Accountability in Teacher Education: Ecological Analyses of VAM Effects
Accountability in Teacher Education:
Ecological Analyses of VAM Effects

A Paper Presented at the Annual Meeting of the
American Educational Research Association
May 2010   Denver, Colorado

Robert Tobias, Barbara Hummel-Rossi, Jane Ashdown,
Ognjen Simic and Kimberly K. Woo

New York University

Jane Ashdown is now at Adelphi University

CRTL Research Paper Series

RP-0410-01

April 2010

Center for Research on Teaching and Learning
Department of Teaching and Learning
The Steinhardt School of Culture, Education, and Human Development
New York University

© Copyright 2010 by the Center for Research on Teaching and Learning
ABSTRACT

Using value-added methods (VAM), a sample of our teacher education graduates (N=191) teaching in New York City public schools was found to have a positive impact on their pupils’ academic achievement beyond predicted gains for these pupils. To examine the stability of these findings, more recent years of program graduates were studied (N=213). VAM analysis revealed further positive findings. However, a sub sample of these graduates for whom student teaching performance assessments were available (N=17) showed a zero or small negative relation between their performance as student teachers and the academic achievement of their pupils as revealed through the VAM analysis. Taking an ecological perspective, we report on these VAM findings and on an intensive case study of the 17 graduates.

OBJECTIVES

As researchers from a large private university in an urban context, we have been conducting a multi-year, multi-phase study with the objective of developing an accountability system linking university-based teacher preparation with the performance of the program graduates. A longitudinal design has been employed that uses value-added methods (VAM). Positive findings have emerged regarding graduates’ (N=191) impact on their pupils’ math and English language arts (ELA) performance as measured by standardized tests, thus providing preliminary evidence of the effectiveness of the teacher education program on teacher quality.
This paper reports on an extension to this research study. We wanted to gain further understanding of the consistency and stability of our initial findings, to link the findings to a pre-service measure of teacher quality, and to study how ecological factors affect measures of teacher quality. VAM have been scrutinized and their assumptions questioned when they are used in accountability systems for teacher education and teacher performance. Mindful of these concerns, we present herein additional VAM findings using an augmented database of teacher education program graduates (N=213) matched to pupil performance on math and ELA tests and an analysis that compares graduates’ VAM effects based on pupil testing with an assessment of their pedagogical proficiency using a student teacher observation scale. Since these two measures of teacher quality are situated in different contexts and different stages of teacher development, we embarked on an intensive study of a small sub-sample of graduates (N=17) using an ecological framework to study the complexity of the relations among the contexts of teacher preparation programs and the schools in which graduates begin teaching and the assessment of their performance. The immediate need for the ecological approach emerged from the observation of an inverse relation with VAM effects and the measure of student teaching proficiency. This intriguing finding serves as the basis for exploring multiple variables associated with these graduates’ teacher education experiences and with their initial teaching positions with a view to generating hypotheses about the relations of interest among these variables. We discuss the implications of these findings for accountability in teacher education.
PERSPECTIVE

Evidence from research demonstrates a strong teacher effect on pupil achievement (Ballou, Sanders, & Wright, 2004; Rivkin, Hanushek, & Kain, 2000; Wright, Horn, & Sanders, 1997), and a cumulative effect on pupil achievement of both good and bad performing teachers (Mendro, 1998; Sanders & Horn, 1998). In order to maximize pupils’ opportunities to learn with highly effective teachers, investigations of school and teaching quality have focused on isolating specific variables such as licensure exam performance and number of years teaching with regard to their predictive properties. These investigations have extended backwards into examinations of the characteristics of teacher education programs as they relate to the teaching effectiveness of the graduates of those programs. In teacher education research, attention to empirical measures of success in terms of evidence of pupil learning has spurred interest in value-added methods (Noell & Burns, 2006; Wilson & Youngs, 2005).

However, VAM despite a history stretching back to 1976 (Bryk & Weisberg, 1976), have been the recent subject of scrutiny by researchers concerned about methodological issues (Amrein-Beardsely, 2008), as well as data availability and management issues (Noell & Burns, 2006). In addition, while Fallon (2006) has argued that evidence of pupil learning should be the primary indicator of high quality teaching, others have raised concerns about reliance on standardized test performance as the sole indicator of that quality (Cochran-Smith, 2006). In practice, putting VAM analyses to work in the service of organizational learning and teacher education program improvement requires robust feedback systems that can turn such analyses into usable
knowledge that can be interpreted meaningfully and acted upon by appropriate stakeholders (X, Y, & Z, 2009).

Although VAM offer important opportunities to more fully understand the relation between pupil achievement and teaching effectiveness, measurement and technical challenges remain. An ecological perspective may offer a broader context within which to interpret findings from our study of our teacher education graduates. Lee (2008) argues for an ecological framework in the study of learning and human development that would position researchers to address the full complexity and diversity of human functioning. Of relevance to our own study is Lee’s synthesis of work in the natural sciences and the study of the brain as a basis for an ecological perspective. In popular parlance an “eco-system” implies an interdependence of various biological, psychological, and cultural systems, and this contrasts with a focus on a single relation or isolated source of information in understanding observed phenomena. In light of the conference theme, we take an ecological perspective on the complexities of teaching and learning. We examine our VAM analyses in relation to student teaching performance data and contextual data available for a sub-sample of 17 teacher education graduates. We describe our decision to develop an intensive case study of these graduates.

