-wide decisions and participation have ticked up starting in Year 2. Collaboration on lesson planning dropped in Year 5, which may, again, be a function of the subject focus: because of the inclusion of math residents in Year 2–4, when teachers were more actively engaged in aligning lessons with the Common Core, there may have been more overall collaborative planning than in Year 5. It may also be that, though special education teachers engaged in a collaborative *co-teaching* relationship with residents, they engaged less in lesson planning. (See Figure 38.)

Figure 38. Mentors' Participation in Collaborative Activities



There are other indications that, either through the mentorship itself or through opportunities that UTR schools offer, that mentors have indeed grown professionally. As one Year 5 mentor noted, describing the mentoring experience, “It was nice collaborating with another person who was in the room every single day and recognizing throughout the year that my resident had different strengths as a teacher than I did.... We used our strengths together.” Among the Year 5 mentors, several also assumed leadership roles: two were lead teachers; one, a dean; one, a department chair; and two, a special education or IEP coordinator. There were also school improvement and curriculum team members, inquiry team leaders, and grade level team leaders.

For residents and mentors, any dissatisfaction with their roles concerned how they were defined or understood in the larger project. Residents have at times felt that schools did not understand that they were also graduate students, and that graduate faculty did not understand that they were also teachers. Mentors have felt that though they play a key role in residents' development and the project's success in schools, channels of communication have not been set up so they can share that expertise more broadly and contribute to project-wide as well as school-wide decisions.

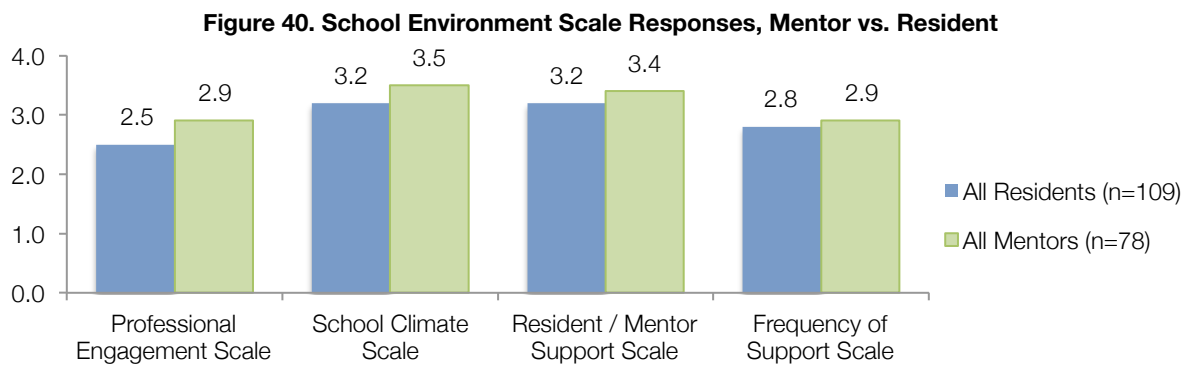
School Environment and Job Satisfaction—Survey Scale Results

For the summative study, we also created three survey scales related to school environment to explore changes over the course of the project, and differences by role, subject area, and school setting (teaching hospital vs. non-teaching hospital; UTR concentration vs. hosting or hiring fewer residents). The scales included:

- *Professional Engagement Scale*, based on items about how often they collaborated with other teachers, on lessons and instructional decisions, and how often they participated in departmental or grade-level meetings.
- *School Climate Scale*, with items from the NYC DOE school survey and the School Assessment Survey (SAS), asking teachers to indicate agreement with statements about school leaders' vision for this school, open communication on school issues, curriculum and assessment alignment, collaboration, and teacher satisfaction.

- *Support Scale*, asking teachers the extent to which they felt supported by their principals, assistant principals, and other teachers.
- *Frequency of Support Scale*, a related scale asking residents how often support activities, such as common planning time and communication from principals or assistant principals, are available to them.

Combined results across four years (Cohorts 2–5) indicate that mentors consistently, though not surprisingly, assign slightly higher ratings than residents. Both groups were most positive about school climate—leaders’ vision, general teacher satisfaction, and curriculum alignment (mentors, $M=3.5$; residents, $M=3.2$). Means were similarly high for levels of support (mentors, $M=3.4$ for mentors; residents, $M=3.2$), in part because teachers indicated strong support from other teachers. Means dropped some for the professional engagement and *frequency* of support scale, which included items related to how *often* teachers had common planning time or regular communication from administrators. (See Figure 40. All items used a four-point agreement or frequency scale, from 1=strongly disagree to 4=strongly agree, or from 1=never/rarely to 4=often/regularly. See Appendix A, p. 71, for individual scale items.)



Source: REA Annual Teacher Surveys

Changes by Project Year

Comparisons of survey responses by year showed similar trajectories for residents and mentors—and improvements in the factors that UTR would likely effect—the sense of professional engagement among teachers and the environment in which they work. Both groups indicated an improvement in school climate, and ratings mirrored each other by 2013–2014: for residents, the mean rose from 2.9 to 3.7; for mentors, from 3.3 to 3.7. There was also an uptick for both groups in the professional engagement scale, with a more notable increase for mentors ($M=2.7$ to 3.3). The factors somewhat out of their control—i.e., the support scales—generally held steady. (See Figures 41 and 42.)

