A New Approach to Educator Preparation Evaluation: Evidence for Continuous Improvement?

Corinne Baron Donovan Jane E. Ashdown Anne M. Mungai Adelphi University

Abstract

The landscape for educator preparation has shifted to accountability models emphasizing performance assessment of teaching, employer feedback reports, newly approved accreditation standards showing impact on K-12 student learning, and expectations of public access to all of this information. This article provides a perspective on the extent to which this change offers promise for improving educator preparation programs and consequently excellence in teaching in K-12 schools. Two accountability reports are used as the empirical evidence for review; one is a pilot institutional feedback report from the Teacher Quality Research Center (Boyd, Lankford, & Wyckoff, 2009) and the second is a new Teacher Preparation Program report prepared by New York City's department of education (NYCDOE, 2013a). Ultimately, a systems perspective is recommended, in which candidates, IHEs, and K-12 schools are involved in the process of how educator preparation is evaluated and how that connects to other aspects of the education profession.

Historically, educator preparation evaluation models have relied on state approval of programs, pass rates on licensure exams, and meeting accreditation standards that privileged operational and descriptive data as a basis for evaluating program quality. That landscape has shifted in educator preparation to accountability models emphasizing performance assessment of teaching practice, employer feedback reports that include growth scores for program graduates based on their students' standardized test scores, newly approved accreditation standards that require evidence of positive impact on K-12 student learning, and expectations of public access to all of this information. This article provides a critical perspective on the extent to which this changing accountability landscape offers promise for improving educator preparation programs and consequently for driving excellence in teaching and learning in K-12 schools as evidence of continuous improvement.

The empirical basis for this article is two reports that establish data linkages between the graduates of one Institution of Higher Education's (IHE) educator preparation program and the school systems where those graduates have been

teaching. The first report is a pilot institutional feedback report from the Teacher Quality Research Center (TQRC; Boyd, Lankford, & Wyckoff, 2009) and the second report is a newly released Teacher Preparation Program report (2013) from the New York City Department of Education (NYCDOE). The overall purpose of the TQRC report is to provide schools and colleges of education in New York State (NYS) with information about where graduates from their teacher education programs are in the teaching profession NYS schools; the purpose of the NYCDOE report is to provide education programs at local colleges and universities (N=12) with a snapshot of their graduates' contributions to the NYCDOE schools after leaving their teacher preparation programs.

Purpose

The assumption behind the reports (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a) and findings presented here is that for schools of education to improve and produce more effective educators, they need to know what happens when graduates finish their programs and become teachers in the classroom. To examine this assumption in the context of the changing nature of teacher education program accountability, we review the following questions. First, we consider the evidence from a program improvement perspective and try to answer the questions: Which features of teacher education programs do the findings from these reports help inform (e.g., sequence and content of academic course work, full or part-time program design, area of preparation)? Will the findings from accountability reports lead to change and enhance the effectiveness of teacher education programs?

The next question to consider is how the findings from these reports impact clinical practice and the school partnerships essential to educator preparation. Educator preparation is not a stand-alone endeavor, but rather requires cooperation from skilled teachers and administrators in the current K-12 school system. We want to know how information about our graduates arising from these reports impacts decisions about school partnerships and clinical experiences which includes the placement of teacher candidates in classrooms, the selection of teacher mentors, the selection of supervisors, the nature of the supervision, and the impact of candidates' teaching on student learning.

Finally, we consider to what extent the data included in these reports (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a) are relevant and actionable from an educator preparation policy perspective. The policy intent of the shift toward accountability models is to drive reform in teacher education by making clear distinctions between teacher preparation programs (U.S. Department of Education, 2011). Therefore, do the findings from these reports provide relevant and actionable information with regard to policy decisions, for example about admissions' standards and selection policies for entry into an educator preparation program? Only when information is relevant to the public and the parties involved (e.g., educator preparation programs, participants in these programs, K-12 schools who hire graduates from educator preparation programs, regulatory bodies such as state education departments,

policy makers, taxpayers) does it provide clarity for policy changes and actions needed for teacher education program improvements.

After analyzing data in these reports and addressing this set of questions, we conclude by recommending a systems perspective on accountability in teacher education (Boulding, 1956; Katz & Kahn, 1966; Parsons, 1951). Systems theory, in this respect, considers individuals and organizations as part of a larger open system, where the environment and all parts of the system have an impact on the survival and success of the system. Early theorists in organizational theory and social sciences sought a new and common paradigm that would allow researchers across multiple disciplines to access common terminology (Hillon, 2005). Much of this work is grounded in biological sciences, considering such concepts as energy to sustain a system, homeostasis, entropy, and system cycles (Katz & Kahn, 1966). Parsons (1951) pointed out the defining feature that holds a system together is the integrated values or norms which drive the system. In the current study, we consider educator preparation the system under review and analysis.

The Research Context

Teacher education is under immense pressure to change and improve, pressure driven in large part by the poor performance of P-12 students in the nation's public schools especially when compared to international student achievement outcomes based on recent data from the Organisation for Economic Cooperation and Development Progress on International Student Achievement (OECD, 2010). This concern is accompanied by a persistent failure to adequately address the widening achievement gap among diverse student groups (Wiseman, 2012).

Historically, research studies show that teacher quality varies and that variation in quality is associated with both student success and with compounding disadvantages for low achieving students (Goldhaber & Anthony, 2004; Lankford, Loeb, & Wyckoff, 2002). This variability has put a spotlight on impact and outcomes in teacher education and has resulted in a paradigm shift away from measuring teacher quality and toward measuring teacher effectiveness (Darling-Hammond, 2000) with the achievement test score gains of students a key component. In turn, teacher education programs are increasingly being held accountable for their graduates' impact on student learning as measured in some instances by the use of value-added measures.

Longitudinal databases and the accompanying capacity to track the impact of education program graduates on student learning have been lacking in many states; however, increasing numbers of states are now able to or are committed to doing this (Henry, Kershaw, Zulli, & Smith, 2012). This development is in part leveraged by federal and state stimulus funding and through accountability requirements associated with Race to the Top (RttT) grant awards that require grant recipients to build comprehensive tracking databases and to recruit and retain effective teachers especially in high needs schools and fields (U.S. Department of Education, 2014).