Our research questions are as follows;

1. Extending VAM analyses from our prior research study (N=191) with VAM analyses using a newly augmented database of teacher education program graduates (N=213), what is the evidence for stability of teacher effects on pupil performance?
2. For a sub sample of these graduates, how do ecological differences in their student-teaching and in-service teaching contexts affect the relation between their value-added
teacher performance indicators and their student teaching performance measures? The contextual variables will include demographic characteristics, school assignment characteristics, academic performance and other ecological variables.

3. What are the implications of these findings for accountability efforts in teacher education?

**METHODS**

**Participants**

Teacher education program graduates from September 2005 to 2008 were matched to a file provided by the New York City Department of Education. Access to this file was part of a cooperative agreement and had been the basis of the VAM analyses conducted for our earlier study. The match of graduates yielded VAM data for 213 cases teaching in grades 4 to 8 in New York City public schools. This database of graduates was matched against VAM estimates of teacher effectiveness based on pupil achievement data in ELA and math standardized tests. Testing in mathematics and ELA is mandated by the New York State Education Department (NYSED). Of these 213 graduates, 183 (86%) were female and 171 (80%) had earned a Master’s degree. Most of the graduates, 148 (69%), were White, while the remainder were African-American (11%), Hispanic (10%), and Asian-American (9%). All of the graduates taught at the elementary and/or middle school level in grades 4 to 8. Value-added estimates in ELA were available for all of the sample and in math for 126 (59%) of the graduates.

The sub sample for the second research question, an in-depth small-sample comparison of VAM estimates with evaluations of student teaching performance and a variety of other demographic and performance measures consisted of 17 graduates. Of
these 17 graduates, 14 were female (82%), 14 were White (82%), and 12 (71%) had graduated with Master’s degrees.

**Instruments**

**Achievement Tests:** Pupils in NYC public schools grades 3 to 8 participate in the New York State testing program in math and ELA. These tests are aligned with NYSED learning standards and have good internal consistency reliability. Standardized test scores in math and English Language Arts for grades 3 through 8 were used as the dependent variables and baseline covariates in the VAM analyses.

**Student Teaching Measure:** In order to situate the findings from the VAM analyses in a more complex, interpretive environment, student teaching performance data were a source of comparison. During student teaching some of the teacher education program graduates in our database were observed by student teaching supervisors who measured their classroom teaching skills using the Domain Referenced Student Teacher Observation Scale-Revised (DRSTOS-R). The DRSTO-R is based on the work of Danielson (1996); its 20 items (expanded to 21 items in 2009) measure four domains of teaching practice: planning and preparation, classroom environment, instruction, and professional responsibilities. Supervisors who use the DRSTOS-R are required to participate in a training session and achieve acceptable inter rater agreement. Evidence accumulated at our research center supports the instrument’s inter rater agreement, internal consistency reliability, and construct validity.

**Analyses:**

VAM were employed to estimate the impact of the graduates on their pupil’s reading and mathematics achievement. By inference, part of the teacher effect was
attributable to their pre-service teacher education program, although the goal of this study was not to isolate this program effect. Analyses of value-added effects were performed with data on 213 graduates using a value-added model developed by the NYCDOE (2008). The analyses employed a four-level hierarchical model of students within classrooms within teachers within schools, using multiple control variables, and generating a series of VAM estimates, of which we focus on two: an estimate positioning the teacher in context of all teachers in the NYCDOE system, and an estimate of value-added in context of a peer group of teachers with the same years of teaching experience and working with similar students. These estimates included one-year and three-year averages, the latter if enough data points were available.

**Stability and Consistency of VAM Estimates:** Analyses were conducted to investigate two sets of relationships that addressed the stability and consistency of the VAM estimates for making inferences about the teaching effectiveness of NYU graduates. The first set of analyses assessed the relationship between VAM estimates in ELA and math and years of teaching experience, as estimated from the date of degree completion for the full sample of graduates with VAM data, using regression analysis and graphs plotting mean VAM estimates against year of graduation. These analyses assessed the replicability of a key finding from the previous study. The second set of analyses used correlational analysis to assess the within-subjects consistency of VAM estimates longitudinally over three years and within a single year for graduates with two VAM estimates from different schools or different grades.

**Ecological Analysis:** The ecological analysis on the sub sample of 17 graduates to gain greater understanding of the VAM effects is still underway and is using a variety of
descriptive statistical methods as well as qualitative case study procedures. Initial investigations have tapped into a number of different databases and explored relations between graduates’ VAM estimates and their academic performance, licensure examination test scores, characteristics of student teaching schools, characteristics of current schools where teaching, nature of academic program, grade and subject area teaching, and numerous other variables. Relationships among the pre-service and in-serve measures of teacher effectiveness and the relationships between context and the effectiveness measures were explored using trend analysis and detailed case studies have been conducted to exemplify the complexity of these relations.

RESULTS

VAM Stability Findings

In order to address our first research question concerning the stability of VAM findings from our earlier study, we conducted two sets of analyses on the VAM effects for each set of matched cases (N=191; N=213) using ELA and math scores as the dependent variables. First, we assessed the nature and degree of the relationships between VAM estimates and years of teaching experience, which was shown to be consistently related to both ELA and math VAM estimates in our previous study. Since reliable information on years of teaching experience was not available to us for the 2008 sample, we have carried out a cross-sectional analysis of classes graduating from NYU over the years (from 2001 to 2006) who have subsequently found teaching jobs in New Your City public schools, with respect to their reported value-added estimates in years 2006, 2007, and 2008. The average VAM values for ELA for each graduating class are
shown in Figures 1-3, and for math in Figures 4-6. All of the six figures show approximately the same trend, albeit with some disruptions: average VAM is greater for teachers of earlier graduating classes who have spent more time teaching in the NYC public schools, and smaller for more recent graduates who have fewer years of teaching experience. It would appear that the findings from our previous study, indicating a positive and statistically significant relationship between VAM estimates and teaching experience, have been confirmed. Notably, this relationship emerges from the data even though the graduating class is only an approximation of years of experience (i.e., some teachers who have graduated recently had been teaching for several years using their previous credentials). Accordingly, the VAM estimates appeared to show stability with respect to general trends in their relationships to an important descriptive variable for our graduates.