Figure 41. School Environment Scale Responses by Year (Residents)

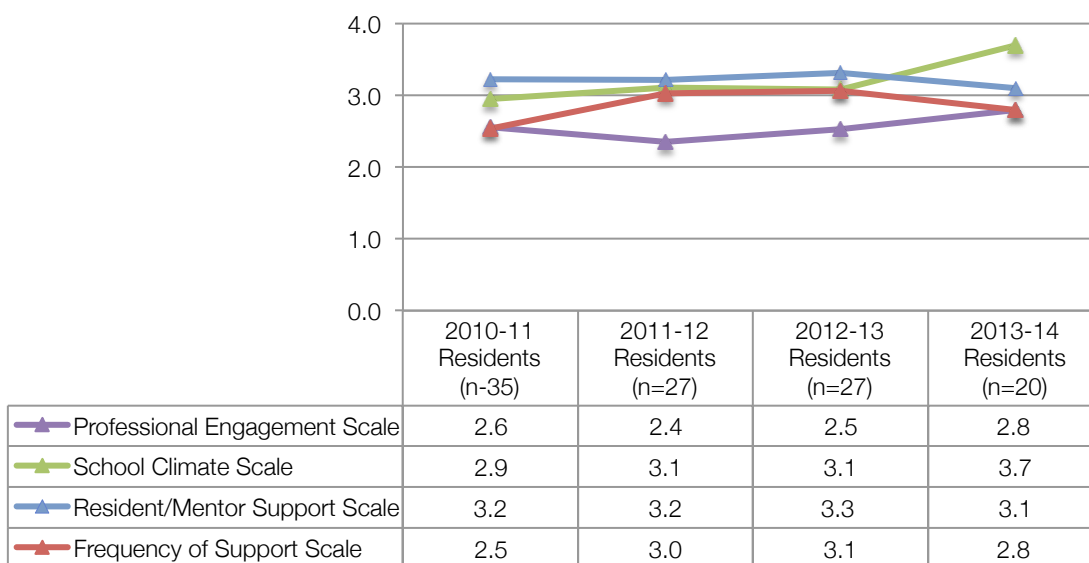
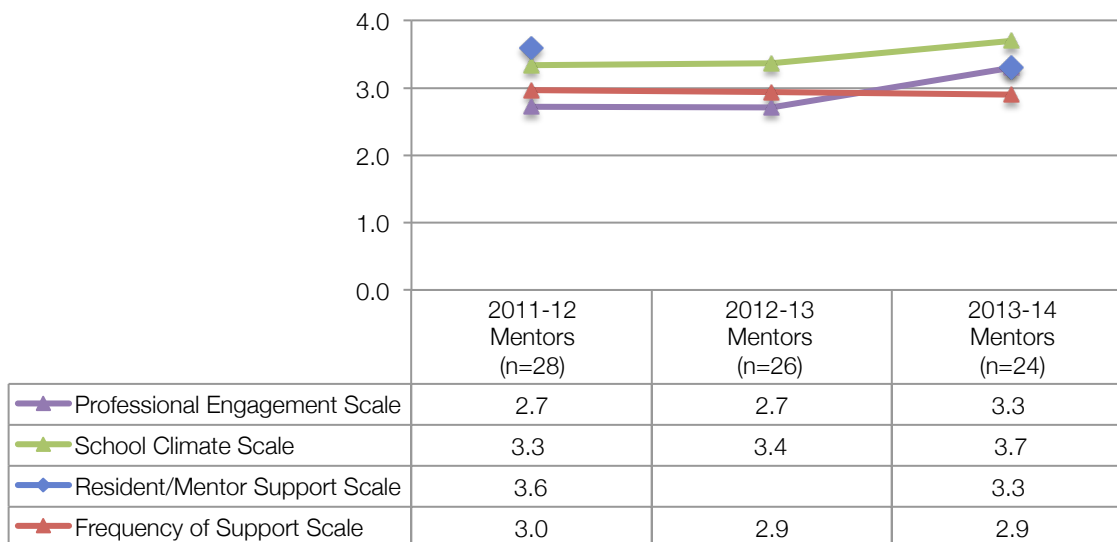


Figure 42. School Environment Scale Responses by Year (Mentors)