Henry et al. (2012) provided a snapshot of 12 state's RttT proposals with regard to the assessment of teacher preparation programs. Their analysis draws attention to the challenges states face in establishing a "true effect" of a preparation program on student test scores (p. 350). The National Center for Education Statistics (NCES) reports that 41 states along with Washington, DC have each received at least one grant for the development of statewide longitudinal data systems providing evidence of this increasing trend toward tracking and linking data (NCES, 2012).

Gansle, Noell, and Burns (2012) provided findings based on one year's analysis from Louisiana's implementation of a Teacher Preparation Accountability System for evaluating the effectiveness of teacher preparation programs. This was one of the earliest accountability systems to incorporate multiple data points, involve the redesign of university-based teacher education programs, and incorporate an evaluation of new teachers' performance based on their students' achievement on standardized tests. Using hierarchical linear modeling with data from this comprehensive tracking database, the authors estimate the degree to which the students of new teachers from different types of teacher preparation programs achieved more or less than predicted outcomes in key content areas on state achievement tests. Results showed variation in achievement gains between students taught by teachers from different educator preparation programs; however, the authors caution that particular results for an institution do not explain why those results occurred. The teacher preparation program is then left with the challenge of unpacking the data and developing hypotheses about which variables are driving particular outcomes.

Plecki, Elfers, and Nakamura (2012) also examined the extent to which value added measures are a useable source of evidence for improving teacher education programs. The authors used fifth grade teachers' value added scores to investigate whether student achievement varies by teachers' preparation program (in-state versus out-of-state programs) and by years of teaching experience. Although the relation between years of experience and teacher value added scores was significant, outcomes in terms of the relation of value added scores to teacher preparation programs were mixed. The authors concluded with important recommendations about the need for cooperation among multiple stakeholders with regard to accountability, a recommendation that is also considered in relation to this investigation.

The Data Bases

The reports (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a) referenced as an empirical basis for this article, reflected a trend in longitudinal database development and represent distinct moments in the history of educator preparation program accountability. The TQRC report is the outcome of a pilot study funded through a partnership between an institution of higher education, a state education department, and a philanthropic foundation. The context for this initiative was twofold. Firstly, in 2001, the Carnegie Corporation launched a major reform initiative in teacher preparation called Teachers for a New Era (TNE; Carnegie Corporation of New York,

Donovan, Ashdown, and Mungai

2001) and offered grant awards to institutions of higher education to implement reforms. These awards were given to selected institutions of higher education (N=11) who were committed to partnerships between their education and arts and sciences programs in the preparation of teachers and also committed to measuring the impact of their education programs in terms of evidence of student learning. This required education programs to rethink their assessment systems with a focus on collecting persuasive evidence of impact on student learning (Fallon, 2006; Kirby, McCombs, Naftel, & Barney, 2005). Secondly, a study of urban public school teachers was undertaken a few years later and examined the effects on student learning of different features of teacher preparation programs (Boyd, Grossman, Lankford, Loeb, & Wykoff, 2009). This study was funded through a partnership between an institution of higher education and several philanthropic organizations and was one of the first studies to use value added modeling to estimate the effects of different teacher education program features in relation to beginning teacher effectiveness. One finding of interest was that preparation programs providing more oversight of student teaching supplied more effective first year teachers to schools.

Both the TNE (Carnegie Corporation of New York, 2001) initiative and the study of urban teachers (Boyd, Lankford, & Wyckoff, 2009) described above placed substantial emphasis on the collection and analysis of evidence about the impact that teacher education candidates and graduates have on student learning as a critical indicator of program effectiveness. The TQRC report under discussion here represents a continuation of that effort to engage more systematically in linking information about K-12 schools and students with graduates of teacher education programs. The Teacher Quality Research Center was housed on the campus of the New York state university system and the TQRC reports were developed through a consultative process with the teacher education community reflected in membership of an advisory group established to provide feedback on report design; one author was a member of that advisory group. The reports were prepared for each approved teacher education program provider in New York State (NYS; N=100), and for the first time provided institutions of higher education with comparative information as well as aggregate measures of student learning. The reports were not publicly available and this was a one-time endeavor as funding was not forthcoming to support multi-year reporting.

The second report discussed here was prepared by a local education authority (LEA), the NYCDOE (2013a), for the IHE of interest. Similar individual reports have also been developed for 11 other IHEs supplying teachers to the NYCDOE public schools as well as a report comparing all 12 IHEs on the selected metrics. All reports are publicly available at http://schools.nyc.gov/Offices/DHR/HumanCapitalData/TPPR. This reporting strategy reflects the increasing capacities of LEAs and state education authorities to collect data and use it to report on variables of interest (Henry et al., 2012). This strategy aligns with current federal education policy exemplified in RttT competitive funding awards to state education departments requiring that teachers be evaluated based on the achievement gains of K-12 students. Similarly, the public

availability of this set of reports aligns with current expectations for transparency in teacher, principal, school, and teacher education accountability.

Table 1 provides a summary comparison of the distinctive design features of each of the reports (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a) used as the data base to address the research questions. The two reports were produced within a relatively short timeframe, yet it is worth noting the differences in features of access and transparency between them. While the reports share a common purpose, differences in their design reflect the rapidly changing landscape of educator preparation program accountability.

Table 1
Summary Comparison of Educator Preparation Program Accountability Reports

	Boyd, Lankford, & Wyckoff, 2009 (University based initiative)	NYCDOE 2013a (Local Education Authority initiative)
Consultation	Advisory group representing IHEs reviewed drafts with study authors	Pre-publication draft shared by LEA with each IHE for feedback
Transparency	All IHEs remained anonymous	All IHEs identified by name
Purpose	To help teacher preparation institutions in program planning, assessment, and program improvement	To help colleges and universities assess and refine their teacher preparation programs
Access	Password protected, zip file in Excel	Available on LEA website in PDF format
Impact on student learning	Uses K-12 school level 4 & 8 grade test scores as a measure of graduates' impact on student learning	Uses individual teacher growth scores based on 4-8 grade student test scores aggregated for each IHE's graduates
Publicity	None	LEA press release and press coverage
Length of report	19 tables	6 tables

In question is whether the right variables are being examined and whether the resulting findings are causally robust and can be used to make significant improvements in teacher education programs (Clotfelter, Ladd, & Vigdor, 2010; Donovan, Ashdown, &

Mungai, 2013). We now consider some of the challenges in using the findings from the TQRC pilot study (Boyd, Lankford, & Wyckoff, 2009) and the NYCDOE Teacher Preparation Program report (NYCDOE, 2013a) for teacher education program improvement purposes in one IHE of interest located in NYS.

