

AASA JOURNAL OF SCHOLARSHIP & PRACTICE



Research and Evidence-Based Practice Advancing the Profession of Education Administration

Winter 2016/Volume 12, No. 4

Table of Contents

Board of Editors	2
Sponsorship and Appreciation	3
Editor's Commentary	4
Research Articles	
A Comparison of Higher-Order Thinking Between the Common Core State Standards and the 2009 New Jersey Content Standards in High School.	5
by <i>Dario Sforza, EdD; Christopher H. Tienken, EdD; Eunyoung Kim, PhD</i>	
Problems with Percentiles: Student Growth Scores in New York's Teacher Evaluation System	32
by <i>Drew Patrick, MEd</i>	
Doctoral Research in Educational Leadership: Expectations for Those Thinking About An Advanced Degree.	50
by <i>David J. Parks, PhD</i>	
The Glass Maze and Predictors for Successful Navigation to the Top Seat to the Superintendency . .	66
by <i>Denise DiCanio, EdD; Laura Schilling; Antonio Ferrantino, EdD; Gretchen Cotton Rodney; Tanesha Hunter, EdD; Elsa-Sofia Morote, EdD; Stephanie Tatum, PhD</i>	
Mission and Scope, Copyright, Privacy, Ethics, Upcoming Themes, Author Guidelines, Acceptance Rates & Publication Timeline.	86
AASA Resources	90

Editorial Review Board

AASA Journal of Scholarship and Practice
2015-2016

Editor

Kenneth Mitchell, *Manhattanville College*

Associate Editors

Barbara Dean, *AASA, The School Superintendents Association*

Kevin Majewski, *Seton Hall University*

Editorial Review Board

Sidney Brown, *Auburn University, Montgomery*

Gina Cinotti, *Netcog Public Schools, New Jersey*

Sandra Chistolini, *Universita` degli Studi Roma Tre, Rome*

Michael Cohen, *Denver Public Schools*

Betty Cox, *University of Tennessee, Martin*

Theodore B. Creighton, *Virginia Polytechnic Institute and State University*

Gene Davis, *Idaho State University, Emeritus*

Daniel Gutmore, *Seton Hall University*

Gregory Hauser, *Roosevelt University, Chicago*

Thomas Jandris, *Concordia University, Chicago*

Zach Kelehear, *University of South Carolina*

Theodore J. Kowalski, *University of Dayton*

Nelson Maylone, *Eastern Michigan University*

Robert S. McCord, *University of Nevada, Las Vegas*

Barbara McKeon, *Broome Street Academy Charter High School, New York, NY*

Margaret Orr, *Bank Street College*

David J. Parks, *Virginia Polytechnic Institute and State University*

Joseph Phillips, *Manhattanville College*

Dereck H. Rhoads, *Beaufort County School District*

Thomas C. Valesky, *Florida Gulf Coast University*

Published by

AASA, The School Superintendents Association

1615 Duke Street

Alexandria, VA 22314

Available at www.aasa.org/jsp.aspx

ISSN 1931-6569

Sponsorship and Appreciation

The *AASA Journal of Scholarship and Practice* would like to thank AASA, The School Superintendents Association, in particular the AASA Leadership Development Office, for its ongoing sponsorship of the Journal.

We also offer special thanks to Kenneth Mitchell, Manhattanville College, for his effort in selecting and editing the articles that comprise this professional education journal.

The unique relationship between research and practice is appreciated, recognizing the mutual benefit to those educators who conduct the research and seek out evidence-based practice and those educators whose responsibility it is to carry out the mission of school districts in the education of children.

Without the support of AASA and Kenneth Mitchell, the *AASA Journal of Scholarship and Practice* would not be possible.



Editor's Commentary

The Winter 2016 issue of the AASA Journal of Scholarship and Practice is about educational leadership. The first two articles—research studies written by Dario Sforza, Christopher Tienken, Eunyoung Kim and Drew Patrick—demand that those leading our schools consider research about the consequences that result from the misguided education laws and regulations set forth by policymakers and politicians. Leaders, especially superintendents, need to be scholar-practitioners, staying informed about the research, but then having the courage to challenge irresponsible and damaging policy, often driven by political agendas, coming from their state houses and the U.S. Department of Education.

This issue is also about ensuring that we not only prepare our school leaders to be such scholar-practitioners via rigorous and intellectually-demanding doctoral programs as found in David Parks study, but understand ways to expand the pool of eligible and talented leaders in the Elsa-Sophia Morote, et.al. study.

A Comparison of Higher-Order Thinking Between the Common Core State Standards and the 2009 New Jersey Content Standards in High School

Dario Sforza, EdD
Principal
Henry P. Becton Regional High School
East Rutherford, NJ

Christopher H. Tienken, EdD
Associate Professor
Department of Education Leadership
Management, and Policy
Seton Hall University
South Orange, NJ

Eunyoung Kim, PhD
Associate Professor
Department of Education Leadership,
Management, and Policy
Seton Hall University
South Orange, NJ

Abstract

The creators and supporters of the Common Core State Standards claim that the Standards require greater emphasis on higher-order thinking than previous state standards in mathematics and English language arts. We used a qualitative case study design with content analysis methods to test the claim. We compared the levels of thinking required by the Common Core State Standards for grades 9-12 in English language arts and math with those required by the New Jersey Core Curriculum Content Standards in grades 9-12 English language arts and math (used prior to the Common Core) using Webb's Depth of Knowledge framework to categorize the level of thinking required by each standard. Our results suggest that a higher percentage of the 2009 New Jersey high school curriculum standards in English language arts and math prompted higher-order thinking than the 2010 Common Core State Standards for those same subjects and grade levels. Recommendations for school administrative practice are provided.

Key Words

Common Core State Standards, standardization, higher-order thinking

According to officials from National Governors Association (NGA) Center for Best Practices and the Council of Chief State School Officers (CCSSO), the Common Core State Standards are “based on rigorous content and application of knowledge through higher-order thinking skills” and “informed by other top performing countries in order to prepare all students for success in our global economy and society” (NGA Center & CCSSO, 2015, About the Standards). An overt message we draw from the Common Core State Standards (CCSS) developers regarding their product is that the Standards are designed to ensure that students will have the knowledge and academic skills necessary to succeed in the global economy. Documentation on the official CCSS website presents “higher-order thinking skills” as a key component of the Standards (NGA Center & CCSSO, 2015, About the Standards). But what constitutes the higher-order skills necessary for success in the global economy?

Mainstream Calls for Higher-Order Thinking

Some commentators from business, economics, and education circles argue that the types of higher-order thinking skills that students need to be globally competitive include creative thinking and strategic thinking. For example, the IBM Corporation (2012), the United States Council on Competitiveness (2012), the Institute for Management Development (2012), the Organisation for Economic Co-operation and Development [OECD] (2013), Pink (2006), Robinson (2011), and Zhao (2012), and others identified variations of creative and/or strategic thinking they believe are important skills that high school graduates need in order to access better options for college, careers, and global economic competitiveness.

Similarly, Cisco Systems Inc., Intel Corporation, Microsoft Corporation, and the University of Melbourne (2010) drew similar conclusions from *The Assessing and Teaching of 21st Century Skills* (ATC21S) study. They found higher-order thinking related to greater global competitiveness. The results from the ATC21S identified and categorized skills that future employees will need in order to remain viable in the global economy. The ATC21S study divided the skills into four categories, one of which was based exclusively on creative and strategic thinking:

- Ways of thinking: creativity, critical thinking, problem solving, decision making, and learning
- Ways of working: communication and collaboration
- Tools for working: information and communications technology (ICT) and information literacy
- Skills for living in the world: citizenship, life and career, and personal and social responsibility

Andreas Schleicher (Asia Society, 2010), OECD’s head of the Programme for International Student Assessment (PISA) echoed the ATC21S findings of a need for higher-order non-routine competencies when he stated, “In the developing knowledge economy, workers are expected not to take orders, but to think in complex ways with ever-changing variables.” Schleicher’s emphasis on critical thinking was repeated in the United States by various business and education lobbying groups. The American Society for Training and Development (2010) identified “innovative thinking and action; the ability to think creatively and to generate new ideas and

solutions to challenges at work” as crucial competencies and skills students will need to succeed in the global economy (p. 13). The National Education Association (NEA), the largest public educator special interest group in the U.S., warned its members that their students will not be able to meet the varied demands of a global economy and join the 21st century workforce unless schools prepare them with the skills to “create and innovate” (NEA, 2012, p. 24).

Although the type of creative and strategic thinking that public school personnel should develop in students can be debated, there seems to be some agreement in the school-reform literature that creativity and strategic thinking have a role to play in P-12 education to prepare students for economic life beyond compulsory schooling. The literature on economic global competitiveness and the shift to a knowledge economy reflects a conviction shared by leading corporate voices and some education officials that successful education will need to place greater emphasis on creative and strategic thinking.

New Jersey context

As in almost 40 other states, the New Jersey education landscape is not immune to the perceived pressure to equip students with higher-order thinking skills. New Jersey Department of Education (NJDOE) officials adopted the Common Core State Standards on June 16, 2010; 14 days after the NGA and CCSSO (2010) officially released the final version of the standards. The NJDOE (2010) reiterated the CCSS creators’ claims on its state education Common Core website that the Standards will prepare New Jersey students for 21st century college and career expectations:

The Common Core State Standards, adopted by the New Jersey State Board of Education in 2010, define grade-level expectations from kindergarten through high school for what students should know and be able to do in English Language Arts (ELA) and mathematics to be successful in college and careers.