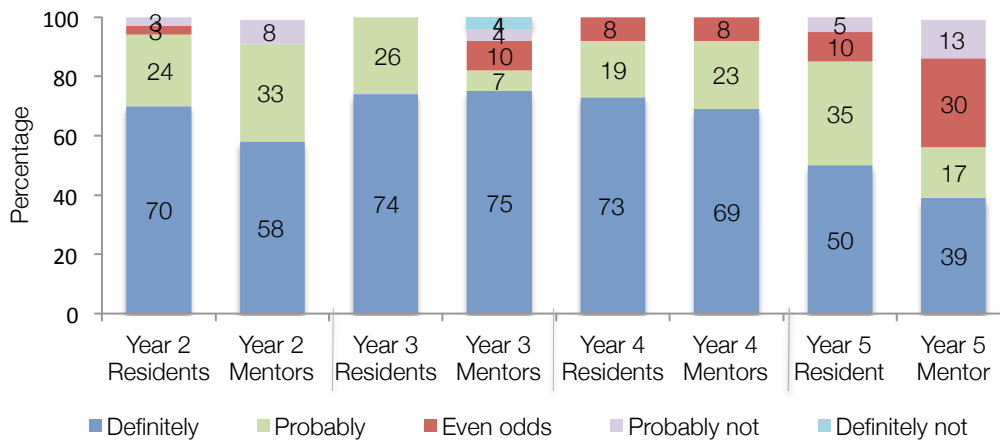
Source: REA Annual Teacher Surveys Scales. Note: There was missing data from mentors in 2012, thus the missing data point in Figure 42.

Teaching as the Chosen Profession

The final measure of job satisfaction is residents' and mentors' responses about their choice to become a teacher. Through Year 4, these figures shifted only slightly: 80–90+% of the residents and mentors said that, if they could start over, they would definitely or probably become a teacher; those who said “definitely” choose the same profession were at 50% or higher.

In Year 5, there were marked shifts, for residents and mentors: half of the residents, and 39% of the mentors said they would definitely become a teacher if they had to do it over. Over a third of residents in the “probably” category put their combined responses on par with previous residents’. Mentors’ responses—at 56% for the combined “definitely”/“probably” ratings were the lowest for all five years, by margins of between 26 and 44 percentage points. (See Figure 43.)

Figure 43. UTR Teachers Thoughts about their Chosen Profession



Teachers did not elaborate on their responses, but it may be that their ambivalence reflect a broader dissatisfaction, and teachers’ sense that respect for them is declining even as expectations—for them and their students—are on the rise. Though the purportedly more rigorous Common Core exams teachers are not an imminent challenge for Cohort 5 ELA and special education teachers, they may be still feeling the heightened stress in schools about new standards and assessments. They are most definitely feeling the pressure of New York’s new teacher evaluation system, and the new Measures of Teacher Practice (MOTP) and school accountability.

One thing that comments from teachers made clear, when asked what factors play the biggest role in retention, is that they see their UTR experience as an antidote, an experience that both prepares them for and provides relief from the pressure of accountability.

Realistic expectations of workload, and clarity around internal accountability. Even more significantly, I think that teachers need to feel VALUED. They are barraged by society, and they need to feel like part of a loyal team. If they struggle, they need to have productive and positive support to grow but remain valued. This year’s mentoring program was a welcome change. –2014 Mentor

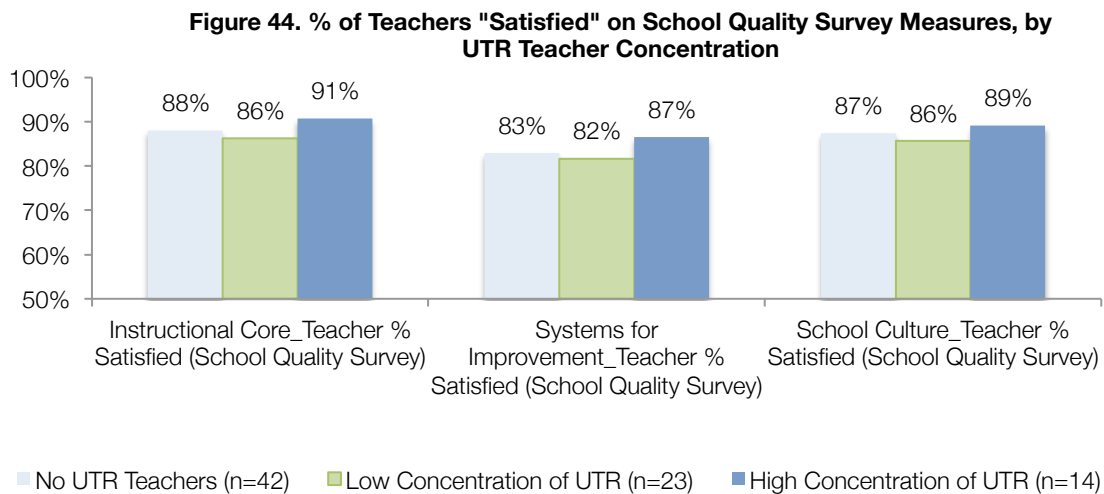
I think the most important factor is acknowledgement from administration that most teachers are working hard at a difficult job....This has been a transformative year for me. I really appreciate the UTR staff and dedication and commitment to excellence and professionalism. –2014 Resident

Are schools with more UTR teachers different?

To explore whether there was a UTR effect—something that might in fact boost teacher morale—we created a UTR concentration index based on the numbers of UTR teachers they had hosted and hired. The UTR concentration index was based upon the "actively teaching" list provided by UTR. Schools with no UTR teachers (not on list) were assigned a concentration of "0"; schools with 1 to 2 UTR teachers, a concentration of 1=low; and schools with 3 or more UTR teachers, a concentration of 2=high. We derived our sample by using the list of schools included in the Regents/Course grade analyses. We conducted multiple comparisons: one looking at job satisfaction, based on publicly available data from the NYC DOE school survey, and comparisons based on other publicly available school quality data, including achievement and progress ratings, graduation rates, and SAT scores.