Data Analyses by Report

The TQRC report (Boyd, Lankford, & Wyckoff, 2009) integrated data across six sources of information about cohorts of teachers (referred to as program completers) in NYS who completed their educator preparation between 2000 and 2005 and allows tracking of early teaching careers through the 2006-2007 school year. The Teacher Preparation Program Report created by the NYCDOE (2013a) analyzes the performance of 12 teacher education programs that supplied the most teachers to New York City's public school system from 2008 to 2012. The report is the nation's first district-level teacher preparation report to analyze the quality, distribution, and retention of new teachers hired from traditional college and university teacher education programs.

TQRC report. Data from six different information sources were integrated to complete a TQRC report for each IHE as follows:

- 1. A Personnel Master File includes information on the schools where the program completers teach in the NYS public system.
- 2. The Exam History File includes NYS certification exam scores.
- 3. The TEACH file lists teaching certificates awarded by NYS.
- The Program Completers File includes information about each individual recommended for teacher certification by a NYS approved preparation program.
- 5. Elementary and Secondary School data file includes demographic information about school populations and accountability status (in terms of need for improvement).
- The College Board File, which includes SAT math and verbal scores for all program completers who took the SAT in NYS between 1980 and 2000, as well as high schools attended.

The report includes summary data for three comparison groups: a sector grouping (Public or Independent IHEs), a regional grouping based on an IHE's geographic proximity, and a statewide comparison. For purposes of this article, we limit the data presented to the IHE of interest and the statewide comparison.

Ultimately, the report (Boyd, Lankford, & Wyckoff, 2009) included 19 tables of information such as demographics (age, gender, ethnicity), results for program completers on the NYS Teacher Certification Exams and NYS Awarded Certifications, initial employment of program completers (general information; percentages by subject,

grade level, and type of school; and demographic attributes of the K-12 schools where program completers taught), K-12 educational outcomes in NYS public schools where program completers initially taught, location of K-12 schools, and retention data. Results are shared here on variables of interest to teacher education leaders at the IHE of interest as a basis for meeting accreditation standards in the areas of assessment (including employer feedback and evidence of student learning), for implementation of the IHE's conceptual framework, and for overall continuous improvement purposes.

The total number of program completers with enough data available for analysis across the six data sources in the report for NYS was 56,000, and the total for the IHE of interest was 2,559. This is about half of the total number of program completers who graduated from the IHE in that timeframe. The report authors explain that data were missing due to incomplete information or errors in the files (e.g., names and identification codes could not be matched across files).

TQRC results. Table 2 provides demographic information for the IHE of interest compared to the NYS total for the 2000-2005 cohort used for analysis. Demographically, the IHE of interest has program completers who are slightly younger than program completers across NYS; about 24% of IHE of interest program completers are younger than 25, compared to the state average of 13%. The IHE of interest has slightly more female program completers (85% female) compared to the state rate of 75%. The IHE of interest has slightly higher rates of program completers who are Hispanic (6.5%) compared to the state average of 5.8% and White (70.5%) compared to 66.5%. The IHE of interest also has a higher rate of program completers who are recommended for initial or provisional certification (78%) compared to the statewide average of 63%.

Table 2

Demographics of Program Completers, 2000-2005 Cohorts, IHE vs. NYS

-	N	IHE	Statewide-NYS
<u>Age</u>			
< 25	612	23.9%	13.2%
25-30	1,012	39.6%	46.8%
31+	934	36.5%	39.9%
Total	2,558	100.0%	100.0%
Gender			
Female	2,165	84.6%	75.4%
Male	352	13.8%	19.0%
Unreported	42	1.6%	5.6%
Total	2,559	100.0%	100.0%
Ethnicity			
White	1,803	70.5%	66.5%
Black	155	6.1%	6.6%
Hispanic	166	6.5%	5.8%
Other	57	2.2%	2.9%
Unreported	378	14.8%	18.1%
Total	2,559	100.0%	100.0%

The report includes general counts and percentages by subject taught for the first year of teaching in NYS public schools (Table 3). The IHE of interest, for example, has a slightly higher percentage of program completers who teach at the elementary level (35%) compared to the percentage of elementary program completers across NYS (33%). The IHE of interest has the same percentage of program completers who teach math at the state level (5.3%) but slightly lower percentages of program completers in English and social studies. The IHE of interest has a slightly higher percentage of program completers in special education (18.7%) compared to the NYS percentage (16.7%) and a higher percentage of other teaching areas (25.4%) compared to the NYS percentage (18.3%), which includes subjects such as physical education and health.

Table 3
Subjects Taught in First Year of Teaching in NYS Public Schools 2000-2005 Cohorts, IHE vs. NYS

Subject Area	IHE FTE	IHE %	NYS %
Elementary	233.1	35.0%	32.7%
Literacy	17.9	2.7%	3.5%
Special Education	124.6	18.7%	16.7%
English	42.8	6.4%	8.0%
Mathematics	35.5	5.3%	5.3%
Science	14.9	2.2%	8.0%
Social Studies	27.4	4.1%	7.4%
Other Teaching	169.1	25.4%	18.3%
Total	665.2	100.0%	100.0%

When examining the data for program completers teaching in high needs schools, the IHE of interest has a higher percentage of program completers working in schools with the poorest students (see Figure 1). For example, when program completers across the state are rank ordered by the percentage of K-12 students eligible for free or reduced price lunch (FRL), the top 25% of IHE program completers are in schools with 71% or more K-12 students receiving free or reduced price lunch compared to the top 25% of NYS program completers who are in schools with 63% of K-12 students receiving free or reduced price lunch. Similarly, the IHE of interest has a higher percentage of program completers who teach in schools with Black and Hispanic students (see Figure 2). For example, when program completers across the state are ranked by number of K-12 students who are Black or Hispanic, the top quarter of program completers from the IHE of interest and across the state are in K-12 schools with 79% or more Black or Hispanic students (IHE) compared to 63% or more Black or Hispanic students (NYS).