By replacing the former New Jersey state standards in ELA and math with the CCSS, New Jersey education officials implied that the CCSS are superior to the former NJ standards in those areas.

The concern with the skills necessary to compete economically in a global economy extends to systemic reform plans in New Jersey. For example, officials from the NJDOE (2012a) issued a warning about the need to improve high school graduates’ higher-order thinking in their *Education Transformation Task Force Final Report*:

The dramatically changed economic environment of the 21st century characterized by increased global competitiveness and a shift from an industrial to a knowledge-based economy has shed a harsh light on another achievement gap. There is a growing chasm between what we require children to learn to be eligible to graduate from high school and what they actually need to learn to be truly ready for college and career. (p. 3)

Officials at the NJDOE created policies that correspond with the CCSS creators’ claims of superior development of higher-order thinking and preparation for the global economy. The NJDOE leadership mandated

that all school district leaders fully align their K-12 curricula in ELA and math with the CCSS shortly after the NJ State Board of Education voted to adopt the Standards, as did other states like California, Tennessee, and Illinois. NJDOE officials also indicated in their state's application for a United States Department of Education Race To The Top Phase III grant that 100% of schools would use CCSS aligned curricula by the start of the 2014-2015 school year (NJDOE, 2012b).

New Jersey provides an example of what took place in almost 40 other states around the nation since the 2010 launch of the Common Core. In essence, it is a microcosm of changes happening at state education agencies across the country. New Jersey was one of the first states to sign on to the Common Core and also a founding member of the Partnership for Assessment of Readiness for College and Careers consortium (PARCC), one of the two national testing bodies that created tests aligned to the Common Core, and thus represents an early adopter of the large-scale curriculum national standardization movement.

High school focus

The CCSS claims of enhancing higher-order thinking and global competitiveness seem to resonate most concretely in high school. High school represents the end of compulsory schooling, and, according to the information posted on the official CCSS website, "The standards define the knowledge and skills students should gain throughout their K-12 education in order to graduate high school prepared to succeed in entry-level careers, introductory academic college courses, and workforce training programs" (NGA Center & CCSSO, 2015, About the Standards). Policies adopted by New Jersey education officials

signal that high school curriculum standards play an important role in ensuring that students will graduate with the skills necessary to compete in the global economy. One example of the NJDOE officials' concern about raising the level of thinking in high school is their continued emphasis on high school exit exams. Not only did they reaffirm their commitment to high school exit exams, NJDOE officials also took the additional step of increasing the number of mandated exams from two to six, all of which must be aligned to the CCSS.

Given the rhetoric regarding the ability of the CCSS to prepare all students for all colleges and careers in a global knowledge economy, one might expect to see creativity and strategic thinking embedded throughout the CCSS high school standards for English language arts (ELA) and mathematics (M) more so than previous versions of New Jersey curriculum standards in those subjects.

Problem, purpose, and questions

No qualitative analytical research has been done to test the assumption that the CCSS are superior to previous state standards in the development of higher-order thinking and creativity at the high school level. Our purpose for this qualitative case study using content analysis techniques was to describe and compare the percentages of the CCSS and former New Jersey Core Curriculum Content Standards (NJCCCS) in ELA and M that require students to demonstrate strategic and/or creative thinking at the high school level.

Three questions guided our study:

1. To what extent are creative and strategic thinking, as defined by Webb's Depth of Knowledge, embedded in the

Common Core State Standards for English Language Arts and Mathematics for grades 9-12?

2. To what extent are creative and strategic thinking, as defined by Webb's Depth of Knowledge, embedded in the New Jersey Core Curriculum Content Standards for English Language Arts and Mathematics for grades 9-12?
3. What differences and similarities exist in creative and strategic thinking between the Common Core State Standards and New Jersey Core Curriculum Content Standards in English Language Arts and Mathematics for grades 9-12?

Significance

Our study includes an important innovation over previous works by not only including all CCSS anchor standards, but also drilling down to the sub-standards or individual learning objectives embedded within each standard. Sat et al.'s (2011) Smarter Balanced Study deviated from Webb's (2005) recommendations by giving multiple ratings to one Common Core anchor standard to account for all the sub-standards.

For example they labeled ELA RL.9-10.1 as a DOK 1, 2, and 3. Therefore one standard could receive credit as a 3 even if it were populated by a majority of Level 1 objectives.

We sought to provide greater precisions with our ratings and gave one DOK rating per standard and rated each sub-standard. Another study, Florida State University's (2012) CPALMS study, gave one rating for each Common Core standard and sub-standard within the Grades 9-12 ELA and Math CCSS

and NJCCCS. The precision in our methods translates to greater precision of the results and more a complete picture of the CCSS.

Literature Touchstones

Conceptual framework

There have been various attempts to define what constitutes higher-order thinking in the public high school curriculum. The mainstream, non-empirical, literature on standards-based education reform tends to group creativity, innovation, entrepreneurship, and strategic or critical thinking together. However, scholarly frameworks allow researchers to deconstruct and categorize curriculum standards according to expected levels of cognition or thinking. Webb's (1997; 2007) Depth of Knowledge (DOK) is one such framework.

According to Webb (1997), Depth of Knowledge encompasses multiple dimensions of thinking, including the "level of cognitive complexity of information students should be expected to know, how well they should be able to transfer the knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas" (Webb, 1997, p. 15). DOK is a way to define and categorize cognitive complexity of curriculum standards and tasks. The "DOK level of an item does not refer to how easy or difficult a test item is for students" (Wyse & Viger, 2011, p. 188). The focus of DOK is on the cognitive complexity of required tasks or curriculum standards.

Complexity Versus Difficulty

Although complexity and difficulty are necessary components of an intended curriculum, the Depth of Knowledge or

complexity of a learning objective is dynamic and encompasses the multiple dimensions of an objective ranging from the “level of cognitive complexity of information students should be expected to know, how well they should be able to transfer this knowledge to different contexts, how well they should be able to form generalizations, and how much prerequisite knowledge they must have in order to grasp ideas” (Webb, 1997, p. 15).

Sousa (2006) defined complexity as the thought processes required to address a task. Complexity can be thought of as the difference between remembering a fact or imitating a procedure and developing an original product, conclusion, or process. Remembering facts and imitating procedures are less cognitively complex than developing an original conclusion, product or process.

Difficulty is a more static component of a learning objective that simply refers to the amount of work or effort a student must use to complete a task, regardless of complexity. For example, asking students to solve an addition problem with two one-digit numbers is less difficult than solving the same problem with four one-digit numbers. The complexity is still at the “remember and imitate” procedure level, but the second problem is theoretically more difficult because it requires more effort to add more numbers. Our concern rests with cognitive complexity.

DOK levels

Webb (1997) described Depth of Knowledge within an educational objective as cognitively complex, involving the numerous connections students make from prior knowledge to current knowledge using strategic and extended forms of thinking in order produce an idea that is

original and purposeful (p. 15). We used Webb’s (1997; 2007) four DOK levels as lenses through which to deconstruct and describe the cognitive complexity of the CCSS and former 2009 NJCCCS in grades 9-12 for ELA and M for this study:

Level 1 (recall): Standards at this level require students to recall a simple definition, term, or fact, or replicate a procedure, or algorithm.

Level 2 (skill/concept): Standards at this level require students to develop some mental connections and make decisions about how to set up or approach a problem or activity to produce a response, apply a recalled skill, or engage in literal comprehension.

Level 3 (strategic thinking): Standards at this level require students to engage in planning, reasoning, constructing arguments, making conjectures, and/or providing evidence when producing a response and require students to do some complex reasoning and make original concepts or draw conclusions.

Level 4 (extended thinking): Standards at this level require students to engage in complex planning, reasoning, and conjecturing, and to develop lines of argumentation. Items at this level require students to make multiple connections between several different key and complex concepts, inferencing, or connecting the dots to create a big picture generalization.

Depth of Knowledge includes multiple forms of knowledge such as declarative, which is based on facts, and procedural, which can be

described as practical “know-how” (Runco & Chand, 1995, p. 245). Declarative knowledge is linked to procedural knowledge; together they form the foundation that structures creative and strategic thinking opportunities. Levels 1 and 2 of Webb’s DOK focus on declarative and procedural knowledge (in other words, recall and basic application). Although basic application of material is the first of many steps involved in creative and strategic thought, thinking does not stop at the declarative and procedural levels.

Webb’s Levels 3 and 4 include creative and strategic thinking and provide opportunities for students to experience deeper, analytical, and more divergent types of thinking. Sternberg (1999) asserts that creativity is the “aptitude to generate work that is unique and original as well as suitable for the specific task or problem one is attempting to solve” (p. 3); this comports with Webb’s higher levels of DOK.

We equated DOK Levels 3 and 4 with the types of thinking that commentators in the mainstream literature on standards-based education reform refer to when they call for students to develop higher-order thinking skills. Webb (1997) views DOK Levels 3 and 4 as the levels at which students have opportunities to be flexible, creative, and strategic in their thinking because they are not bound to converge on one correct answer or to imitate one procedure.

If a set of curriculum standards does not have an appropriate flexible mix of cognitive complexity, including various DOK levels of thinking, students have fewer opportunities to gain the consistent learning experiences they need in order to think effectively at Webb’s DOK 3 and 4 levels of cognition. Their

thinking can become somewhat rigid if they receive a predominance of declarative and procedural thinking opportunities (Runco and Chand, 1995; Sternberg, 2003). If cognitive flexibility is not embedded in the standards and they are over-weighted with Level 1 and 2 standards, students will reach what Runco and Chand (1995) call “functional fixedness” (as cited in Ward, Smith, & Finke, 2010, p. 201, p. 247).