Figure 1. Mean VAM effects for NYU graduates on their pupils’ 2006 ELA achievement test scores by year of NYU graduation

![Graph showing mean VAM effects](image)
Figure 2. Mean VAM effects for NYU graduates on their pupils’ 2007 ELA achievement test scores by year of NYU graduation

Figure 3. Mean VAM effects for NYU graduates on their pupils’ 2008 ELA achievement test scores by year of NYU graduation
Figure 4. Mean VAM effects for NYU graduates on their pupils' 2006 Math achievement test scores by year of NYU graduation

Figure 5. Mean VAM effects for NYU graduates on their pupils' 2007 Math achievement test scores by year of NYU graduation
The second set of stability investigations used correlational analysis to assess the within-subjects consistency of VAM effects over three years and within the same year. The results are summarized in Table 1, including the analysis of one-year estimates (for 06, 07 and 08) and cumulative three-year estimates (06-08) made available for 2008, as well as the comparison of 2008 cumulative value-added estimates with the 2007 cumulative estimates reported in our previous study. Correlations are reported separately for ELA and math estimates, and it is apparent that correlation coefficients follow a similar pattern, with math correlations being somewhat larger, but overall rather low for one-year estimates. Correlation of VAM estimates separated by two years is non-existent (.01 fro ELA, .03 for math), while VAM values from consecutive years are in .2 range for ELA (.18 and .23) and .3 range for math (.34 and .35). Correlations of math estimates from consecutive years are the only correlations of one-year VAM values reaching statistical significance.

Comparisons of three-year estimates from two consecutive school years yields statistically significant correlations of moderate size, .46 for ELA and .39 for math. This
is an expected finding, since these estimates are mutually related by design: both sets of three-year estimates are presumed to incorporate the data from years 06 and 07, i.e. to have about two-thirds of common variability. However, the difference between these two sets of cumulative estimates may be attributed to differences in VAM methodology used from one year to another.

The last set of analyses reported in Table 1 were performed on a subset of teachers who have taught in different grades or in different schools during the school year 07-08, and hence for whom different sets of independent VAM estimates were generated using the same methodology. These correlations are rather small (.21 for ELA and .25 for math), and not statistically significant. It must be noted that these were obtained on rather small subsets of teachers (48 for ELA and 36 for math) for whom these data were available.

Table 1. Correlations of value-added modeling (VAM) estimates for NYU graduates teaching in NYC public schools (2006-2008)

<table>
<thead>
<tr>
<th>Year of VAM</th>
<th>ELA</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-year VAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r (N)</td>
<td>2006 (48)</td>
<td>2007 (39)</td>
</tr>
<tr>
<td>0.19</td>
<td>.01</td>
<td>.34*</td>
</tr>
<tr>
<td>2007</td>
<td>.23</td>
<td>.39*</td>
</tr>
<tr>
<td>2008</td>
<td>.35*</td>
<td></td>
</tr>
<tr>
<td>Three-year VAM</td>
<td>2008 (06-08)</td>
<td>2008 (06-08)</td>
</tr>
<tr>
<td>r (N)</td>
<td>2007 (05-07)</td>
<td>(113)</td>
</tr>
<tr>
<td>.46*</td>
<td>.39*</td>
<td></td>
</tr>
<tr>
<td>Three-year VAM for teachers teaching in different grade/school during the same year</td>
<td>2008 (06-08)</td>
<td>2008 (06-08)</td>
</tr>
<tr>
<td>r (N)</td>
<td>2008 (06-8)</td>
<td>(48)</td>
</tr>
<tr>
<td>.21</td>
<td>.25</td>
<td></td>
</tr>
</tbody>
</table>

* p< .05
In summary, the consistency of the relationships between VAM estimates and a proxy measure for years of experience suggests that the VAM data may have utility for revealing broad, general trends in the effectiveness of teachers who have certain common characteristics. However, the low consistency of single-year estimates, lower-than-expected consistency of three-year estimates in consecutive years, and low consistency of estimates for teachers working in different environments in the same school year, suggests that caution must be exercised in using these VAM estimates alone for making high-stakes decisions about teacher effectiveness.

Ecological Analysis of VAM Effects

The second research question focused on a subsample of 17 graduates and addressed how ecological differences in the graduates' student-teaching and in-service teaching contexts affect the relation between their value-added teacher performance indicators and their student teaching performance measures. This analysis also looked at the interactions among a comprehensive array of pre-service cognitive, affective, and performance indicators and the success of graduates in their first teaching experiences. To examine this question we fortunately were able to examine numerous University and NYCDOE data bases that included information on the graduates' demographics, master's degree course grades including student teaching grades, and undergraduate GPA’s, teaching certification test scores, student teaching placement school demographics, self-evaluations of the extent to which their student-teaching experiences helped them to develop their teaching knowledge and skills, DRSTOS-R scores, and current teaching
placements and status, including demographics of these schools. These data sources provided a wealth of information on each graduate.