96

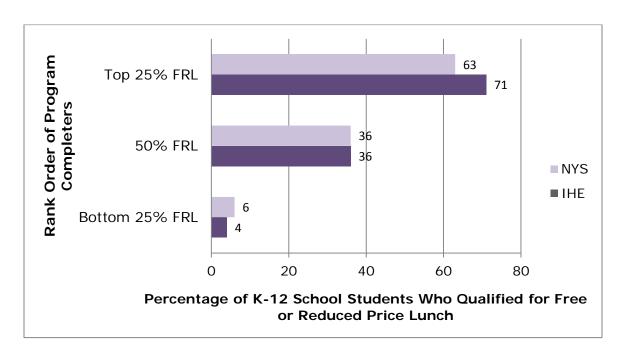


Figure 1. Rank ordering of program completers by the percentage of their K-12 school's students on free and reduced price lunch status.

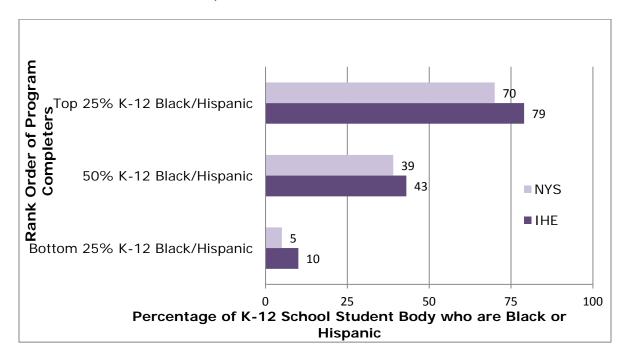


Figure 2. Rank ordering of program completers by their K-12 schools' percentage of Black or Hispanic students.

Retention is presented in the report relative to initial employment in NYS public schools (Table 4). The data reveal that roughly 81% of teachers statewide and at the IHE of interest are still teaching in the NYS public school system three years after initial employment. However, these data must be interpreted with caution, since the number of program completers that are tracked after three years declines by roughly 50% (407 compared to 824) from the number tracked at initial employment. This is the same attrition rate for both the IHE of interest and IHEs statewide. It is difficult to interpret the 81% rate without further context and details about the dataset.

Table 4
Teacher Transfers and Attrition Relative to Initial Employment in NYS Public Schools, IHE vs. NYS

	Initial Employment	1 Year Later	3 Years Later
<u>IHE</u>			
Number of Program Completers	824	769	407
Still Teaching in NYS Public School System		90.2%	81.3%
Teaching in Same District		80.9%	73.2%
Teaching in Same School and District		73.0%	71.3%
NYS			
Number of Program Completers	16,740	15,533	8,565
Still Teaching in NYS Public School System		88.8%	81.2%
Teaching in Same District		79.0%	75.3%
Teaching in Same School and District		72.6%	71.4%

One of the more interesting tables in the report includes Educational Outcomes data for students in NYS public schools where program completers initially teach (Table 5). Outcomes data are reported for students in grades 4 and 8 on math and English Language Arts (ELA) tests, as well as English and math Regents exams (NYS high school proficiency exams by subject). Outcomes on the grade 4 tests indicate that program completers of the IHE of interest teach in schools where the students have slightly higher rates of proficiency compared overall to IHEs across the state. In 4th grade math, the IHE of interest has 80% of students proficient compared to 78% for NYS. In ELA, the IHE of interest has 70% of students proficient compared to 66% for NYS. This trend continues on the grade 8 exams with approximately a 6-point difference. Program completers from the IHE of interest have students with higher percentages performing at the proficient level compared to the state overall. In 8th grade math, the IHE of interest has 61% of students proficient compared to 55% for NYS. In ELA, the IHE of interest has 56% of students proficient compared to 50% for NYS.

The data shown in this report are difficult to interpret with regard to K-12 student achievement since outcomes are presented at a group level. In other words, K-12 student achievement is calculated and reported across an entire grade level for a school where program completers from the IHE of interest teach, and each program completer may be one of several teachers providing instruction at the 4th or 8th grade level.

Table 5
Educational Outcomes of Students in NYS Public Schools Where Program Completers Initially
Taught, IHE vs. NYS

Percentage of Student Proficiency Outcomes by Grade	IHE	NYS	Percentage Points Difference
Grade 4 Math Exam	80.1%	77.8%	+2
Grade 4 ELA Exam	70.2%	66.0%	+2
Grade 8 Math Exam	61.3%	55.2%	+4
Grade 8 ELA Exam	55.8%	49.5%	+6
Students Scoring ≥ 65 on English Regents	90.2%	90.0%	
Students Scoring ≥ 65 on Math Regents	85.8%	85.2%	

As a summary of the TQRC pilot study (2009) results, the IHE of interest has program completers who teach at the elementary level at roughly the same rate as program completers across the state. Program completers from the IHE teach at K-12 public schools that have a greater percentage of students receiving free or reduced price lunch and students who are Black or Hispanic. The K-12 schools with teachers from the IHE show a greater number of students scoring proficiently on the 4th and 8th grade math and ELA exams. Finally, program completers from the IHE persist in the public school system at the same rate as program completers across the state.

NYCDOE teacher preparation program report. The NYCDOE Teacher Preparation Program Report (2013a) not only provides an analysis of graduates from the teacher education programs at the IHE of interest, but also analyzes in separate reports the performance of 11 other IHE teacher education programs that supplied the most educators to the NYCDOE city public school system from 2008-2012 (NYCDOE, 2013b).

All reports are publicly available, and thus, comparative data across teacher education programs can be accessed. In the time frame covered by the report, over 10,000 new teachers were hired by the NYCDOE (N = 10,135) from traditional pathways (e.g., graduated from college and university education programs) with 51.6% (N = 5,229) of new hires graduating from the 12 schools included in the reports. The reports provide analysis of the quality, distribution, and retention of new teachers and focuses on promoting awareness and cultivating productive partnerships between local schools of education and the NYCDOE. Specifically, the reports aim to evaluate the education programs' contributions toward preparing teachers to meet the diverse

recruiting needs of the NYCDOE. The reports describe the hiring and retention of effective teachers as being inextricably tied to partnerships between the NYCDOE and local schools of education as the certifying institutions.

The reports provide metrics in three areas: (1) Meeting the Needs of the NYCDOE, (2) Performance, and (3) Retention. There are a total of 6 measures across these three areas. Representatives from the NYCDOE met with each educator preparation program to review a draft of the report, and there are plans for continued collaboration around the reports and their uses.