Functional fixedness is “the rigidity or mental set that locks thinking so an individual cannot see alternatives” (Runco and Chand, 1995, p. 247). A curriculum standard with functional fixedness would be categorized as a Level 1 recall or, a Level 2 basic application in terms of Webb’s DOK. Standards at levels 1 and 2 do not have the divergent thinking opportunities needed to develop cognitive flexibility and they are dominated by convergent thinking aimed at finding one correct, pre-determined answer based on imitation processes.

If the purposeful cognitive design of curriculum standards and the dangers of functional fixedness are understood during the creation of curriculum standards, then standards can potentially increase cognitive “originality and flexibility,” by ensuring that a mix of cognitive levels appears throughout the standards in each subject and for each grade level (Runco & Chand, 1995, p. 245). Although curriculum standards focused on procedural and declarative knowledge are not the lead actors in fostering creative and strategic thinking, they do play a supporting role.

Procedural and declarative knowledge provide a foundation needed to reach complex and extended forms of thinking; however, too much focus on the lower levels of thinking can

crowd out opportunities for more divergent thinking and turn students into “intellectual clones” (Sternberg, 2003, p. 335). If deeper levels of cognitive demand are absent and content is repetitive in nature, standards can jeopardize complex efforts to help students become creative and original thinkers (Runco & Chand, 1995, p. 245).

DOK Examples in the Content Areas

Attributes and key words for each DOK level provide descriptive language and concrete

boundaries for abstract concepts like strategic thinking. Each DOK level in Webb’s framework describes a specific type of thinking and its associated cognitive complexity. In general, the higher the cognitive complexity of a standard, the more creativity and strategic thinking will be embedded in it.

Below are example descriptions we used to frame the parameters of the levels of thinking for the purposes of this study:

Mathematics DOK Level 1. Standards at Level 1 require the recall of information such as basic facts, definitions, mathematical terms, as well as the ability to follow through a process by performing a simple algorithm or applying a formula. A one-step, well-defined algorithmic procedure should be included at Level 1.

Mathematics DOK Level 2. A Level 2 standard requires students to make some decisions regarding how to approach the problem or activity; whereas, a Level 1 only requires students to demonstrate a rote response, perform a previously learned algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that might distinguish a Level 2 item include “classify,” “organize,” “estimate,” “make observations,” “collect and display data,” and “compare data.” These prompt students to perform multi-step procedures.

Mathematics DOK Level 3. Curriculum standards at this level require reasoning, planning, using evidence to generate an original thought or interpretation, and doing more complex and inventive thinking than the previous two levels. Problems that ask students to explain their thinking by making original inferences or conclusions, beyond regurgitating memorized steps or processes, and make conjectures can be classified as Level 3. The cognitive demands at Level 3 are non-standard, complex, open-ended, and more abstract. The complexity results from the standards requiring more demanding creative reasoning.

Mathematics DOK Level 4. Students must demonstrate complex reasoning, planning, developing, and strategic thinking, usually over an extended period of time. Extended time is not a requirement for Level 4, but it is often a component of the type of cognitive work done at this level. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2. However, if the student conducts a river study that requires interpreting and drawing conclusions from data and proposing original solutions, based on evidence, to a non-standard problem, based on multiple variables and data points collected over time, the problem would be Level 4. The

work is complex and divergent. More often there is not a single answer, as much as there are original conclusions or interpretations that are reached and multiple, nonstandard ways to arrive at them. Students are generally required to make several connections within the content area and among content areas.

Reading Level 1. This level requires students to remember or recite facts or to use simple skills or abilities. Oral reading and basic comprehension of a text (but not analysis of a text) are included. Questions require only a shallow understanding of the text presented and often consist of verbatim recall, slight paraphrasing of specific details from the text, or simple understanding of a single word or phrase.

Reading Level 2. Level 2 involves some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text. Inter-sentence analysis or inference is required. Questions at this level might include words like “cite evidence,” “summarize,” and “explain.” Students might also be asked to determine whether a statement is a fact or an opinion. Literal main ideas are stressed.

Reading Level 3. Deep knowledge becomes a greater focus at Level 3. Students must show an understanding of the ideas in the text and are encouraged to go beyond the text to make connections. Students might be prompted to explain, generalize, or connect ideas. Standards at Level 3 involve reasoning and planning; students must be able to support their conclusions or interpretations. Questions might involve abstract theme identification, inference across an entire passage, or the application of prior knowledge to form a generalization.

Reading Level 4. Higher-order thinking is central and deep knowledge is required at Level 4. The standard at this level will probably require participation in a longer-term activity that is non-repetitive and requires the application of significant conceptual understanding and divergent thinking. Students must take information from at least one passage of a text and apply this information to a new task or in an original way or to create and support original conclusions and interpretations. They might also be asked to develop hypotheses and perform complex analyses of the connections among texts in order to develop original ideas, uses, processes, or productions from knowledge.

Writing Level 1. Level 1 requires the student to develop basic ideas and write facts from recall. The students might be asked to list ideas, words, or simple sentences, the way one might work during a brainstorming activity. They might also be required to copy notes from a pre-made source. Students are expected to write, speak, and edit using the conventions of Standard English and they are required to demonstrate a basic understanding and appropriate use of reference materials, such as a dictionary or thesaurus.

Writing Level 2. Level 2 requires some degree of mental processing. At this level, students engage in first-draft writing or brief extemporaneous speech for a limited number of purposes

and audiences. Students are expected to begin connecting ideas to form paragraphs and might also need to work at independent note taking, outlining, or summarizing.

Writing Level 3. Students develop original works with multiple paragraphs that include complex sentence structure and demonstrate some synthesis and analysis of a topic. Students show awareness of their audience and purpose through focus, organization, and the use of appropriate compositional elements such as voice. At this stage, students use known criteria to independently engage in editing and revising to improve the quality of the composition.

Writing Level 4. A curriculum standard at this level would involve writing a multi-paragraph composition that demonstrates the ability to synthesize, analyze, and develop complex ideas or themes. Students should demonstrate a deep awareness of purpose and audience. For example, informational papers should include hypotheses and supporting evidence and original interpretations or conclusions.

Methodology

We used a qualitative case study design with content analysis methods to describe and compare the percentages of the CCSS and of the former New Jersey Core Curriculum Content Standards (NJCCCS) in ELA and M that require students to demonstrate strategic and/or creative thinking.

Qualitative content analysis refers to research methods for interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns” (Hsieh & Shannon, 2005, p. 1278).

The content analyzed in this study consisted of CCSS and NJCCCS documents presenting the curriculum content standards for grades 9-12 mathematics and English language arts.

Deductive category application was used to connect Webb’s existing Depth of Knowledge framework to the high school CCSS and NJCCCS in ELA and M (Mayring, 2000). Figure 1 shows the *Step Model* of deductive category application, as described by

Mayring (2000), that we used to guide the process of coding and analyzing the standards. Hsieh and Shannon (2005) stressed that the “success of a content analysis depends greatly on the coding process” (p. 1285). The coding activities for each set of standards in each subject area and grade level followed the same procedure as described by Mayring (2000).

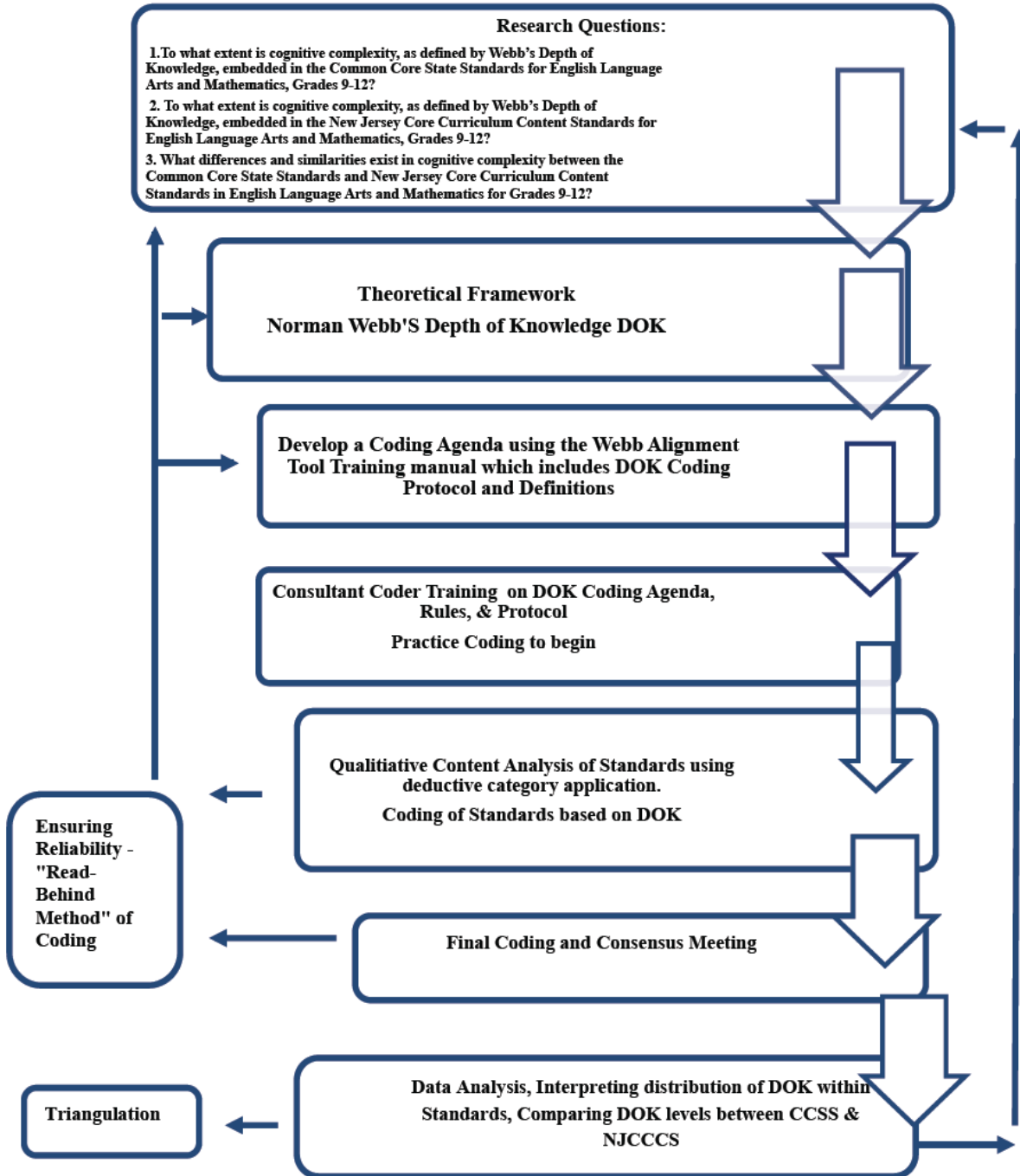
Instead of aligning the standards with an external assessment, as is commonly done in alignment studies, we compared the cognitive complexity of one set of curriculum standards to another based on DOK levels. School districts across the country are mandated by their state education agencies to align their curriculum to the CCSS, not an assessment.