Initially, ELA and math VAM estimates from two different models (one that does not control for teacher experience and one that does) were correlated with DRSTOS-R scores for the sub-sample of 17 cases. Pearson product-moment correlation coefficients, while not statistically significant, were all negative, -.30 and -.35 for uncontrolled and controlled ELA VAM estimates, respectively, and -.10 and -.13 for uncontrolled and controlled math VAM estimates, respectively. Inspection of scatter diagrams revealed a much more complex picture than that shown by the correlation coefficients; for example, for ELA there was one outlier graduate with a high DRSTOS-R score and a substantially negative VAM effect. This same outlier graduate had among the highest positive math VAM effects for the group of 17. Clearly, this was an indication that a much more in-depth analysis of the data was required.

An ecological perspective compels consideration of adaptability and survival (Lee, 2008). Therefore, the cases first were first categorized into those in which the graduates showed some success in induction, as indicated by retention in the same schools in which they were hired (n = 11) versus those who had evinced some problems in adapting to their new teaching assignments (n = 6). The latter included graduates who were terminated from their positions (n = 2), were on leave (n = 2), or transferred to a different NYC public school (n = 2). While there are not specific data available as to why these six graduates were not continuing their service in the same schools, new-teacher retention is a serious problem in NYC public schools, as it is in inner-city schools in general, and attrition frequently signifies problems in adapting to the rigors of
induction into the profession. We conducted an in-depth analysis of a wide array of evidence for these 17 graduates, including their undergraduate and graduate transcripts, New York State Teacher Certification Exam (NYSTCE) scores, self-reports of the quality of their clinical (student teaching) experiences, as reported in the End-of-Tem Feedback Questionnaire (ETFQ), DRSTOS-R scores of pedagogical proficiency demonstrated during student teacher placements, VAM scores, and demographic data describing the student teacher placement schools and the schools which first hired them.

The case-study analyses revealed a complex interaction among the various data elements in determining the career trajectory of these 17 graduates. The selected case studies reported in the appendix provide a window into the unique interplay of pre-service and in-service experiences of each graduate. Nevertheless, several categories of underlying factors emerge as typical of the forces that are related to the early career success of the graduates.

The first category involves the degree of similarity in the demographics of the schools in which the graduates had their clinical experiences and the schools in which they were first hired. For example, one graduate completed student teaching initially in a school with an English Language Learner (ELL) population of 1%, followed by a second student teaching in a school with 8% ELL, followed by a teaching position in a school with 16% ELL. Another graduate had two student teaching placements in schools with low Title 1 eligible populations of 12% and 7% and special education populations of 12% and 13%, respectively, followed by a teaching position in a school with a Title 1 population of 94% and a special education population of 39%. The ecological perspective raises the question of the extent to which the graduates were accumulating the necessary
skill sets to thrive in these first teaching positions. In three of the six cases of disrupted teaching continuity, there was a demographic mismatch between student-teaching and in-service schools; this was true for only one of the 11 schools in which the graduates were continuing to serve at the time this report was written. In all cases of demographic mismatch, the in-service school had more high needs students (i.e., English language learners, special education students, minority students, and Title 1-eligible students) than the schools in which they student taught. Case studies 12 and 16 in the appendix involve this factor.

The second category involves cases where there is clear and compelling pre-service evidence that the graduate has or does not have the skills and knowledge that hold promise for success in teaching. The evidence comes from a variety of indicators, including undergraduate and graduate transcripts, quality of undergraduate institution and, NYSTCE test scores. This category is typified by Case Study 2 (see appendix), a 2006 graduate of the Masters in English Education program who had a mixed undergraduate record and failed the English content specialty test twice.

The third category is related to the perceived quality of the student-teaching experiences, as reported by graduates through the ETFQ. Two graduates, who are continuing to serve in their school of first hire, reported on the ETFQ that they experienced consistent quality with regards to their cooperating classroom teacher, their University supervisor, and the seminar that accompanied their school-based placement. One of these graduates currently is teaching English in a middle school serving a Title 1 eligible population of 85%, a special education population of 28%, and with a School Environment rating of B. Both student teaching placements for this graduate were in the
same school, a school with a low Title 1 population (15%), a relatively low special education population of 7%, and an Environment rating of A. The second graduate is currently teaching special education in an elementary school having had four student teaching placements and experience in schools with varying demographics in terms of service to Title 1 and special education students.

Eight of the graduates had at least one negative experience with regard to one of the three components of their student teaching experience – the quality of their cooperating teacher, the quality of the university supervisor, and the quality of an accompanying seminar taught by the university supervisor. One issue the examination of these cases raises is the interrelation between initial and final student teaching experiences and how these experiences influence and shape the graduates’ developing pedagogical skills. Negative experiences concerning the cooperating teacher included a lack of feedback, poor organization, lack of knowledge, and lack of availability. Regarding the university supervisor, these negative experiences included lack of availability, lack of appropriate expertise in the field of special education or the academic content area, and lack of relevance in topics for the student teaching seminar. Two of the cases with disrupted teaching careers, one who was terminated (Case 2 in the appendix) and one who is on leave (Case 3 in the appendix), reported extremely negative student-teaching experiences.