NYCDOE teacher preparation program results. In evaluating contributions to meeting the staffing needs of the NYCDOE, two personnel metrics are presented: Highest Needs Schools and Subject Shortage Licenses. The first metric shows the percentage of new teachers hired into the Highest Needs Schools across the city. This includes all schools designated for special education, as well as schools from the top 25% of need as measured by a prior year progress report using a peer index developed by the NYCDOE. Overall, across the NYCDOE, 30% of the 10,135 teachers were hired into highest needs schools, with a range from 16% to 48% hired into highest need schools across the 12 IHEs; the IHE of interest has 24% of program completers hired into highest needs schools out of a total of 264 hired from the IHE across the four years. The second metric presents the number of teachers hired by highest need license area which includes teachers of English as a Second Language (ESL), Math, Science, Special Education, and Other (bilingual and other foreign language certification areas). The NYCDOE hired 69% of new teachers over the past four years into one of the highest need license areas. Across all 12 IHEs from 55% to 92% of graduates' hired held licenses in these areas; the IHE of interest has the third highest rate with 75% hired in highest need areas out of 175 completers hired from 2009 to 2011.

Three metrics are used to assess performance of the recently hired NYCDOE teachers: Tenure Decision, Unsatisfactory Ratings at year one performance reviews, and Growth Scores. The tenure findings identify the percentage of teachers who achieve three different ratings: *approved, extended*, or *denied*. Results in the report only include the first tenure decision point for teachers hired by October 31 in each of three academic years 2008-2009, 2009-2010, and 2010-2011. The mean percent of NYCDOE teachers approved at the first tenure decision point is 60% and ranges from 51% to 67% across the 12 IHEs. The IHE of interest has the highest percentage of approval at the first decision point with 67% approved out of 150 hired from 2008 through 2010. The second metric reports the percent of teachers rated Unsatisfactory in their first year of teaching for those hired in each of four academic years from 2008-2009 through 2011-2012. The NYCDOE average rating of first year teachers receiving an Unsatisfactory is 3.1% across the four years and ranges from .7% to 4.8% across the 12 IHEs. The IHE of interest has 2.3% of its 264 alumni rated Unsatisfactory in their first year teaching.

The third metric used to assess performance is the 2011-2012 Growth Score. For the NYCDOE report (2013a), this score is only calculated for those teaching 4th and 8th grade math and English Language Arts (ELA) during the academic year 2011-2012. The Education Department growth scores are calculated using 4th and 8th grade student scores on state exams in math and ELA. Scores are only calculated for students who have two consecutive years of test data on record and are determined by a statistical model that rates students' growth in each content area relative to other similar students. This growth calculation, therefore, is designed to assess *relative* student growth and not achievement, which prevents teachers from being penalized for students who are not performing at grade level. Growth scores for individual teachers are then calculated by taking the average of student growth percentiles taught by each teacher.

There are four rating levels on the growth score model: Ineffective, Developing, Effective, and Highly Effective. Using only growth scores for 4th and 8th grade math and ELA teachers leaves a limited group of students and teachers being assessed (N = 1,466) relative to the total number of teachers hired (N = 10,135). The report notes that results should be interpreted with caution. The NYCDOE overall reports 82% of teachers are rated in the top two categories: Effective (75%) or Highly Effective (7%). This same rating ranges from 61% to 91% across the 12 IHEs. The IHE of interest has 80% of its 41 math and ELA teachers rated in the top two categories (68% Effective, 12% Highly Effective) based on the growth scores, an overall percentage slightly below the average for the NYCDOE, although it has a higher percentage of graduates rated highly effective compared to the NYCDOE. At the same time, 12% of the teachers from the IHE of interest were rated as Ineffective compared to 6% of the overall NYCDOE teachers; the percentage with ineffective ratings vary between 2% and 14% across the 12 IHEs.

The final metric presented in the NYCDOE report is the retention of teachers after three years of teaching, and therefore includes only those hired by October 31 for 2008-2009 and 2009-2010. The NYCDOE three year retention rate is 80% and across the 12 IHEs ranges from 72% to 94%. The IHE of interest has a retention rate of 83% for its 131 program graduates.

As a summary of the NYCDOE report, graduates from the IHE of interest are teaching in highest need schools at slightly lower rates than for newly hired teachers in the NYCDOE, but a greater percentage of hires are in the highest need subject fields compared to the NYCDOE teachers overall. Graduates from the IHE of interest are retained at about the same rate as the NYCDOE newly hired teachers, are pacing slightly ahead of newly hired NYCDOE teachers overall in first time approved tenure decisions, and are less likely to receive an unsatisfactory rating than NYCDOE teachers. Their ratings in terms of student growth scores pace ahead on highly effective ratings compared to the NYCDOE teachers, but there is a higher percentage of ineffective teachers from the IHE of interest compared to the NYCDOE.

Discussion and Implications

Implications and limitations of the reports are considered for both the IHE of interest and for other IHEs facing the implementation of new evaluation systems for their educator preparation programs.

Implications of the Reports Beyond the Local Context

The results of the reports presented here reference IHEs in one geographic region of the country. However, for IHEs in states that have not yet implemented an educator preparation evaluation system or are facing newly implemented evaluation systems, this particular set of reports is illustrative of both what to expect in terms of increasingly sophisticated measures available for accountability purposes and in terms of the limitations of these newly developing systems.

Of most significance in comparing these reports in terms of measures of impact on student learning is that the earlier TQRC report (2009) uses group level K-12 test scores as one measure of program completers' impact on students' learning in 4th and 8th grade. The NYCDOE report (2013a), on the other hand, uses more advanced teacher impact measures now available, such as teacher growth scores. The shift to this advanced analytical capability allows analysis at the individual teacher level rather than at the group level as the latter approach confounds inferences of teacher effectiveness. While the more sophisticated growth score results are only available for math and ELA teachers at this point, the advanced capability to look at individual teachers facilitates more accurate ratings of individual teachers on their own merit, although these methods remain controversial among researchers and educators alike. For example, Henry et al. (2012) caution about generalizing from state teacher preparation program reports given the challenges of obtaining unbiased estimates of programs' impact on student learning.