We analyzed and coded the grades 9-12 Common Core English language arts and mathematics standards and the grades 9-12NJCCCS in English language arts and mathematics based on their corresponding DOK levels. Each standard was assigned a 1-4 Depth of Knowledge level based on Webb’s Depth of Knowledge methodology. Utilizing Mayring’s (2000) step model as the guide (see

Figure 1), a coding agenda was created using the DOK definitions, examples, and coding

rules as described in the Webb Alignment Tool (WAT) training manual (Webb, et al., 2005).

Figure 1. Mayring's (2000) step model used to guide analyses.



Coding

Webb's Alignment Tool (WAT) training manual contains definitions, explanations, and examples for coders to reference and specifically understand how the DOK levels should read for English Language Arts and Mathematics objectives. We used two trained coders to analyze and code each set of standards. Webb's definitions of each DOK level helped ensure the coders' reliability and consistency as they rated each standard (Webb, et al., 2005, p. 36).

Below are samples of the rules—adapted from the WAT training manual—that the two coders followed when assigning DOK levels to each standard.

- The DOK level of an objective should be the level of work students are most commonly required to perform at that grade level to successfully demonstrate their attainment of the objective.
 - The DOK level of an objective should reflect the *complexity* of the objective, rather than its *difficulty*. The DOK level describes the kind of thinking involved in a task, not the likelihood that the task will be completed correctly.
 - In assigning a DOK level to an objective, coders should consider the complete domain of items that would be appropriate for measuring the objective and identify the depth-of-knowledge level of the most common of these items.
 - If there is a question regarding which of two levels an objective matches, such as Level 1 or Level 2, or Level 2 or Level 3, it is usually appropriate to select the higher of the two levels.
- The team of reviewers should reach consensus on the DOK level for each objective before coding any items for that grade level.

Two coders using Webb's coding protocol have already proven to be effective in two large-scale studies that used the WAT to analyze and code standards based on DOK complexity (Yuan & Le, 2012; Sato et al., 2011). Each deductive category within Mayring's (2000) step model (See Figure 1) has explicit descriptions, examples, and DOK coding rules adapted from the WAT (Webb, et al., 2005) training manual.

The descriptions, examples, and coding rules helped to increase the probability that coders understood thoroughly which DOK level should be assigned to each standard. Mayring's step model was adapted and revised for this study to include descriptions of Webb's depth of knowledge (DOK) levels excerpted from the Web Alignment Tool (WAT) training manual (Webb, 2005, p. 45-46, 70-75). Two coding agendas were developed, one for all mathematics standards and one for all English language arts standards. Webb's DOK wheel was used as an additional reference tool to increase the reliability and consistency of the coding process.

Reliability

According to Merriam (2009), documentary data are persuasive, allowing little room for the researcher to "alter what is being studied" (p. 155). A document content analysis is valid in the context of this study because it is "grounded in the product in which it was produced and therefore grounded in the real world" (Merriam, 2009, p. 156). In order to

increase the reliability of the findings between coders and the overall credibility of the results, the findings of this study were compared to previous studies, for which researchers coded the Common Core State Standards using the WAT for alignment purposes.

Another step we took to increase the coders' reliability was a "double-rater read behind consensus model," which proved effective in coding standards for other studies (Miles, Huberman, & Saldaña, 2014, p. 84; Sato, Lagunoff, & Worth, 2011, p. 11). Maxwell (2005) recommended using member checks to ensure the credibility of research. We used member checks as an additional inter-rater reliability strategy. The member checks allowed us to validate the coding analyses completed by the first coder using those of the second coder (p. 111). Both analysts in this study used the same data, coding agenda, and rules of coding.

Content clustering or grouping of standards, similar to those used in Sato et al.'s (2011) study, was used in coding the standards for this study. We used content clustering in cases when the content of one standard or a portion of a standard overlapped with another standard or strand (Sato, et al., 2011). The content clustering allowed us to make more reliable decisions about the DOK of overlapping standards.

Niebling (2012) provided an important warning that we heeded in preparing our coding standards: "Perhaps the most complicated work involved in using the Webb alignment model is helping coders of standards, objectives, and test items understand and reliably code them according to the DOK framework" (p. 12). Along with the preparation described above, we held preparatory meetings with coders prior to

the coding sessions to discuss the methods. Coders also participated in practice coding sessions to ensure that they fully understood the coding process as well as the member check and double-rater read behind methods.

The analysts completed two practice sessions prior to the formal coding meetings. The practice sessions allowed time for the coders to familiarize themselves with the specific coding situation comparing one set of standards to another and allowed for inter-rater reliability calibration.

After the initial training meetings, the coding team read and coded the grades 9-12 NJ M and ELA standards (2009), using the "double-rater read behind consensus model" (Sato, Lagunoff, & Worth, 2011, p. 11). The second analyst reviewed the DOK findings of the first analyst and noted agreements or disagreements with each coded standard. Any disagreements were noted and discussed in follow-up meetings.

The double-rater read behind consensus model continued with the grades 9-12 CCSS in ELA and Math. Following the completion of all coding for the NJCCCS and CCSS, the coders compared their CCSS findings with Florida State University's CPALMS (2012) study, which rated all CCSS based on DOK.

This triangulation strategy of using the double read behind method and comparing the coders' results with those from previous studies increased the validity of our findings. A final member check meeting was held at the completion of each coding session to compare the completed findings of the coded CCSS from our sessions to those of the results from the study of Florida's state mandated standards, known as CPALMS (2012), in an effort to

increase reliability among coders and to external results.

During instances of disagreement, the coders followed a protocol to attempt to reach consensus. For example, there was initial disagreement on the DOK level for a CCSS ELA standard. One coder rated it a Level 3 and the other rated it at Level 2. Although one rater felt the ELA standard could be rated at a DOK Level 2, the rater who coded the standard at a DOK Level 3 explained why it should be rated at a DOK Level 3, providing specific examples and descriptions from the WAT training manual to support the rating.

The rationale was that students had to use strategic skills in order to analyze the specific literature listed in the standard; therefore, a DOK Level 3 rating was appropriate because it satisfied more of the

descriptions found in Level 3 than Level 2. Coders followed Webb's et al. (2005) recommendation and used the higher of the two DOK levels in rare cases in which they could not reach consensus.

Findings

Overall, the high school Common Core State Standards in ELA and M contained fewer standards rated at DOK Levels 3 and 4 than the 2009 New Jersey high school standards in ELA and math. That is, the standards that NJ had in place prior to adopting the Common Core provided more of the Level 3 and 4 higher-order skills cited in mainstream business and education publication as necessary capabilities for competing in a global economy.

The following sections provide an account of the results for each subject area as they relate to each research question.

CCSS high school standards

Our first research question asked: To what extent is cognitive complexity, as defined by Webb's Depth of Knowledge, embedded in the high school Common Core State Standards for English Language Arts and Mathematics for grades 9-12?

CCSS English language arts

Level 1 and 2 Depth of Knowledge complexity accounted for 72% of the high school ELA Common Core State Standards. Thirty-seven percent (37%) of the 9-12 CCSS ELA standards were rated at Level 1. Two examples of grades 9-12 CCSS ELA standards coded at a DOK Level 1 were:

Reading, grades 9-10: 9-10.RL.10. By the end of Grade 9, read and comprehend literature, including stories, dramas, and poems, in the grades 9–10 text complexity band proficiently, with scaffolding as needed at the high end of the range.

Writing, grades 11-12: 11-12.W.3.d. Use precise words and phrases, telling details, and sensory language to convey a vivid picture of the experiences, events, setting, and/or characters.

The distribution of ELA standards coded at a DOK Level 2 in the grades 9-12 ELA CCSS was 35%. Two examples of grades 9-12 ELA standards coded at a DOK Level 2 were:

Writing, grades 9-10: 9-10.W.9. Draw evidence from literary or informational texts to support analysis, reflection, and research.