The fourth category pertains to the developing pedagogical proficiency of the graduates as measured by the DRSTOS-R. The DRSTOS-R is the culminating pre-service measure of teaching effectiveness. It assesses the extent to which student teachers integrate and apply their content knowledge, pedagogical content knowledge,
pedagogical knowledge, and knowledge of learners and learning in actual teaching
classroom teaching situations. The case-study data suggest that this measure of
pedagogical proficiency is more than just the sum of the component skills. Although
there is a relationship between DRSTOS-R scores and ETFQ data, the relationship is only
moderate. DRSTOS-R scores may be predictive of problems in early teaching careers for
graduates who show strong positive evidence on other indicators. Such an example is
Case Study 3 (see appendix), a beginning teacher currently on leave who had a strong
academic record but low (Mean = 2.75) DRSTOS-R scores.

It is important to note that while the above typology is a powerful heuristic for
discussing consistent themes in the case-study data, most cases do not simply fall into
one category or another. Virtually all case studies evince a complex interaction among
these four thematic categories leading to nuanced career trajectories that defy simple
explication and explication.

**DISCUSSION**

In summary, the consistency of the relationships between VAM estimates
and a proxy measure for years of experience suggests that the VAM data may have utility
for revealing broad, general trends in the effectiveness of teachers who have certain
common characteristics. However, the low consistency of single-year estimates, lower-
than-expected consistency of three-year estimates in consecutive years, and low
consistency of estimates for teachers working in different environments in the same
school year, suggests that caution must be exercised in using these VAM estimates alone
for making high-stakes decisions about teacher effectiveness. Moreover, the in-depth
analysis of a sub-sample of graduates to examine ecological differences in their student-
teaching contexts, demographic characteristics, academic performance, and student teaching evaluations in relation to their VAM indicators revealed no strong correlates of VAM performance. However, we were fortunate to be able to access a fairly large number of variables in this examination and found that four categories of thematic factors appeared to act in complex ways to determine the career trajectories of a sample of 17 graduates. An ecological perspective did lead to questions about the interdependence of graduates' multiple student teaching experiences and the extent to which graduates engaged with a coherent set of experiences building toward early career survival and success in teaching. Others (Grossman, Hammerness, McDonald, & Ronfelt, 2008) have noted that the concept of coherence is under-explored in teacher education programs and based on the case study findings from the ETFQ, we agree that this topic could benefit from further examination in the service of program improvement. It also would have been desirable to gain still more in-depth information by interviewing the graduates, their student-teaching supervisors, and their current principals. However, practical constraints prevented this.

Many policy makers currently advocate the evaluation of teachers and schools based on VAM effects. While this position has the attractiveness of objectivity, it is not completely without a certain element of statistical, measurement, and model error. Unfortunately, our current state of knowledge concerning VAM is such that we are unsure of the nature and size of this error. In 2003 in an insightful analysis, Kupermintz raised questions about the validity of the Tennessee Value Added Assessment System (TVAAS) and stressed the need for further validity work on the system. Such analyses
need to be applied to all VAM systems to ensure that conclusions drawn from the systems have minimal error.

We are most encouraged by the potential of the DRSTOS-R for use in predicting graduates' performance as beginning teachers. Clearly, further validity work needs to be done in this area, but our findings suggest the possibility of using data from this instrument obtained on an individual during student teaching to tailor the nature of the mentoring provided to him/her as a new teacher. The findings also suggest the possibility of developing a form of the DRSTOS-R for use exclusively with new teachers to help pinpoint their strengths and weaknesses so as to be able to more efficiently provide targeted assistance.

Also, the finding concerning the importance of the match between pre-service and in-service school demographics for the early career success of graduates has important implications for teacher education programs. Program decision-makers and faculty need to analyze carefully the schools for which they are preparing their graduates and make sure students have experiences working in these types of schools in order to prepare them for successful induction.

Evaluation of teachers in training and of new teachers is a complex undertaking and results from this study serve to underscore the importance of treading carefully.
REFERENCES


Ecological Analyses of VAM Effects


APPENDIX: CASE STUDIES

Case 2

Program Experiences

Case 2 (C2) was a graduate-level student in English Education who entered the program with a moderate undergraduate GPA. Her academic record in the graduate program was accomplished, with grades primarily in the A-B range. Her lowest grade was a B, earned in her first student teaching placement.

For her initial student teaching experience, C2 was placed with a first-year 8th grade teacher in a K-12 school. This school had a privileged student body (12.6% Title I), which was approximately a third composed of minority students and which contained very few English Language Learners (1.3%) or students in special education (3.28%). C2’s ratings of this placement on the ETFQ indicated that this was a very poor experience. The cooperating teacher only moderately included C2 into the school environment and rarely offered her the opportunity to take control of the class. Furthermore, the cooperating teacher offered little by way of supporting C2’s developing organization, reflection and class management.

C2’s ratings of her second placement were considerably more positive. In this latter experience, she was placed in a high school with a cooperating teacher with more experience (6-10 years). The school had a comparably less privileged population (65.0% Title I) made up of predominantly minority students (95.1%) and notable special education and English Language Learner populations (19.15% and 7.30%, respectively). This placement’s cooperating teacher included C2 into the school community and provided pedagogical support and opportunities for C2 to take over instruction. Additionally, in contrast to the first placement’s cooperating teacher, the second placement mentor was reported as being more available and having more rapport with C2.

The student teaching supervisors affiliated with each of the student teaching placements were both given positive feedback, although the latter placement’s supervisor received higher ratings and met with C2 more frequently.

Consideration of C2’s experiences as a whole indicates that she experienced difficulties in both pedagogy and content knowledge. Although she received predominantly A-grades in her methods coursework, the B she received in her first placement of student teaching is unusually low for both C2’s academic record and typical grading for student teaching. Given that this grade was assigned by C2’s supervisor who was aware of the challenges posed by placement, this suggests a clear need for improvement. By her second (and last) placement, C2 continued to show weakness in pedagogical skill, earning an overall average DRSTOS-R score of 2.8, below the 3.0 benchmark for entry-level proficiency in pedagogical skill.