Job Satisfaction

We found that high percentages of teachers were satisfied, that differences were not dramatic, and that schools with fewer or no UTR teachers looked very similar. However, in all three areas we considered—the instructional core items, systems for improvement, and school culture—schools with the highest UTR concentration had the most satisfied teachers. (See Figure 44.)



Source: Publicly available School Quality data, NYC DOE

Student Performance

We again did not find striking differences, based on the three performance measures. The progress score for the high UTR concentration schools was higher than scores in the other two categories (60 vs. 56 and 54); on the achievement scores, the low concentration schools had the highest score (67 vs. 64). (See Figure 45.) Graduation rates were higher for the UTR schools by a few percentage points (77% and 76%, vs. 74%). Average SAT scores also varied by only a few points. (See Figures 46 and 47.)

Figure 45. Average 2014 "Achievement" and "Progress" Scores from the NYC School Quality Report, by UTR Teacher Concentration

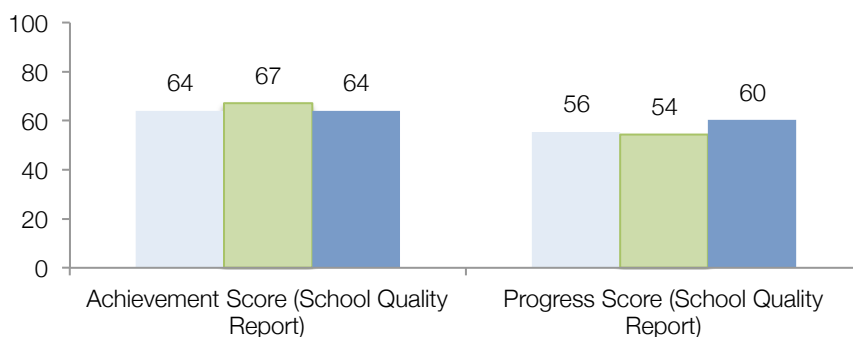


Figure 46. 4-year Graduation Rate for 2013, UTR Teacher Concentration

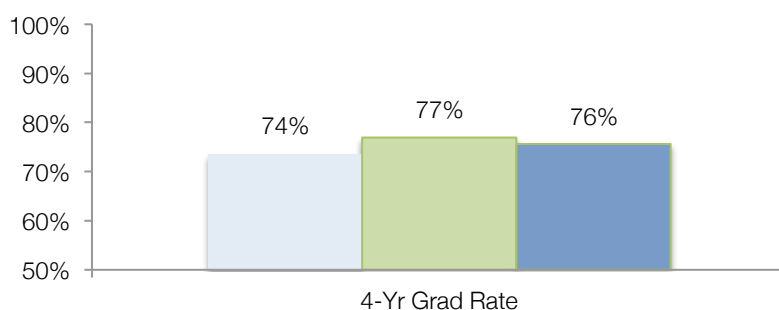
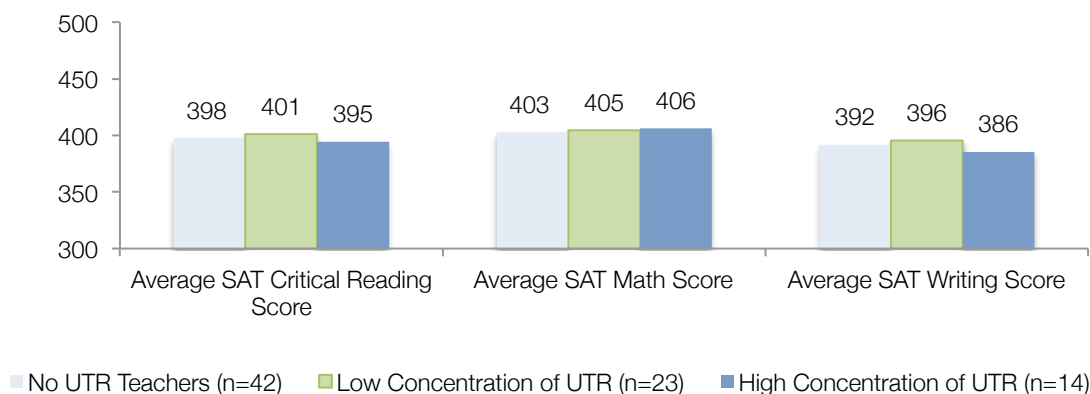


Figure 47. Average 2014 SAT Scores for College Bound Students, by UTR Teacher Concentration



Source: Publicly available School Quality data, NYC DOE

These examinations were in part exploratory, as we looked for ways to gauge the impact of UTR participation, and school commitment to accountable practice and collaborative, evidence-based decision-making. For the supplemental study, and for the new Teacher Quality Partnership grant, we will continue to explore both measures and evidence of success.