Reading, grades 11-12: 11-12.RI.2. Determine two or more central ideas of a text and analyze their development over the course of the text, including how they interact and build on one another to provide a complex analysis; provide an objective summary of the text.

DOK Level 3 standards made up 26% of the CCSS grades 9-12 ELA. Deeper cognitive processing, strategic thinking, and more complex understanding are emphasized in ELA standards coded at a DOK Level 3. “Editing and revising” to add original ideas, not error identification, as well as the ability to provide evidence of student thinking were important components of ELA standards coded at a DOK Level 3. Furthermore, standards coded at DOK Level 3 prompted students to look beyond the required text and create essays by explaining, generalizing, and connecting ideas. Two examples of grades 9-12 ELA standards coded at a DOK Level 3 were:

Reading, grades 9-10: 9-10.RI.7. Analyze various accounts of a subject told in different mediums (e.g., a person’s life story in both print and multimedia), determining which details are emphasized in each account.

Writing, grades 11-12: 11-12.W.2.a. Introduce a topic; organize complex ideas, concepts, and information so that each new element builds on that which precedes it to create a unified whole; include formatting (e.g., headings), graphics (e.g., figures, tables), and multimedia when useful to aid comprehension.

The distribution of standards rated at a DOK Level 4 in the grades 9-12 ELA CCSS was only 2%. Extended activities with multi-paragraph essays and the ability to apply, analyze, critique, create, and connect ideas with empirical evidence were strong components of ELA standards coded at DOK Level 4. Two examples of grades 9-12 ELA standards coded at a DOK Level 4 were:

Writing, grades 9-10: 9-10.W.7. Conduct short as well as more sustained research projects to answer a question (including a self-generated question) or solve a problem; narrow or broaden the inquiry when appropriate; synthesize multiple sources on the subject, demonstrating understanding of the subject under investigation.

Reading, grades 11-12: 11-12.RI.9. Analyze seventeenth-, eighteenth-, and nineteenth-century foundational U.S. documents of historical and literary significance (including The Declaration of Independence, the Preamble to the Constitution, the Bill of Rights,

and Lincoln's Second Inaugural Address) for their themes, purposes, and rhetorical features.

CCSS mathematics

As was the case with the CCSS ELA standards, lower-level declarative and procedural thinking dominated the mathematics CCSS, with 90% rated as either DOK Level 1 or 2. The distribution of standards rated at a DOK Level 1 in the grades 9-12 Mathematics CCSS was 19%. Two examples of grades 9-12 Math CCSS coded at a DOK Level 1 were:

Math, grades 9-12 (The Real Number System): N.RN.2. Rewrite expressions involving radicals and rational exponents using the properties of exponents.

Math, grades 9-12 (Congruence): G.CO.7. Use the definition of congruence in terms of rigid motions to show that two triangles are congruent if and only if corresponding pairs of sides and corresponding pairs of angles are congruent.

The distribution of standards rated at a DOK Level 2 in the grades 9-12 mathematics CCSS was 71%. DOK Level 2 mathematics standards had language that prompted students to make judgments and observations about how to solve problems and to classify and compare different data sets (Webb, et al., 2005). Two examples of grades 9-12 Math CCSS coded at a DOK Level 2 were:

Math, grades 9-12 (Vector and Matrix Quantities): N.VM.3 (+). Solve problems involving velocity and other quantities that can be represented by vectors.

Math, grades 9-12 (Similarity, Right Triangles, And Trigonometry): G.SRT.11 (+). Understand and apply the Law of Sines and the Law of Cosines to find unknown measurements in right and non-right triangles (e.g., surveying problems, resultant forces).

The distribution of standards rated at a DOK Level 3 in the grades 9-12 Mathematics CCSS was 10%. To be rated a DOK Level 3, math standards needed to include language that created a valid argument for complex problems and situations that could yield more than one right answer or original conclusion. Two examples of grades 9-12 Math CCSS coded at a DOK Level 3 were:

Math, grades 9-12 (Seeing Structure in Expressions): A.SSE.4. Derive the formula for the sum of a finite geometric series (when the common ratio is not 1), and use the formula to solve problems. For example, calculate mortgage payments.

Math, grades 9-12 (Building Functions): F.BF.1.b. Combine standard function types using arithmetic operations. For example, build a function that models the temperature of a cooling body by adding a constant function to a decaying exponential, and relate these functions to the model.

None or 0%, of CCSS mathematics standards in grades 9-12 were rated as DOK Level 4 in grades 9-12.

New Jersey high school standards

Our second research question asked: To what extent is cognitive complexity, as defined by Webb's Depth of Knowledge, embedded in the New Jersey Core Curriculum Content Standards for Language Arts Literacy and Mathematics for grades 9-12?

NJ high school English language arts (ELA)

DOK Levels 1 and 2 accounted for 62% of the NJ ELA standards. The distribution of DOK Level 1 in the grades 9-12 ELA NJCCCS was 22%. Two examples of grades 9-12 ELA NJCCCS coded at a DOK Level 1 were:

Reading, grades 9-12: 3.1.12.D.1. Read developmentally appropriate materials (at an independent level) with accuracy and speed.

Writing, grades 9-12: 3.2.12.A.6. Review and edit work for spelling, usage, clarity, and fluency.

The distribution of NJCCSS standards coded at a DOK Level 2 in grades 9-12 ELA was 40%. ELA standards coded at a DOK Level 2 often required comprehension and continued processing of reading, along with unplanned speaking and simple writing tasks. Two examples of grades 9-12 ELA NJCCCS coded at a DOK Level 2 were:

Reading, grades 9-12: 3.1.12.A.2. Identify interrelationships between and among ideas and concepts within a text, such as cause-and-effect relationships.

Writing, grades 9-12: 3.2.12.B.13. Write sentences of varying length and complexity, using precise vocabulary to convey intended meaning.

The distribution of standards coded at a DOK Level 3 in the grades 9-12 ELA NJCCCS was 33%. Two examples of grades 9-12 ELA NJCCCS coded at a DOK Level 3 were:

Reading, grades 9-12: 3.1.12.E.1. Assess and apply reading strategies that are effective for a variety of texts (e.g., previewing, generating questions, visualizing, monitoring, summarizing, evaluating).

Writing, grades 9-12: 3.2.12.B.3. Draft a thesis statement and support/defend it through highly developed ideas and content, organization, and paragraph development.

The distribution of standards rated at a DOK Level 4 in the grades 9-12 ELA NJCCCS was 5%. Two examples of grades 9-12 ELA NJCCCS coded at a DOK Level 4 were:

Reading, grades 9-12: 3.1.12.G.2. Analyze how our literary heritage is marked by distinct literary movements and is part of a global literary tradition.

Writing, grades 9-12: 3.2.12.D.2. Write a variety of essays (e.g., a summary, an explanation, a description, a literary analysis essay) that develop a thesis; create an organizing structure appropriate to purpose, audience, and context; include relevant information and exclude extraneous information; make valid inferences; support judgments with relevant and substantial evidence and well-chosen details; and provide a coherent conclusion.

NJ high school mathematics

Levels 1 and 2 represented 62% of the NJCCSS math standards in high school. The distribution of standards rated at a DOK Level 1 in the grades 9-12 Mathematics NJCCCS was only 8%. Two examples of grades 9-12 Math NJCCCS coded at a DOK Level 1 were:

Math, grades 9-12 (Geometry and Measurement): 4.2.12 C.3. Find an equation of a circle given its center and radius, and, given an equation of a circle in standard form, find its center and radius.

Math grades 9-12 (Patterns and Algebra): 4.3.12 D.2. Select and use appropriate methods to solve equations and inequalities (e.g. linear equations and inequalities – algebraically; quadratic equations and factoring including trinomials when the coefficient of x^2 is 1, and using the quadratic formula; literal equations; solve all types of equations and inequalities using graphing, computer, and graphing calculator techniques).

The distribution of standards rated at a DOK Level 2 in the grades 9-12 Mathematics NJCCCS was 54%. Two examples of grades 9-12 Math NJCCCS coded at a DOK Level 1 were:

Math, grades 9-12 (Numbers and Numerical Operations): 4.1.12 A.2. Compare and order rational and irrational numbers.

Math, grades 9-12 (Mathematical Processes): 4.5 F.4. Use calculators as tools to problem-solve (e.g., to explore patterns and validate solutions).

The distribution of standards rated at DOK Levels 3 and 4 was 38%. Level 3 standards accounted for 28% of the NJ 9-12 mathematics standards. Two examples of grades 9-12 Math NJCCCS coded at a DOK Level 3 were:

Math, grades 9-12 (Patterns and Algebra): 4.3.12 C.2. Analyze and describe how a change in an independent variable leads to change in a dependent one.

Math, grades 9-12 (Mathematical Processes): 4.5 A.2. Solve problems that arise in mathematics and in other contexts (i.e. open-ended problems; non-routine problems; problems with multiple solutions; problems that can be solved in several ways).

The distribution of standards rated at DOK Level 4 in the grades 9-12 Mathematics NJCCCS was 10%. Two examples of grades 9-12 Math NJCCCS coded at DOK Level 4 were:

Math, grades 9-12 (Mathematical Processes): 4.5 B.3. Analyze and evaluate the mathematical thinking strategies of others.

Math, grades 9-12 (Data Analysis, Probability, and Discrete Mathematics): 4.4.12 A.2. Evaluate the use of data in real-world contexts (e.g. accuracy and reasonableness of conclusions drawn; correlation versus causation; bias in conclusions drawn; statistical claims based on sampling).

Comparisons

Our third research question asked: What differences and similarities exist in creative and strategic thinking between the Common Core State Standards and the New Jersey Core Curriculum Content Standards in English Language Arts and Mathematics for grades 9-12?