As C2’s graduate coursework was predominantly methods-based, there is little data available as to the kind of content knowledge she gained while in the program or with which she graduated. Her ETFQ responses indicate that the both her student teaching experiences only offered “average” support in this area. C2’s scores on the English Content Specialty exams required for licensure, however, suggest struggles; C2 required three attempts to pass.
In-Service Experience

C2 was hired within a year of graduation. The school in which she was hired had a large population of disadvantaged students (84.6% Title I) and was predominantly made of pupils from minority backgrounds (96.4%). The school also had sizeable special education and English Language Learner populations, making up 18.8% and 20.3%, respectively.

VAM data indicate that her students performed better than predicted on the ELA exams, both with respect to the city scores and scores of the school’s peer group. However, as of November 2008, C2’s status with Human Resources was listed as “Terminated”. The data available do not allow us to know the conditions surrounding C2’s termination aside from the fact that she is no longer teaching as part of the New York City Department of Education.

Discussion

C2’s pre-service profile indicates that she was a student teacher whose experience and performance improved between semesters but whose pedagogical and content knowledge showed limitations. Her DRSTOS-R score (2.8) as of her final placement particularly suggested that she was not yet performing at a level expected of a recent graduate.

This profile is consistent with the outcome of C2’s in-service experience insofar as her (non-)retention in the system. However, these data are at odds with the VAM data, which indicated success given that her students received higher-than-predicted scores on the ELA.

Under the ecological model, it is important to consider the context of C2’s experiences. Although the school in which she was hired had a considerably higher population of disadvantaged students than either of her two placements, the characteristics of the hiring school were more similar to those of her second placement than with her first. Both the hiring school and second-placement school were largely composed of minority students, had notable populations of ELL and special education students, and received Bs on performance and progress in their annual school progress reports.

Given C2’s positive experience in her second student teaching placement and its relative similarity with the hiring school, her termination cannot be attributed to a mismatch between her preparatory and in-service experience. If anything, this correspondence suggests that her performance in the second student teaching placement might be predictive of her performance in the field. The data at hand suggest that this may be the case the DRSTOS-R data indicating pre-Entry level proficiency better corresponds with C2’s termination outcome than her VAM data indicating instructional success.
Case 12

Program Experiences

Case 12 (C12) was a graduate-level student in the Childhood Education-Special Education dual-program of study who entered the program with relatively low undergraduate GPA. In contrast, her academic record in the graduate program was accomplished, with grades primarily in the A-B range.

In her four semesters of study in the program, C12 had three student teaching experiences, the first of which focused on common-branch childhood education and the latter two emphasizing special education at the childhood level.

School data were available only for the first two placements and ETFQ data were available only for the first. The first two student teaching placements were conducted in public elementary schools with privileged student populations; the first placement had 24% of students under Title 1, while the latter had only 8.6% Title 1 students. Both schools also had very low minority and English Language Learner (ELL) populations. Fewer than 50% of students were minorities in the first placement and 2.3% were ELLs; in the second placement 16.1% of students were from minority backgrounds and 3.7% were ELLs. Both schools also contained a notable special education population (14.8% in placement 1, 21.8% in placement 2).

For the first placement, for which ETFQ data were provided, C12 indicated that she had an overall positive experience both with regard to her cooperating teacher and university supervisor. Her cooperating teacher was experienced (2-5 years), was readily available, and was successful in incorporating C12 into the school environment and providing opportunities for controlling the class. Additionally, the cooperating teacher excelled in supporting C12’s developing knowledge of content, organization, pedagogy, class management and teaching practice.

In all three student teaching placements C12 received a grade of A. DRSTOS-R data were available only for the second placement and her overall average rating of 3.30 indicated that as of the beginning of her second year in the graduate program, she was performing above the 3.0 benchmark for entry-level proficiency expected at graduation.

Despite being in the dual-certification program, C12 only took the three state license exams required for childhood certification (LAST, ATS-W and the Childhood Content Specialty exam) and did not take the Students with Disabilities Content specialty exam required for certification in special education. She passed all three exams although her score of 226 on the Elementary Education Content Specialty Exam only just met the 220 cutoff for passing.

In-Service Experience

C12 was hired in the summer following graduation. The school in which she was hired had a student body whose socioeconomic status was typical of city schools (67.6% Title 1). The school had a relatively low population of minority students (46%) but sizeable special education and English Language Learner populations, both at approximately 22%. As of November 2008, C12’s Human Resources status was listed as “On Leave,” however, the nature of this leave is not indicated in the data available.

On the VAM data, C12’s students appear to be doing better than predicted on the Math exams, both in comparison to the city and peer schools. On the ELA, however,
C12’s students are performing lower than predicted with respect to both comparison groups.

**Discussion**

The point of interest in this case is the mixed VAM data. Given that the data available do not allow us to make inferences regarding the nature of C12’s leave, VAM data serve as the measure of C12’s in-service performance. Although the DRSTOS-R data indicated that even prior to her final placement C12 demonstrated strong pedagogic skill, it could not predict the mixed nature of her performance on the VAM. The DRSTOS-R could not predict this outcome because it is a general assessment of teaching skill; it does not make distinctions in instruction for different content areas.