CONCLUSIONS

Five years and close to 150 new teachers later, UTR project partners have successfully designed, implemented, and tested a model for teacher preparation—with consistently positive results. UTR has defined “well-prepared,” “effective” teachers as those who are proficient in data-driven inquiry, committed to accountable practice, and equally committed to being part of—and cultivating—this kind of collaborative instructional community in their schools. UTR has also created a set of replicable processes and tools including: a selective recruitment process, which for UTR yielded five diverse cohorts of residents confident in their abilities and committed to urban schools; graduate coursework that integrates content and pedagogy with clinical practice; and an ongoing coaching and monitoring system that guides preparation, sets benchmarks for success, and provides individualized support and intervention plans for residents who need help to succeed.

Final figures are not in, but numbers so far indicate that UTR partners have met their goals, which stipulated that:

- 90% of the graduates to complete the program. To date, 89% have.
- 90% would obtain jobs in high-need schools. Of the 91% of UTR grads currently teaching, 86% are in high-need NYC DOE schools.
- 90% would successfully complete first year of teaching, and 89% would complete their third. 100% of the UTR graduates met the first milestone, and 93% the second—after not just three but four years.

Perhaps the most compelling result from UTR’s first five years is that the focus on data-driven inquiry and assessment has paid off in student performance, especially on high-stakes Regents’ exams.

What may sustain the positive impact on student performance and teacher turnover is the broader impact on schools and the factors that keep teachers there: a positive, supportive school climate that involves teachers in school decisions and recognizes their efforts. Over the years of the UTR program, partners and past participants have built up a network of principals and schools that welcome UTR-trained teachers and share their commitment to data-driven practice. Ongoing capacity-building and induction support can continue to transform teaching and learning and help students succeed.

The task may, however, be getting harder, as schools face the challenges of implementing the Common Core State Standards and making students college and career ready, and as teachers adapt their practice to the Measures of Teacher Quality. New teachers may need further training and support from Hunter and New Visions as the high-stakes exams roll out. While they may be confident in their abilities, more information and instructional strategies that lead to improved performance can reinforce the guidance they receive from their mentors during their residencies and coaches during their induction years—and make them the emissaries who share new strategies with peers.

The 2013-2104 teacher surveys tapped into some ambivalence—not about the UTR model or the accountability it defines—but about a broader accountability. Pleased, confident about residents’ growth and their own, mentors suggest that teacher effectiveness can only be accelerated so fast, especially when the gauges keep changing. What will sustain them and preserve the sense of excellence and professionalism gained through their UTR experience are, in their words, “realistic expectations,” “clarity around internal accountability,” and a sense of being part of “a loyal team.” Having defined effectiveness, UTR may be called upon to redefine and recalibrate it for teachers feeling the mounting pressure of heightened accountability.

References

"Gathering Feedback for Teaching: Combining High-Quality Observations with Study Surveys and Achievement Gains." Bill & Melinda Gates Foundation, 2012. Available online at http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf.

"Measuring UTRU Network Program Impact 2014." Urban Teacher Residency United, April 2014, p. 2. Available at: http://www.utrunited.org/EE_assets/docs/UTRU_Network_Trends_Report_2014.pdf

NYC Teaching Fellows' statistics see https://www.nycteachingfellows.org/about/program_statistics.asp.

"A Statistical Portrait of New York City's Public School Teachers." NYC Independent Budget Office Report, 2014. Available at: <http://www.ibo.nyc.ny.us/iboreports/2014teacherdemographics.pdf>.

"Teachers Matter: Understanding Teachers' Impact on Student Achievement" (2009). Santa Monica, CA: The RAND Corporation. Available at http://www.rand.org/pubs/corporate_pubs/CP693z1-2012-09.html.

"Transforming Teacher Education through Clinical Practice: A National Strategy to Prepare Effective Teachers" (2010). The National Council for Accreditation of Teacher Education. Available at <http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OoqPk%3D&tabid=715>.

Appendices

APPENDIX A. SURVEY SCALES (SECTION 2)

All items used a 4-point scale: (1=not at all prepared and 4=very well prepared; 1=almost never and 4=almost daily/on a regular basis)

Overall Preparedness Scale

How prepared or confident are you to:

- Teach your subject matter
- Handle a range of classroom management or discipline situations
- Motivate and engage students
- Use data to inform instruction
- Differentiate Instruction based on students' needs
- Create assessments that measure short-term growth
- Create assessments that measure long-term growth
- Improve student achievement

Overall Best Practices Scale

How often do you:

- Check for understanding during lessons or use "in the moment" assessments
- Provide opportunities for students to process information by writing or talking
- Involve students in assessing their own learning
- Ask questions that push students toward higher-level thinking
- Check for understanding at the end of the lesson
- Create opportunities for students to work both independently and collaboratively
- Maintain class routines and ensure that discipline problems don't interfere with lessons
- Keep students on task from bell to bell

Supporting Engagement Scale

How prepared are you/how often do you:

- To motivate and engage students
- Provide opportunities for students to process information by writing or talking?
- Ask questions that push students toward higher-level thinking?
- Create opportunities for students to work both independently and collaboratively?