More broadly for those IHEs in other states facing newly implemented program evaluation, the TQRC report (2009) reflects a challenge facing many longitudinal data systems, that of following consistent cohorts of teachers across time and geography. As noted in the TQRC report, 50% of the teachers initially tracked into their first teaching position were not in the database three years later. Without inter-state education agency cooperation, it is not clear whether these individuals left the profession, relocated to another school system within the state (e.g., parochial, private, or charter), or simply relocated to another part of the country. An additional complexity is pointed out by Plecki et al. (2012) whose analysis of teacher preparation programs in Washington state related to value added measures revealed a significant, positive relationship between teacher experience and value added scores. The positive relationship did not change in magnitude or significance even after accounting for differences in gender, education level, or race/ethnicity. Novice teachers, with less than three years of experiences, had a lower value-added measure by more than one point each for reading and math. Therefore, if a 50% teacher attrition rate is a reasonable

estimate three years after initial employment *and* if, as Plecki et al. suggest, less than three years into a job teachers have lower value-added measures, it is critical to ask if these new evaluations of preparation programs are looking at the most accurate evidence to measure program effectiveness.

Finally, the ability to link data across multiple systems and take into account various individual factors in order to calculate student and teacher growth scores requires careful and time-consuming work which must be supported by adequate resources. The TQRC (2009) reports were only available for one year due to a lack of further funding. The report results served as a baseline of educator preparation program performance, but without further reports the findings are of limited value from a program improvement perspective. The NYCDOE report (2013a) was produced under LEA leadership that is no longer in place. There is some uncertainty about how the new LEA leadership will act with regard to the development of future reports. It is encouraging from one perspective that this level of evaluation is being promoted; however, the return on investment is not obvious as the extent to which the findings from either of these reports can lead to change and improvement in the effectiveness of teacher education programs is not well established.

Implications for Program Improvement

As we consider the first set of questions posed for these report findings (TQRC, 2009; NYCDOE, 2013a) and examine the evidence from a program improvement perspective, it is difficult to determine which educator preparation program features are implicated by the findings from these reports. For example, the IHE of interest has a higher percent of graduates rated as ineffective compared to NYCDOE teachers overall based on growth scores and a slightly lower percent of graduates rated as unsatisfactory in their first year compared to the DOE teachers overall (NYCDOE, 2013a). A logical program improvement goal for the IHE of interest in the service of producing more effective teachers is to identify program changes that would eliminate the likelihood of any graduates being rated unsatisfactory in their first year teaching and to eliminate the number of teachers rated as ineffective based on student achievement gains. The logical question then is which features of teacher education programs do these findings point to in terms of change? The unsatisfactory rating is based on classroom teaching observations conducted by a building leader, but without more proximal information about the evaluation rubrics used, identifying program changes would only be speculative. The ineffectiveness ratings based on growth scores are for math and ELA teachers only. One possible area for consideration for program improvement is the teaching methods courses (math and language arts) that graduates complete during their preparation program. Another possible action step is to compare educator preparation program features at the IHE of interest with program features at IHEs in the NYCDOE sample with the fewest teachers rated as unsatisfactory and ineffective. It is not clear, however, for either of these possible actions, which features of course work or program designs should be studied and changed in order to achieve the goals identified above.

Results from both reports provide some findings for the IHE of interest that reflect positively on the institutions' mission and core values of social justice and inclusive community. Graduates from the IHE of interest were reported as teaching in highest need schools, in highest need subject fields (e.g., special education), and in schools with greater percentages of children receiving free or reduced price lunch at similar or greater rates than comparative graduates. This appears to provide some confirmation that graduates are reflecting the institution's core values in their career trajectories (e.g., Inclusive Community and Social Justice are two of the core values).

The more recent NYCDOE report (2013a), however, shows that over the past four years, when compared to the DOE overall, a smaller percentage of graduates from the IHE of interest have been hired into highest need schools in the city. Across all 12 IHEs, the range is from 16% to 48% of graduates hired into highest need schools, suggesting that for the IHE of interest there is still room for improvement. An appropriate goal, therefore, might be to at least match the NYCDOE percent hired into highest need schools overall. Of note, the IHE with the biggest percentage of graduates teaching in highest needs schools is itself geographically located in the vicinity of high needs schools. It is well established that teachers typically work close to home and to where they attend school. For the IHE of interest, which has its main campus in a suburban neighborhood, to improve on the percent of graduates teaching in highest need schools would require strategizing about ways to counter this employment trend among teachers.

Implications for Clinical Practice and School Partnerships

It is difficult to determine what the implications are for the design and implementation of clinical experiences and for school partnerships from the reports (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a). The IHE of interest has two distinct tracks for clinical experiences both of which comply with state regulations, but one of which goes far beyond state requirements and involves a yearlong immersion in the K-12 school environment. The question of whether graduates who participate in this immersion track are more likely to be retained in teaching arises. Similar questions surround how the quality of supervision and mentoring and the characteristics of the student teaching placements associated with the IHE of interest might shed light on the outcomes for graduates who receive unsatisfactory ratings or who are ineffective or conversely highly effective in terms of student growth scores.

In the absence of a roster tracking individual graduates from either report (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a), it is difficult to know what aspects of clinical practice would benefit from change and improvement. As noted earlier, teacher preparation is not a standalone endeavor, and it would be reasonable to assume, for example, that graduates from the IHE of interest teaching in the NYCDOE might well have undertaken clinical experiences in the NYCDOE public school system. Further then, might it be reasonable to assume that the effectiveness of graduates from the IHE of interest in part reflects the quality of those K-12 clinical placements? If so, actions to

improve educator preparation programs and the effectiveness of their graduates do not rest with IHEs alone. A systems analysis is required to provide greater clarity in understanding which type or types of field experiences help prepare teacher candidates most effectively as they begin their teaching careers (Katz & Kahn, 1966; Parsons, 1951).