C12’s strong overall showing across data sources in her preservice profile makes the negative VAM scores on the ELA exam surprising. However, closer examinations offer potential explanations. For instance, C12’s just-passing score on the Elementary Education Content Specialty exam may indicate a weakness in literacy content knowledge. Additionally, one of C12’s few lower course grades (being in the B-range) was in “Language and Literacy in the Early Years” hints at struggles with literacy instruction, although she did earn an A in the later-grades version of the course.

Perhaps most compelling is the option offered by a comparison of the instructional context of her hiring school with that in which she student taught. At least two of the schools in which C12 student taught had student populations with a very small proportion of English Language Learners (less than 5% each). In contrast, the school in which C12 was hired had nearly a quarter of its students designated as English Language Learners. This mismatch would mean that C12’s graduate experience offered few practical opportunities for her to learn how to instruct students in need of language support. Given that students’ language proficiency is more likely to impact performance on the ELA than on the math exam and C12’s limited preparation in working with an ELL population, the negative VAM ELA findings are

**Case 3**

Case 3 (C3) was a graduate-level student in the Childhood/Special Education dual-program of study. She entered the program with a strong undergraduate GPA and maintained a high record of academic achievement in the graduate program with grades of A and A-minus.

While in the program, C3 completed three student teaching placements, the first focusing on general childhood education and the latter two focusing on special education in the childhood setting. ETFQ data were available only for the first two placements as they were conducted during the academic year; the third placement was conducted during the summer. The first two placements were both conducted in public elementary schools with cooperating teachers with 2-5 years of experience.

C3’s first student teaching placement occurred in a privileged school (18.4% Title 1) with few minority students (34.0%) and English Language Learners (6.3%), and an
average-sized population of special education students (11.0%). Her feedback on this placement indicated a poor overall experience. C3 had a weak rapport with her cooperating teacher; although the cooperating teacher included C3 into the school community, no extra effort was taken to provide her with opportunities to take control of the class. The cooperating teacher was noted as providing only adequate support in developing C3’s reflection, organization, class management and content knowledge, although a slightly higher rating was given with regard to support in general teaching skill. C3’s ratings of her university supervisor for this placement were equally low-to-average with regard to support, with the exception of supporting classroom management, which was much weaker. However, she indicated a slightly higher level of rapport with the supervisor than the cooperating teacher.

C3 also expressed a complaint regarding the on-site student teaching seminar conducted by university supervisors. These on-site seminars are unique to the university’s childhood education program; while student teaching seminars are required of pre-service students in all teacher education programs, in most programs the course is typically conducted on-campus by an instructor who is not necessarily the student teachers’ supervisor. For C3, the seminar was perceived as a “waste of time” which did not adequately allow her to share and discuss her observations of the field.

Feedback on C3’s second placement was considerably higher, with positive ratings of rapport and overall experience provided for both the cooperating teacher and university supervisor. In relation to the first placement school, the second placement was comparably more diverse (64.7% minority), less privileged (38.2% Title 1), and had a greater special education population (21.1%). Like the first placement, the second placement school had few English Language Learners (7.9%).

The second placement cooperating teacher was reported as being highly available and making an effort in both incorporating C3 into the school community and deferring control of the classroom. Solid support was provided in all areas and the cooperating teacher was rated as excelling in creating a positive classroom environment. Likewise, the university supervisor was highly supportive of C3’s development; praise was given to the supervisor for her availability and involvement both in the placement and beyond.

C3 received a grade of A in each of her student teaching placements. DRSTOS-R data were available only for the second placement. On the DRSTOS-R, C3 received a rating of 2.75, which was below the 3.0 benchmark for entry-level proficiency.

During the semester of her second student teaching placement, C3 took and passed all four state exams required for her dual certification with scores well above the 220 cutoff.

**In-Service Experience**

C3 was hired in the interim between the summer session in which she completed her coursework and student teaching requirements and the fall semester of her official graduation date. She was hired by a public elementary school containing a privileged population of students (6.4% Title 1) with few English Language Learners (3.1%) and minority students (14.5%) and a moderately high special education population (17.3%). As of November 2008, C3 was listed by Human Resources as being “On Leave”. The data available do not indicate the nature of this leave.
C3 received average VAM scores for her students on the ELA in both the peer and citywide comparisons, indicating that her students were performing close to their predicted level. On the math test, C3 received positive VAM scores indicating that her students were performing higher than predicted.

**Discussion**

With a strong academic record and high certification exam scores, C3’s preservice profile showed signs that she would be a candidate for success in the field. The similarities in C3’s preservice placements with her hiring school also suggest potential for success; despite having had varying student teaching experiences due to her relationships with the cooperating teachers, both placements reflected the school in which she was hired: privileged, with moderate-to-average special education populations, and relatively few students who were English Language Learners or of minority backgrounds.

Given that the data available do not allow us to make inferences regarding the nature of C3’s leave, VAM data serve as the measure of her in-service performance. The VAM data indicate that C3 was successful in bringing her students to at least predicted levels of performance and that she was slightly more successful in her math instruction.

In this case, most predictors from the teacher’s preservice profile pointed toward success in the field and evidence of success was found by way of VAM data. While not solely attributable to C3’s success the similarity between her two preservice placements and her hiring school were likely to have supported her performance in the field, having provided preparation, or at least familiarity with the school context in which she would be teaching.