Assessing Learning/Understanding Scale

How prepared are you/how often do you:

- Use data to inform instruction?
- Create assessments that measure short-term growth?
- Create assessments that measure long-term growth?
- Check for understanding during lessons?
- Check for understanding during lessons or use in the moment assessments?
- Check for understanding at the end of the lesson?
- Involve students in assessing their own learning?

Differentiating Instruction Scale

How prepared are you to differentiate instruction based on students' learning needs?

To what extent do your lessons include instruction that is differentiated to meet individual student needs?

Classroom Management Scale

How prepared are you/how often do you:

Handle a range of classroom management or discipline situations

Maintain class routines and ensure that discipline problems don't interfere with lessons

Keep students on task from bell to bell

General Instructional Efficacy Scale

How prepared or confident are you to:

Teach your subject matter

Improve student achievement

Professional Engagement Scale

How often do you engage in the following activities:

Collaborate with other teachers on lessons and classroom planning

Participate in departmental or grade-level meetings

Collaborate with colleagues on school-wide instructional decisions

School Climate Scale

In your school, to what extent do you agree with the following statements:

School leaders communicate a clear vision for this school

School leaders encourage open communication on important school issues

Curriculum, instruction, and assessment are aligned within and across grade levels

School leaders invite teachers to play a meaningful role in setting goals and making important decisions

There is a great deal of collaboration among staff members

Most of my colleagues share my beliefs and values about the central mission of the school

Teachers are recognized for their efforts

Teachers at this school like being here

Teachers are reflective and think critically about their practice

Teachers see themselves as members of a professional learning community

APPENDIX B: CORRELATIONS BETWEEN COURSE GRADES AND REGENTS (SECTION 3.)

Table 1. Correlation between Cohort 5 Course Grades and Regents Exam Scores

General Subject Area	Overall			UTR			Non-UTR		
	N	r	Sig.	N	r	Sig.	N	R	Sig.
Math (regular Regents)	186	.70	p<.001	50	.68	p<.001	136	.71	p<.001

Table 2. Correlation between Cohort 4 Course Grades and Regents Exam Scores

General Subject Area	Overall			UTR			Non-UTR		
	N	r	Sig.	N	r	Sig.	N	R	Sig.
Math (regular Regents)	487	.52	p<.001	190	.58	p<.001	297	.48	p<.001
Math (CC)	212	.61	p<.001	142	.68	p<.001	70	.52	p<.001
Science	1045	.52	p<.001	609	.53	p<.001	436	.52	p<.001

Table 3. Correlation between Cohort 3 Course Grades and Regents Exam Scores

General Subject Area	Overall			UTR			Non-UTR		
	N	r	Sig.	N	r	Sig.	N	R	Sig.
Math (regular Regents)	473	.49	p<.001	208	.57	p<.001	265	.44	p<.001
Math (CC)	138	.53	p<.001	41	.35	p<.05	97	.56	p<.001
Science	287	.54	p<.001	116	.41	p<.001	171	.56	p<.001

Table 4. Correlation between Cohort 2 Course Grades and Regents Exam Scores

General Subject Area	Overall			UTR			Non-UTR		
	N	r	Sig.	N	r	Sig.	N	R	Sig.
English (regular Regents)	389	.45	p<.001	231	.55	p<.001	158	.31	p<.001
English (CC)	157	.47	p<.001	97	.54	p<.001	60	.47	p<.001
Math (regular Regents)	118	.58	p<.001	101	.63	p<.001	17	.54	p<.001
Math (CC)	95	.35	p<.001	95	.35	p<.001	NA	NA	NA
Science	639	.70	p<.001	302	.69	p<.001	337	.67	p<.001
Global Studies / History	103	.65	p<.001	26	.74	p<.001	77	.62	p<.001