Whether considering changes to academic programs for educator preparation or changes to the design of clinical experiences, timing plays an important role in deciding when and how changes are made. For the TQRC report (2009), three-year retention data were not available for May 2005 program completers until 2008. Similarly, K-12 student test data require time for analysis and reporting. Even current state teacher growth scores for 2013 require two consecutive years of test data on file with the state (2011-2012, and 2012-2013) in order to calculate the growth score of students, and in turn, teachers, from one year to the next. Such a time lag impacts the ability to provide real time analysis of the education system for any parties involved, including schools of education, K-12 teachers, school leaders, students, and parents. Time also is required to monitor any changes made to evaluate their impact. At best the reports could offer a baseline from which to monitor performance over time; however, only the NYCDOE has indicated plans to continue with the reports; the TQRC report was a one-time effort. Such one-shot reports are severely lacking in their ability to provide meaningful feedback to the educator preparation profession. Systems invariably are a product of the environment and the people in them (Boulding, 1956; Katz & Kahn, 1966; Parsons, 1951; Senge, 1990). Systems, especially if they are to be considered effective, must establish feedback mechanisms and cycles, provide opportunities to set goals, and monitor progress towards those goals (Emery, 2000; Senge, 1990). Senge describes different types of feedback cycles, reinforcing or balancing, which either (1) reinforce or amplify a process within a system or (2) balance and stabilize a process in a system. In a similar way, the educator preparation in the United States, and specifically in New York State, is a system that needs continuous feedback mechanisms which will help balance or reinforce this profession. If the ultimate goal of educational change is to enhance the quality of teachers, then adjusting entry variables (e.g., licensure qualifications) is one part of the system. Others include adjustments to tenure and promotion decisions. As each of these new or revised processes are put in place, oneshot evaluations will not provide consistent and continuous input to help regulate the system as a whole.

Considerations for Policy Implications

Finally we consider the extent to which either of these reports (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a) provides relevant and actionable information for making broad policy changes with regard to educator preparation programs.

The current education reform agenda at both federal and state levels is focused on making every high school graduate ready for college or a career (U.S. Department of Education, 2011). As noted earlier, the assumption behind this agenda is that more rigorous accountability models will drive improvements system-wide.

It is not obvious, however, that either report discussed here has had or will have an impact on educator preparation policies or on policies relating to other components of the system. In part we hypothesize that effecting changes is difficult within a system and associated accountability model that is highly segmented, and that policy changes occur within, rather than across, system components. For example, the NYSED has already taken action to make revisions to the licensure process in the state, partly through the addition of a performance assessment, edTPA, that requires candidates to submit a portfolio with sections on planning and preparation; evidence of teaching practice, including video segments; and assessment of their own impact in the classroom. The state education department has also made revisions to the in-service teacher performance review process via the Annual Professional Performance Review (APPR), which requires multiple points of assessment, including leader and peer observations, growth scores, and other local evaluation metrics. These two evaluation tools, edTPA and APPR, are not linked to provide a developmental trajectory of teaching, however, and each only provides feedback within one particular component of the system: teacher education programs *or* the K-12 school.

In addressing results from these reports (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a) regarding clinical experiences and school partnerships, we noted that additional information that could be useful to assess the quality of pre-service preparation is tracking where teachers conduct their clinical practice component of their preparation. Many have argued that this component of pre-service preparation is critical, as highlighted in recommendations set forth by the report of the Blue Ribbon Panel on Clinical Preparation and Partnerships for Improved Student Learning commissioned by the National Council for Accreditation of Teacher Education (2010). One significant policy change that could emerge by examining evidence from each report from a system perspective is to require all student teachers to be placed only with those teachers rated as highly effective. For the first time, these new accountability models could allow cross-sector policy changes. It is now possible to identify highly effective teachers, and it would be possible to track outcomes over time from a policy change that paired teacher candidates with those most effective in the classroom. Considering and using information from multiple components of the state education system could enhance effectiveness within the system overall. In this case, selecting

highly effective mentor teachers could enhance the ability to transition entry-level teachers into the profession.

In summary, neither the TQRC (Boyd, Lankford, & Wyckoff, 2009) nor the NYCDOE (2013a) report appears to have had any direct influence on NYS educator preparation policy changes. As noted above, the TQRC report was a one-time effort, but it has likely served as a precursor to a new feedback report that the NYSED plans to issue for each IHE across the state in the coming year. It is not clear which variables and metrics will be used for these new state reports, and the NYSED would do well to review recommendations from Henry et al. (2012) regarding concerns about accuracy and fairness.

Overall there is a lack of a system perspective with regard to the evidence emerging from these new accountability models. What is clear is that educator preparation accountability will continue to be the subject of ongoing reports such as the ones discussed here, and thus, they reflect a policy trend in state wide comprehensive database development.

Conclusion

In reviewing both reports (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a), it is clear that the evidence for program improvement is sparse and that causality is difficult to determine. In fairness, neither report set out to provide a comprehensive set of causal linkages for program improvement purposes. It is also clear from a review of the reports' findings that the reports are limited in impact because, by their design, the responsibility and accountability for preparing effective educators is placed largely at the foot of IHEs rather than treated from a system perspective as an endeavor involving multiple stakeholders.

We recommend treating educator preparation from a system perspective involving candidates, IHEs, K-12 schools, and policy makers across a timeframe from program entry to early career. A system perspective could lead to more productive outcomes from accountability reports. For example, the absence of a systems approach to the evaluation of educator preparation programs means that program elements that might be most critical to producing effective teachers, such as the quality of faculty and program design, the quality of the clinical placement, the quality of the supervision, and the selection of mentor teachers do not fit neatly into currently available measures, but rather fall into gaps that currently exist in the accountability system. Plecki et al. (2012) in the conclusion to their study noted the need for cooperation among and across programs and institutions about "what elements matter" (p. 331), and therefore what variables and measures should be consistently obtained and used as a basis for improvement. This level of cooperation would help move educator preparation program accountability beyond a task undertaken by one agency about another agency.

A further recommendation is that all stakeholders should have the opportunity to develop specific questions to investigate. Neither report (Boyd, Lankford, & Wyckoff, 2009; NYCDOE, 2013a) presented here was driven by the development of specific questions about preparation program improvements by any of the stakeholders involved. For example, Plecki et al. (2012) were able to use a state database to address specific questions of interest concerning teachers' value-added scores in relation to years of teaching experience. Again, this would involve cooperation and collaboration among stakeholders about developing both the questions and suitable measures.

As noted there have been several changes in reporting capabilities in the four years between these reports. These changes include enhanced transparency and greater public access to information about program completers. A second major change is the capability to track impact in the classroom to individual teachers, allowing evaluation at an individual rather than group level. While facets of this process are controversial, we believe that the capability in and of itself is useful to help answer the question of teacher effectiveness. While there has been increased sophistication in the development of such measures to capture teachers' impact on student learning, this article points to a lack of sophistication in the processes by which educator preparation accountability is being developed. We recommend a system approach, driven by specific improvement questions, developed through collaboration around the critical dimensions of effective teaching.