The DRSTOS-R data for C3 present an interesting inconsistency in this case. While other sources of data indicate C3 to be a solid teacher-candidate, her DRSTOS-R score indicated that she was not yet at entry-level proficiency, a potential sign of trouble for a graduating student teacher. It is important to recognize that C3’s DRSTOS-R was administered during the second of three placements and was the first of two student teaching placements in a special education setting. As a result, it may not fully have reflected her pedagogical proficiency as it appeared at the time of graduation. Without DRSTOS-R scores from the previous and subsequent placement, it is difficult to determine how this score reflects her overall development of pedagogical performance. (C3’s student teaching grades, all A’s, cannot be used for this purpose as the grades are assigned at the discretion of the individual supervisor without necessarily referencing a shared descriptive rubric as in the DRSTOS-R.) While useful for tracking student teachers developing pedagogical proficiency, mid-preparation DRSTOS-R scores may not be as useful for predicting in-service success as those administered in the final placement.

**Case 16**

**Program Experiences**

Case 16 (C16) entered the University as an undergraduate, having graduated from a performing arts high school and declaring a major in Instrumental Performance. Her first year of coursework focused on the requirements of her music major and the
University’s undergraduate liberal arts requirements. Her academic record shows average grades in music coursework and weak liberal arts grades.

In the fall semester of her second year, C16 changed her major to the Childhood Education-Special Education dual program of study. Following this shift, C16’s academic record improved, with grades in teaching methods and liberal arts in the A-B range. By the end of her fourth year, C16 added Italian as a third major.

As per the requirements of the undergraduate Childhood Education-Special Education program, C16 completed four semesters of student teaching. These placements occurred over the course of her fourth and fifth year in the University. Data on the school characteristics for all four placements were available; however, ETFQ data depicting the quality of the student teaching experience were available only for the first three.

All four student teaching placements were conducted in public elementary schools, with the first and fourth focusing on special education instruction and the second and third focusing on general childhood instruction. The first and third placements were conducted in relatively privileged schools with less than 12% Title 1 students. The second and fourth placements were in more disadvantaged schools with upward of 60% Title 1 students. C16’s second placement was by far the most diverse of her placements, with the highest percentages of minority students, English Language Learners and students in special education.

Given the ETFQ data available, for the first three placements C16 was placed with experienced classroom teachers and supervisors with whom she had strong relationships and rapport. Her first placement was rated as average overall, although she indicated that the cooperating teacher failed to include her into the school community. The second placement was the most highly rated, despite having a relatively less experienced mentor (2-5 years of experienced as opposed to 6-10 years). This cooperating teacher excelled and supporting C16’s development of class environment and organization, was very available, and provided adequate opportunities for reflection and taking control of the class. In contrast, the third placement was rated the most poorly, with the cooperating teacher offering little support in either teaching practice or content knowledge.

Despite variation in her experiences, C16’s student teaching grades reflect adequate performance in each placement. C16 earned a grade of A for each of her first placements, and an A-minus in her last. Although ETFQ data are not available to elaborate on the details of her final placement, this placement is the only one for which DRSTOS-R data are available. C16’s overall average DRSTOS-R rating corresponds with the relatively low student teaching grade; at 2.85 she was below the 3.0 benchmark for entry-level proficiency.

During the winter of her final year and the following summer, C16 took and passed all four state certification exams required of the dual-program major, including the Elementary Assessment of Teaching Skills - Written (ATS–W), the Elementary Education Content Specialty Exam, Students with Disabilities Content Specialty Exam, and the Liberal Arts and Sciences Test (LAST).

**In-Service Experience**

C16 was hired within a year of graduation under her common branch license. The school in which she was hired had a large population of disadvantaged students (76.0%
Title I) and was predominantly made of pupils from minority backgrounds (98.9%). The school also had sizeable special education and English Language Learner populations, making up 17.2% and 38.5%, respectively.

After half a year of working in this school, C16’s status with Human Resources was listed as “Terminated”. The data available do not allow us to know the conditions surrounding C2’s termination aside from the fact that she is no longer teaching as part of the New York City Department of Education. VAM data collected from C16’s students in this time indicate that her students were performing lower than predicted on the ELA exams, both with respect to City scores and scores in the school’s peer group.

**Discussion**

C16’s pre-service profile indicates an average teacher education student whose coursework grades and certification exam scores indicate conceptual understanding of what is needed in teaching, but whose practical performance upon the time of graduation was still in need of improvement.

Upon hire, C16 worked in a school whose population most closely resembled those she encountered in her second and fourth placements, the schools with greater populations of disadvantaged students and sizeable groups of English Language Learners and special education students. The similarity between placement 2 and the in-service experience carries particular weight as the two are the placements in which the focus corresponded with the license under which she was hired. The similarity of these placements to the hiring school indicates that poor performance cannot be attributed to unfamiliarity with the school environment.

Given C16’s performance in her second placement (earning a grade of A), her termination and low VAM scores come as a surprise. These outcomes are somewhat less surprising however when compared to the fourth placement, which as the final placement is an arguably better predictor of how she would perform following graduation. The fourth placement produced C16’s lowest student teaching score (A-) and a DRSTOS-R rating of below entry level proficiency (2.85).

In this case, VAM scores and retention are consistent outcomes of teacher success, and the DRSTOS-R is a better predictor of these outcomes than course grades or license exams. However, what remains unclear is whether her lack of success in the field is reflective of a lack of proficiency or limited time to develop and demonstrate ability. Although she received a lower student teaching grade and pre-entry level DRSTOS-R rating in her final placement, her higher student teaching grades in earlier placements in conjunction with her license exam scores and coursework suggest potential. Given that she worked in her hiring school for only half a year and the lack of data detailing the circumstances of her termination, it is uncertain what her performance outcomes (VAM and termination) really reflect.