Table 5. Correlation between Cohort 1 Course Grades and Regents Exam Scores

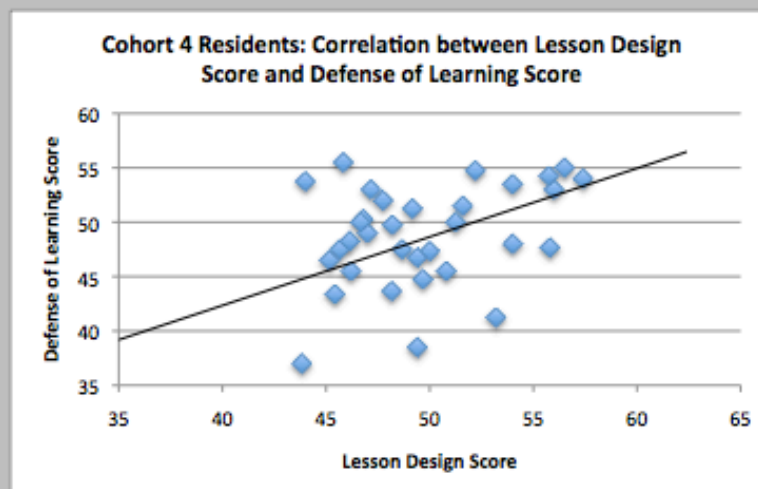
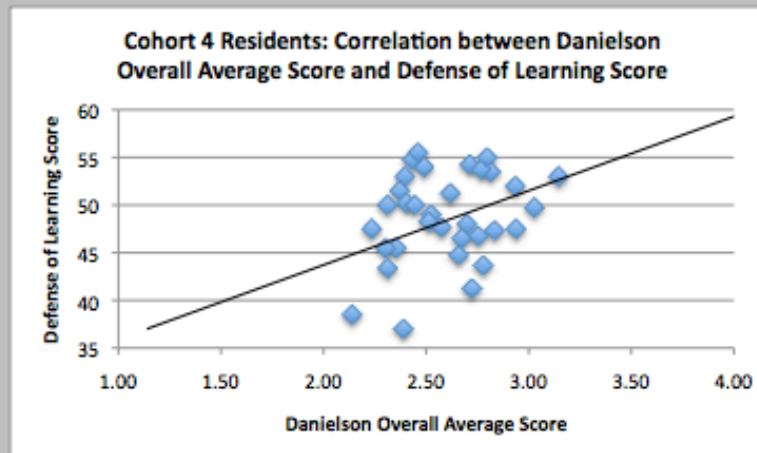
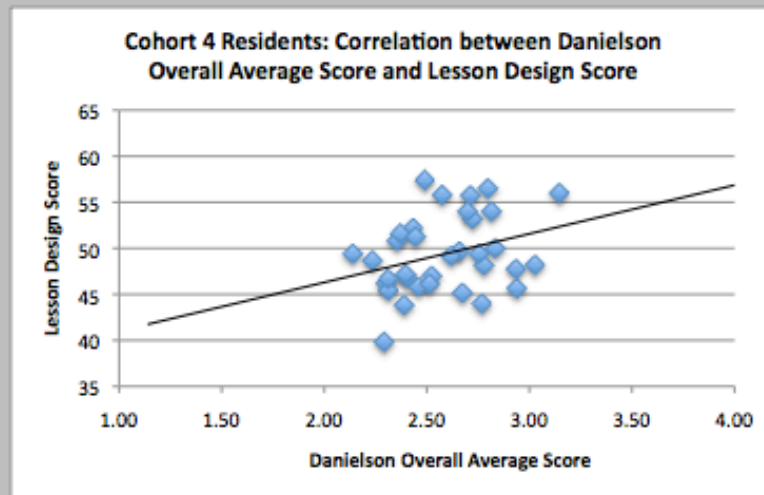
General Subject Area	Overall			UTR			Non-UTR		
	N	r	Sig.	N	r	Sig.	N	R	Sig.
Math (regular Regents)	59	.64	p<.001	21	.71	p<.001	38	.60	p<.001
Math (CC)	64	.43	p<.001	23	.30	ns	41	.50	p<.001
Science	302	.41	p<.001	179	.44	p<.001	123	.63	p<.001

**Table 6. Correlation between Course Grades and Regents Exam Scores
across all GRADUATES (Cohorts 1 – 4)**

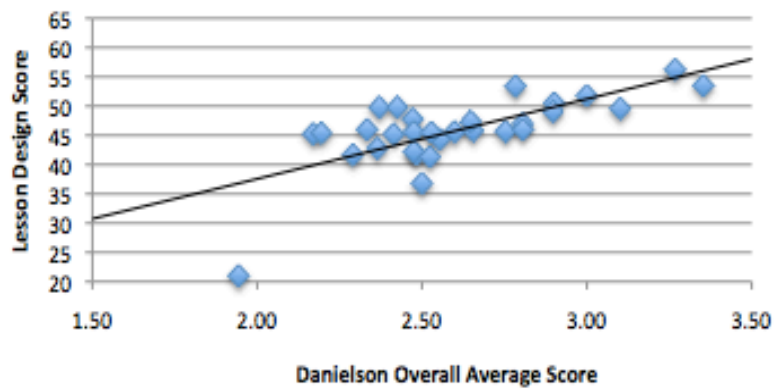
General Subject Area	Overall			UTR			Non-UTR		
	N	r	Sig.	N	r	Sig.	N	R	Sig.
English (regular Regents)	1137	.52	p<.001	520	.55	p<.001	617	.50	p<.001
English (CC)	509	.51	p<.001	301	.51	p<.001	208	.52	p<.001
Math (regular Regents)	389	.45	p<.001	231	.55	p<.001	158	.31	p<.001
Math (CC)	157	.47	p<.001	97	.54	p<.001	60	.47	p<.001
Science	2273	.54	p<.001	1206	.53	p<.001	1067	.54	p<.001
Global Studies / History	103	.65	p<.001	26	.74	p<.001	77	.62	p<.001

APPENDIX C: SCATTERPLOTS, CORRELATIONS BETWEEN MEASURES, COHORTS 3 & 4
(SECTION 4.)