We found a 10% difference in high school ELA standards categorized as Level 3 or 4 favoring the former NJ standards compared to the CCSS. There was a 26% difference in higher-order thinking favoring the NJ math standards compared to the CCSS (See Table 1 & Figures 2 -5).

Table 1

DOK Comparisons for High School CCSS and NJ ELA and M Standards

	Levels 1 & 2	Levels 3 & 4
CCSS ELA	72%	28%
NJ ELA	62%	38%
CCSS M	90%	10%
NJ M	62%	38%

Figure 2. Comparison of cognitive complexity between the Grades 9-12 ELA CCSS and Grades 9-12 ELA NJCCCS.

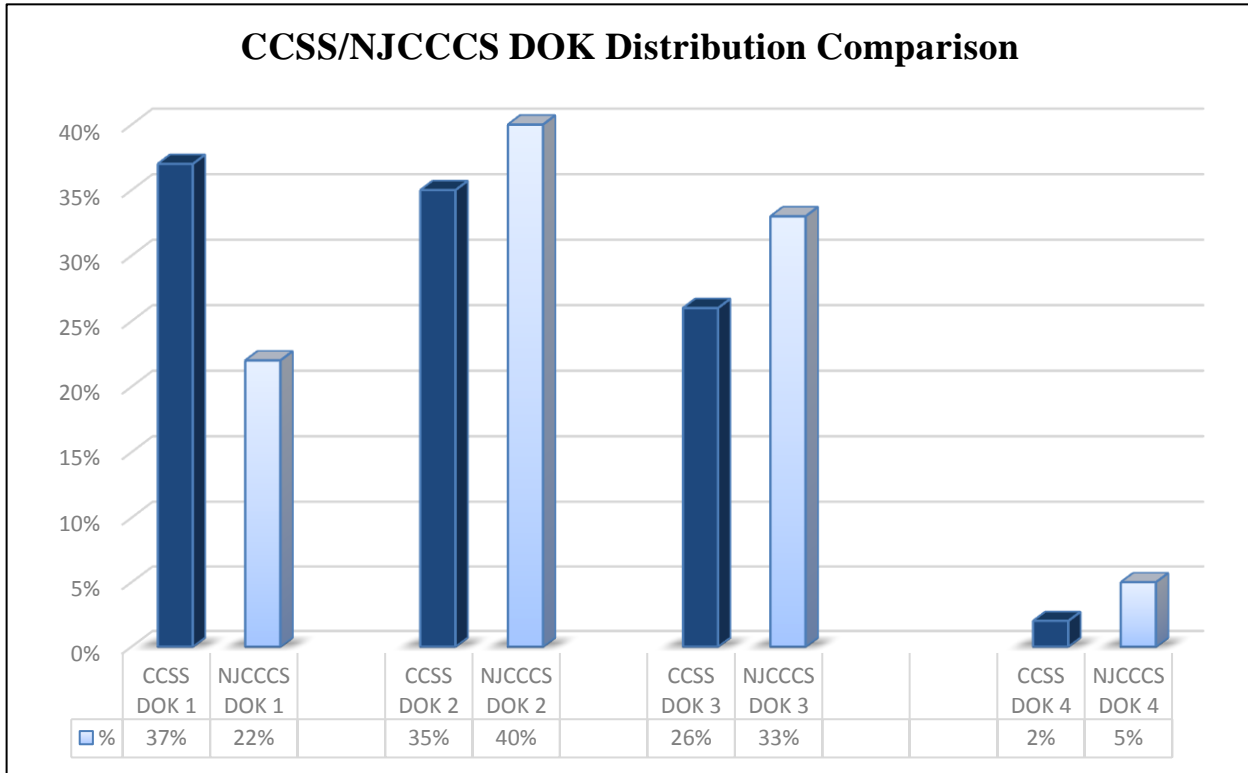


Figure 3. Grades 9-12 ELA CCSS/NJCCCS DOK distribution comparison.

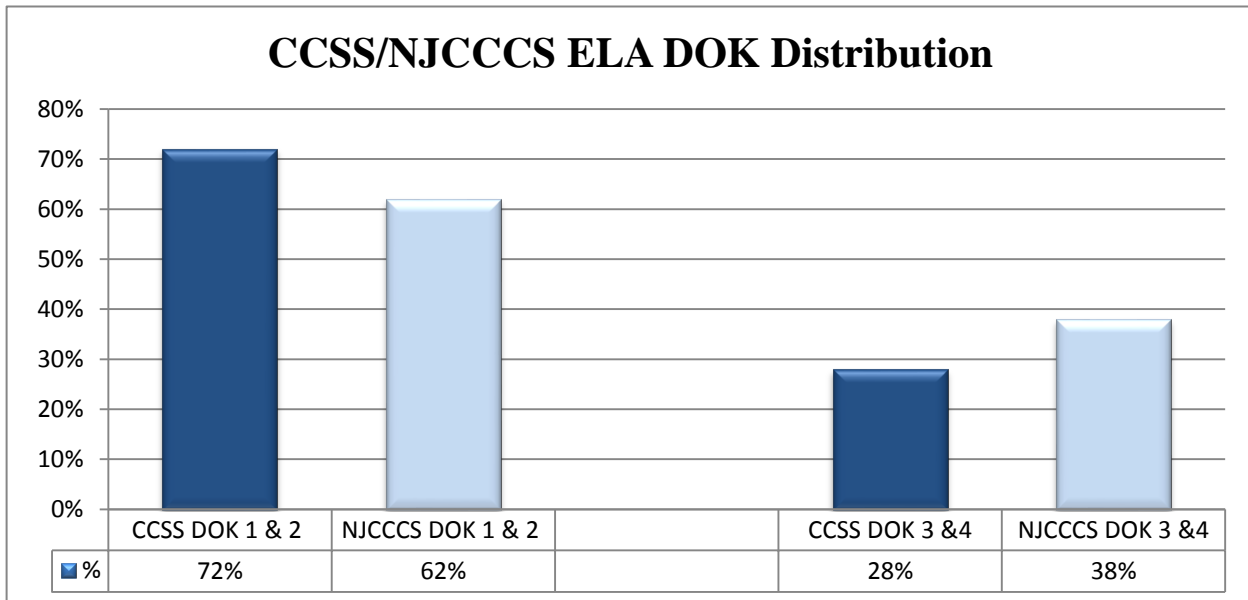


Figure 4. Comparison of cognitive complexity between the grades 9-12 Math CCSS and grades 9-12 Math NJCCCS.

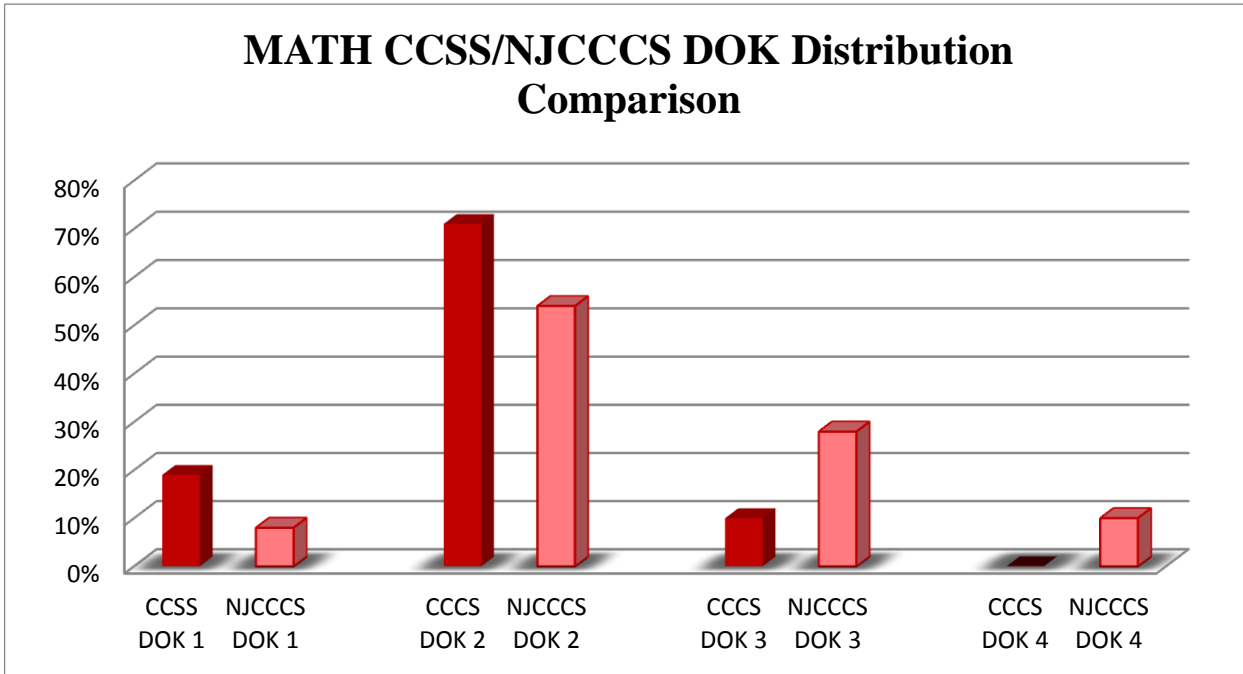
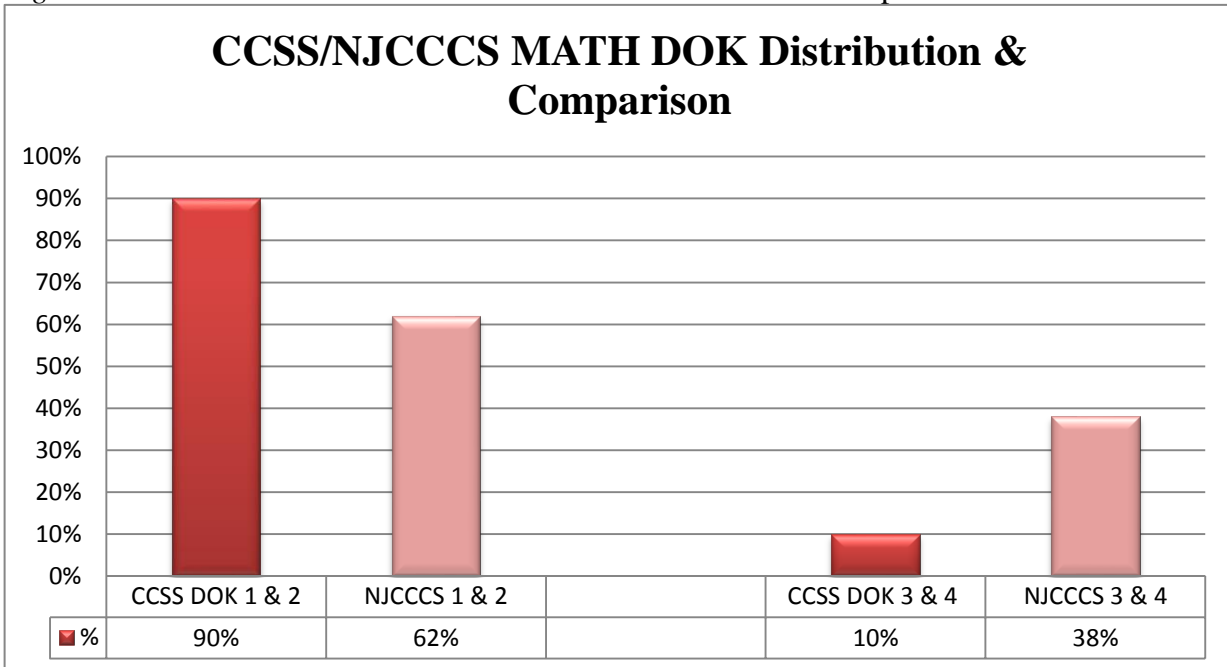