References

- Boulding, K. E. (1956). General systems theory the skeleton of science. *Management Science*, *2*(3), 197-208. doi:10.1287/mnsc.2.3.197
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, *31*(4), 416-440. doi:10.3102/0162373709353129
- Boyd, D. J., Lankford, H., & Wyckoff, J. (2009). *Tracking program completers: Teacher quality research pilot report*. Unpublished report.
- Carnegie Corporation of New York. (2001). Teachers for a new era. *Carnegie Reporter,* 1(3), 36. Available from http://carnegie.org/publications/carnegie-reporter/single/view/article/item/36/
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school. A cross-subject analysis with student fixed effects. *The Journal of Human Resources*, *45*(3), 655-681. doi: 10.1353/jhr.2010.0023

- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(10). Retrieved from http://epaa.asu.edu/epaa/v8n1
- Donovan, C., Ashdown, J., & Mungai, A. (2013, February). Useful and relevant? One institution's findings from a state-wide, career path, pilot study linking teacher education data with school and student level data. Paper presented at the Annual Meeting of the American Association of Colleges of Teacher Education, Orlando, FL.
- Emery, M. (2000). The current version of Emery's open systems theory. *Systemic Practice and Action Research, 13*(5), 623-643. doi:
- Fallon, D. (2006). The buffalo upon the chimneypiece: The value of evidence. *Journal of Teacher Education*, *57*(2), 139-154. doi:10.1177/0022487105285675
- Gansle, K. A., Noell, G. H., & Burns, J. M. (2012). Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana. *Journal of Teacher Education*, *63*(5), 304-317. doi:10.1177/0022487112439894
- Goldhaber, D., & Anthony, E. (2004). *Can teacher quality be effectively assessed?*Washington, DC: University of Washington and the Urban Institute. Available from http://www.urban.org/publications/410958.html
- Henry, G. T., Kershaw, D. C., Zulli, R. A., & Smith, A. A. (2012). Incorporating teacher effectiveness into teacher preparation program evaluation. *Journal of Teacher Education*, *63*(5), 335-355. doi:10.1177/0022487112454437
- Hillon, M. E. (2005). A comparative analysis of socio-ecological and socio-economic strategic change methodologies. (Doctoral dissertation. Lyon 3).
- Katz, D. & Khan, R. L. (1966). *The social psychology of organizations*. New York, NY: Wiley.
- Kirby, S. N., McCombs, J. S., Naftel, S., & Barney, H. (2005). *Implementing teachers for new era: Some promising indicators of change* (Working Paper, W-R 278). Santa Monica, CA: Rand Corporation.
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37-62. doi: 10.3102/01623737024001037
- National Center for Educational Statistics (NCES). (2012). *State wide longitudinal data systems*. Retrieved from http://nces.ed.gov/programs/slds/stateinfo.asp

- National Council for the Accreditation of Teacher Education (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers.* Washington, DC: Author.
- New York City Department of Education (NYCDOE). (2013a). *Teacher preparation program reports 2013*. Retrieved from http://schools.nyc.gov/Offices/DHR/HumanCapitalData/TPPR.
- New York City Department of Education (NYCDOE). (2013b). Teacher preparation program reports: Raising the bar for students in schools. Retrieved from http://schools.nyc.gov/NR/rdonlyres/D9840D7D-7A36-4C66-817C-C48CFE5C017C/0/NYCDOETeacherPreparationProgramPresentation_August_2_013.pdf
- Organisation for Economic Cooperation and Development (OECD). (2010). *PISA 2009 Results: What students know and can do: Student performance in reading, mathematics and science* (Volume 1.) Paris, France: PISA, OECE Publishing. doi:10.1787/9789264091450-en
- Parsons, T. (1951). The social system. New York, NY: The Free Press.
- Plecki, M. L., Elfers, A. M., Nakamura, Y. (2012). Using evidence for teacher education program improvement and accountability: An illustrative case of the role of value added measures. *Journal of Teacher Education*, *63*(5), 318-334. doi:10.1177/0022487112447110
- Senge, P. M. (1990). The fifth discipline: The art and practice of the learning organization. New York, NY: Currency Doubleday.
- U.S. Department of Education (2011). *Our future, our teachers: The Obama administration's plan for teacher education reform and improvement.*Washington, DC. Retrieved from http://www.ed.gov/sites/default/files/our-future-our-teachers-accesible.pdf
- U.S. Department of Education (2014). *Race to the Top Fund*. Washington, DC. Retrieved from http://www2.ed.gov/programs/racetothetop
- Wiseman, D. (2012). The intersection of policy, reform, and teacher education. *Journal of Teacher Education*, 63(2), 87-91. doi:10.1177/0022487111429128

About the Authors



Corinne Baron Donovan, PhD, is the Assistant Dean of Research and Evaluation for the Ruth S. Ammon School of Education, Adelphi University, New York. Dr. Donovan's research includes a focus on career development of pre-service through in-service teachers, evaluation of teacher education, and development of teacher dispositional and attitude instruments. She has a broad background in adult learning and development as well as program evaluation and a degree in Industrial and Organizational Psychology. Email: cdonovan@adelphi.edu



Jane E. Ashdown, PhD, is dean and professor at the Ruth S. Ammon School of Education, Adelphi University. She has been an early childhood classroom teacher and directed a university-based early literacy intervention project that partnered with high need school systems to improve teaching effectiveness and children's literacy achievement. Her research interests include teacher development and strategies for improving educator preparation. Email: jashdown@adelphi.edu



Anne M. Mungai, PhD, is a Professor of Education, Chair of the Curriculum and Instruction Department and Director of the graduate Special Education program at the Ruth S. Ammon School of Education, Adelphi University, New York. Her research agenda over the last several years has revolved around the concept of multicultural issues, Special Education inclusion issues, staff development, gender issues, and learning. She is the author of the book Growing up in Kenya: Rural Schooling and Girls, the co-editor of the books Pathway to Inclusion: Voices from the Field, and co-editor of In the Spirit of Ubuntu: Stories of Teaching and Research. She has published articles on Multicultural Education and issues of Special Education Inclusion. Dr. Mungai has served as a consultant to several schools that have started inclusion and has provided service to schools as a professional developer and field researcher in the area of cognition and classroom instruction with racially,

ethnically, and linguistically diverse populations. Email: MUNGAI@adelphi.edu