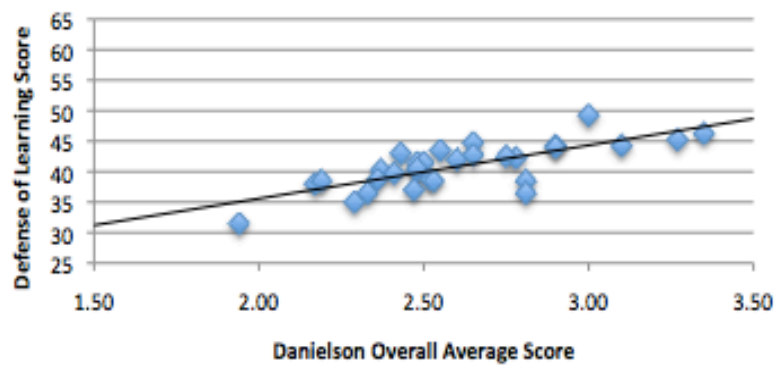
Correlations between Types of Ratings



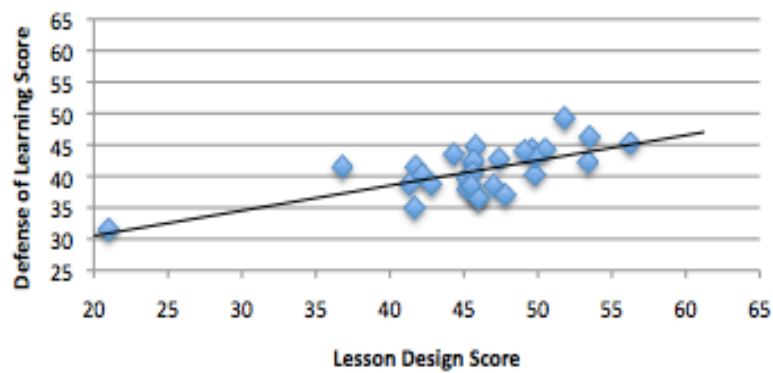
Cohort 3 Residents: Correlation between Danielson Overall Average Score and Lesson Design Score



Cohort 3 Residents: Correlation between Danielson Overall Average Score and Defense of Learning Score



Cohort 3 Residents: Correlation between Lesson Design Score and Defense of Learning Score



APPENDIX D. CORRELATIONS FOR TEACHER QUALITY RATINGS, BY ASSESSMENT TOOL/SCORE, COHORT, STUDENT OUTCOME MEASURE, AND YEAR (SECTION 5.)

The table below shows the correlation coefficients for all the analyses examining links between residents' performance and student outcomes. The areas shaded in blue indicate statistically significant associations; those in gray are approaching significance, with r values at .5 or higher. The yellow shadings indicate statistically significant negative correlations.

	Cohort 3				Cohort 4				Cohort 5		MASTER	
	Residency SY2011-2012		Induction SY2013-2014		Residency SY2012-2013		Induction SY2013-2014		Residency SY2013-2014		Residency SY2013-2014	
	REGENT S	GRADES	REGENT S	GRADES	REGENTS	GRADES	REGENT S	GRADES	GRADES	Attend.	REGENT S	GRADES
Danielson Designing Coherent Instruction	-0.59	0.16	0.53	0.47	0.22	0.07	0.10	0.28	0.4	0.28	0.39	0.25
Danielson Establish a Culture of Learning	-0.60	0.30	0.41	0.66	0.68*	0.46	-0.03	0.25	0.58*	0.24	0.61*	.63**
Danielson Managing Student Behavior	-0.58	-0.01	0.32	0.50	0.69*	0.50*	0.22	-0.04	0.34	0.07	0.52	.60**
Danielson Using Questioning / Prompts and Discussions	-0.61	-0.12	0.65	0.35	0.25	0.03	-0.72*	0.21	0.44	0.26	0.57	0.12
Danielson Engaging Students	-0.46	-0.40	0.74*	0.39	0.40	0.48	-0.25	0.00	0.39	0.19	0.28	0.49*
Danielson Using Assessment	-0.16	-0.17	0.65	0.20	0.32	0.13	-0.19	0.18	0.08	-0.08	0.29	0.21
Danielson Average Score	-0.72	-0.05	0.64	0.51	0.58	0.37	-0.14	0.17	0.47	0.26	0.54	0.44
Lesson Design	0.36	0.31	0.28	0.46	0.15	-0.03	-0.19	0.42	0.42	0.57*	0.70	0.09
Unit Design	0.57	-0.08	0.69	0.37	0.01	0.18	-0.19	0.53	-0.02	-0.15	0.19	-0.22
Professionalism	0.22	-0.80*	0.05	-0.45	0.42	0.30	0.51	0.56	0.16	0.37	0.21	0.34
Defense of Learning	-0.31	0.31	0.59	0.56	0.27	0.37	0.36	0.74*	0.56*	0.64*	0.24	-0.17
Resident Rating Composite (avg of % Dan, LD, UD, Prof, DoL)	0.15	-0.33	0.58	0.20	0.38	0.36	0.06	0.65*	0.15	0.21	0.40	0.18
Resident Rating (high, med, low)			0.35	-0.09	0.46	0.36	-0.06	0.58*	0.36		0.33	0.21