Figure 5. Grades 9-12 Math CCSS/NJCCCS DOK distribution comparison.



Common Core Less Complex

The results suggest that the previous versions of the NJ high school ELA and math standards included more complex, higher-order thinking and provided more opportunities to practice the types of thinking valued in the mainstream education reform literature as necessary to compete in the global economy. Although some have noted the CCSS as being more difficult than some previous states' standards, difficulty is not a proxy for creativity and strategic thinking (e.g. Porter, McMaken, & Hwang, 2011). Convoluted prompts and questions and unclear portions of some standards do nothing to foster creative or strategic thinking (Wiggins, 2014).

The CCSS are not superior to the previous version of the NJ high school standards in ELA and math in the areas of creative and strategic thinking. If a goal of the high school CCSS is to provide more opportunities for complex thinking then that goal has not been achieved compared to what existed previously in NJ. Our results suggest that a majority of the high school CCSS include procedural and declarative knowledge as opposed to necessary strategic and creative thinking. The intended curriculum of the CCSS requires students to more often engage in convergent thinking and use facts to imitate processes in order to find one correct answer than they were with the previous high school ELA and math standards in NJ.

Recommendations for School Leaders

Regardless of whether they support or reject the CCSS, school leaders in New Jersey and other states should work with their professional staff to review their schools' and districts' curriculum and augment it to include opportunities for creative and strategic thinking beyond those required by the CCSS in ELA

and math if their curricula are directly aligned to the CCSS. School leaders, in collaboration with their professional staff, might endeavor to revise and customize existing objectives and activities in their state mandated ELA and math curricula to generate more creative and strategic thinking opportunities for students.

The results of our study suggest a preponderance of procedural and declarative knowledge and thinking in the ELA and math CCSS. The danger we fear is that the CCSS ELA and math standards in high school might instill functional fixedness in student thinking and hinder their ability to enter the post-secondary global economic environment (Runco & Chand, 1995).

One way to inject creativity and strategic thinking into curricula is to add activities that focus on socially conscious problem solving. Problem-based activities derived from issues found in American society, as well as international issues, have a long track record of providing students opportunities to engage in creative and strategic thinking, while also producing superior results on traditional measures of academic achievement (e.g., Aikin, 1942; Boyer, 1987; Dewey, 1938; Isaac, 1992). Although such activities can be decidedly unstandardized, allowing for various processes and answers, state mandated curriculum standards can be infused into them without violating compliance laws.

Another way to inject more higher-order thinking in the CCSS would be to put the previous NJ ELA and math standards categorized as Level 3 or 4 back into the New Jersey school curricula. School leaders in NJ could add at least 10% more higher-order thinking in ELA and 20% in math just by

reusing a curricular “wheel” that already exists instead of trying to reinvent one. A drawback to this approach, though, would be the challenge of finding room in already over-stuffed ELA and math curricula. Perhaps school leaders can work with their professional staff to de-emphasize some of the procedural and declarative knowledge in the CCSS and replace it with some higher-order NJ standards or other quality standards and problem-based activities.

We are sensitive to the fact that NJ education officials instituted six new high school exit exams in grades 9-11 in ELA and math and to the fact that those exams are aligned to the CCSS standards. Given the high stakes for students (not graduating from high school) and for teachers and school administrators (lower evaluation ratings if student standardized test scores are low) attached to the high school exit exams, we understand the trepidation some superintendents and other district administrators might feel about de-emphasizing the CCSS.

We leave the moral and professional decision making about this issue up to them.

However, we do remind our colleagues that students do not have a voice at the policy making table, and thus their rights to a high quality, comprehensive education are protected only by educators who take their duty to provide that comprehensive education seriously. We see equipping students with the ability to think creatively and strategically as moral and professional duties. Following ineffective or untested education policy simply to not upset state education officials is not leadership in our opinion.

School leaders, education officials, and policy-makers in other states might also take notice of our results. They might choose to engage in a review of their previous state standards in ELA and math to determine if they contained more higher-order thinking compared to the CCSS. As we were somewhat surprised to learn from the results of this study in New Jersey, high school administrators should not rely on the claims of others regarding the ability of the CCSS to provide superior levels of higher-order thinking. We suggest they adopt the mantra “show us the data” when it comes to this claim.

Author Biographies

Dario Sforza is a high school principal at Henry P. Becton Regional High School in East Rutherford, NJ . He works with stakeholders to develop progressive and innovative education programs for students. Sforza's research interests include the influence of creativity and critical thinking on curriculum and instruction locally and nationally. E-mail: dbsforza@gmail.com

Christopher Tienken is an associate professor at Seton Hall University in South Orange, NJ. His books include *The School Reform Landscape: Fraud, Myth, and Lies* and *Education Policy Perils: Tackling the Tough Issues*. For additional information on Tienken, visit his website at <http://christienken.com/>. E-mail: Christopher.tienken@shu.edu

Eunyoung Kim is an associate professor at Seton Hall University, in the department of education leadership, management, and policy in South Orange, NJ. E-mail: eunyoung.kim@shu.edu

Authors' note:

Portions of this article are adapted from Sforza, D., Tienken, C.H., & Kim, E. (2015). *Common Core and Creativity: A Webb's Depth of Knowledge Analysis*. Paper presented at the National Council Professors of Educational Administration, August 6, 2015, Washington, DC.

References

- Aikin, W. M. (1942). *The story of the eight-year study with conclusions and recommendations*. New York, NY: Harper & Brothers.
- American Society for Training and Development (ASTD). (2009). *Bridging the skills gap*. Retrieved from <http://www.astd.org/%20About/~media/Files/About%20ASTD/Public%20Policy/%20BridgingtheSkillsGap2010.pdf>
- Asia Society. (2010). *PISA chief explains the data*. Retrieved from <http://asiasociety.org/education/learning-world/pisa-chief-explains-data>
- Boyer, E.L. (1987). *College: The Undergraduate Experience in America*. New York: Harper & Row.
- Cisco Systems Inc., Intel Corporation, Microsoft Corporation, & University of Melbourne. (2010). *Assessment and teaching of 21st century skills (ATC21S)*. Springer. DOI 10.1007/978-94-007-2324-5
- Council of Chief State School Officers [CCSSO]. (2010). *Press release: National Governors Association and state education chiefs launch common state academic standards*. Retrieved from http://www.ccsso.org/News_and_Events/Press_Releases/NATIONAL_GOVERNORS_ASSOCIATION_AND_STATE_EDUCATION_CHIEFS_LAUNCH_COMMON_STATE_ACADEMIC_STANDARDS_.html#sthash.IwRxgnGo.dpuf
- Dewey, J. (1938). *Experience and education*. New York, NY: Macmillan.
- Florida State University. (2013). *CPALMS*. Retrieved from <http://www.cpalms.org/Downloads.aspx>
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288.
- IBM Corporation. (2012). *Global chief executive officer study*. Retrieved from <http://www-935.ibm.com/services/us/en/c-suite/ceostudy2012/>
- Isaac, K. (1992). *Civics for democracy. A journey for teachers and students*. Washington, DC: Essential Books.
- Maxwell, J. A. (2005). *Qualitative Research Design: An Interactive Approach* (2nd Ed.). Thousand Oaks, CA: Sage.
- Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research, 1*(2). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385>

- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks, CA: Sage.
- National Education Association. (2012). *Preparing 21st century students for a global society: An educator's guide to "the four Cs."* Washington, DC: NEA. Retrieved from <http://www.nea.org/assets/docs/A-Guide-to-Four-Cs.pdf>
- National Governors Association and Council of Chief State School Officers. (2015). *About the standards*. Washington, DC: NGA and CCSSO. Retrieved from: <http://www.corestandards.org/about-the-standards/>
- New Jersey Department of Education. (2012a). *Education transformation task force final report* Retrieved from <http://www.state.nj.us/education/reform/ETTFFinalReport.pdf>
- New Jersey Department of Education. (2012b). *New Jersey application for funding under Race To The Top phase 3*. Retrieved from <http://www.state.nj.us/education/rttt3/about/Application.pdf>
- New Jersey Department of Education. (2010). *Common core state standards: Preparing students for college and careers*. <http://www.state.nj.us/education/sca/>
- New Jersey Department of Education. (2009). *New Jersey core curriculum content standards in mathematics*. Retrieved from <http://www.state.nj.us/education/cccs/2009.htm>
- New Jersey Department of Education. (2009). *New Jersey core curriculum content standards in language arts literacy*. Retrieved from <http://www.state.nj.us/education/cccs/2009.htm>
- Niebling, B. C. (2012). Using Webb's Alignment Model to Measure Intended-Enacted Curriculum Alignment: A Brief Treatment. *Midwest Instructional Leadership Council*, (1) 1-17.
- Organisation of Economic Co-operation and Development [OECD]. (2013). *PISA 2012 results. What students know and can do: Student performance in reading, mathematics and science* (Vol. I). PISA, OECD Publishing. Retrieved from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-I.pdf>
- Pink, D. (2006). *A whole new mind. Why right brainers will rule the future*. New York: Riverhead Books.
- Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common Core Standards: The new U.S. intended curriculum. *Educational Researcher*, 40(3), 103–116.

- Robinson, K. (2011). *Out of our minds: Learning to be creative*. North Mankato, MN: Capstone Publishers.
- Runco, M. A., & Chand, I. (1995). Cognition and creativity. *Educational Psychology Review*, 7(3), 243-267.
- Sato, E., Lagunoff, R., & Worth, P. (2011). SMARTER Balanced Assessment Consortium Common Core State Standards analysis: Eligible content for the summative assessment. (Final Report). WestEd. Retrieved from <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/Smarter-Balanced-CCSS-Eligible-Content-Final-Report.pdf>
- Sternberg, R.J. (2003). Creative thinking in the classroom. *Scandinavian Journal of Educational Research*, 47, 325-338.
- Sternberg, R.J. (1999). *The handbook of creativity*. New York, NY: Cambridge University Press.
- Sousa, D. (2011). *How the brain learns*. (4th Ed.). Thousand Oaks, CA: Corwin.
- Ward T. B., Smith, S. M., & Finke, R. A. (1999). Creative cognition. In R. J. Sternberg (Ed.), *Handbook of Creativity*. (p. 189-212). New York, NY: Cambridge University Press.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6). Washington, DC: Council of Chief State School Officers.
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20, 7–25.
- Webb, N.L., Alt, M., Ely, R., & Versperman, B. (2005). *Web alignment tool training manual*. Wisconsin Center for Education Research. Retrieved from <http://wat.wceruw.org/index.aspx>
- Wiggins, G. (2014, Nov. 24). Failure: The 8th grade NYS Common Core math test. *Granted and...* Retrieved from <https://grantwiggins.wordpress.com/2014/11/24/failure-the-8th-grade-nys-common-core-math-test/>
- Wyse, A.E. and Viger, S.G. (2011). How item writers understand depth of knowledge. *Educational Assessment*, 16(4), 185-206.

Problems with Percentiles: Student Growth Scores in New York's Teacher Evaluation System

Drew Patrick, MEd
Assistant Superintendent, Curriculum & Instruction
Bedford Central School District
Mount Kisco, NY

Abstract

New York State has used the Growth Model for Educator Evaluation ratings since the 2011-2012 school year. Since that time, student growth percentiles have been used as the basis for teacher and principal ratings. While a great deal has been written about the use of student test scores to measure educator effectiveness, less attention has been paid to how value added models have played out in schools, school districts, and states since their widespread adoption associated with Race to the Top. This study employs univariate and multivariate statistical procedures to examine model results at the student level in one district, and across districts, and identifies problems associated with the model. Policy implications and recommendations are discussed.

Key Words

Growth models, value-added modeling, student growth percentiles, educator evaluation ratings, evaluation policy

Introduction

The use of test scores to evaluate teachers and principals has increased tremendously during the Race to the Top era (Baker, Oluwole, & Green, 2013). Generally referred to as *value-added modeling* (VAM), the technique relies on complex statistical models to predict future student test scores based on prior scores and various other demographic and school-related factors. Teachers and principals of students who beat their predictions are considered to have “added value”, or contributed substantially to student learning, relative to teachers whose students miss their predicted scores.

In many systems, teachers are then assigned to a rating category (i.e., “ineffective” or “highly effective”). Not surprisingly, lines have been drawn and a debate is underway between proponents of VAM and those who argue against its utility for gauging educator effectiveness (Goldhaber, 2015; Holloway-Libell & Amrein-Beardsley, 2015). While it is important to understand the arguments for and against, briefly outlined in the next section, there remains a substantial dearth of information about the performance of state- or district-specific VAMs over time. There is a clear and present need for a determination as to whether or not these models are capable of producing the results intended by the policy makers who adopted them.

The purpose of this exploratory study was to gauge the extent to which the New York Growth Model for Educator Evaluation provides meaningful student level growth data to inform educator practice. Furthermore, since these student-level data are aggregated at the teacher and school level to make effectiveness

determinations, the study also attempted to identify potential problems with their use for this purpose. Overall, the analysis raises concerns about the meaning of student growth percentiles (SGPs), along with questions about year-to-year stability and performance-level bias, such that using these measures to assign a teacher or principal growth score deserves closer examination, and supports the call for a broader and deeper study.

Context of the Problem

The use of student growth for accountability purposes first entered the education policy arena in the context of the school and district-level performance, as opposed to teacher performance (Betebenner, 2011). In 2005, the USDOE gave states opportunities to begin measuring and reporting student growth-toward-proficiency as a strategy to meet AYP (adequate yearly progress) as part of the Growth Model Pilot Program (Hoffer et al., 2011). As the accountability gears kept grinding, the methodologies associated with this (i.e., VAM) were turned toward the classroom (Betebenner, 2011). Since this time, economists and educational researchers have been debating over the use of these models for teacher-level accountability.

Research that favors the use of VAM to make judgments about educator effectiveness generally argue that the potential for good outweighs the negatives, and is constructed around the following ideas (Chetty, Friedman, & Rockoff, 2014a; Chetty, Friedman, & Rockoff, 2014b; Hanushek & Rivkin, 2010; Holloway-Libell & Amrein-Beardsley, 2015; Rockoff & Speroni, 2010; Tyler, Taylor, Kane, & Wooten, 2010):

- Teachers' effectiveness varies as measured by value-added
- Teacher value-added is an educationally and economically meaningful measure
- Teacher effects can be discerned from VAMs in an unbiased manner
- The models and the results they produce can adequately control for non-classroom or teacher effects
- Using teacher value-added improves achievement more than not using it

On the other side, those opposed to using VAM for educator effectiveness decisions argue there is a high risk of unintended negative consequences, including false positives and negatives, narrowing of the curriculum, class roster and student test manipulation.

These arguments are constructed around the following ideas (Baker et al., 2013; E. L. Baker et al., 2010; Ballou & Springer, 2015; Braun, 2015; Darling-Hammond, 2015; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rothstein, 2010; Strong, Gargani, & Hacifazlioglu, 2011):

- Value-added estimates are biased, and are invalid- they do not measure what they purport to measure
- Value-added estimates have unacceptably high error to be used in making high stakes decisions about teachers
- Value-added estimates are unstable over time, limiting their reliability and therefore usefulness
- Value-added estimates are too complex to be understood in meaningful ways by those for whom they are intended (i.e., teachers and school leaders)

- There are underlying biases in the student-level estimation of growth that create problems for aggregating to teacher-level effects
- Even if you identify bad teachers with VAM, the current workforce does not support the idea that low performers can regularly be replaced by higher performers.

Regardless of viewpoint, models that rely on student test scores to make educator effectiveness determinations are in use across the country (Baker et al., 2013). One of the gaps in the literature is a lack of research focused on the value-added models that are currently in place in states and districts. Specifically, it is important to examine how these models have performed over time with respect to their ability to predict student performance in a meaningful way, and therefore contribute toward an understanding of teacher influence on that performance.

New York's Growth Model for Educator Evaluation

Student growth on state tests as determined by New York's Growth Model for Educator Evaluation, developed by American Institutes for Research (AIR), has been used over the past four years to generate one of the multiple measures used in deriving an overall teacher score and rating (American Institutes for Research, 2014). This model results in state-provided growth scores (SPGS) for teachers of ELA and mathematics in grades 4-8. This score represents 20% of an overall composite score that also includes locally-determined measures of student growth or achievement (20%) and other measures based on classroom observation (60%).

While only impacting some 15% of classroom teachers, the use of SGPs has implications beyond just 4-8 ELA and mathematics teachers because some school districts elected for the simplicity of applying state-provided scores to all teachers, something allowable under New York's evaluation law (New York State Education Department, 2012).

The model uses grade-specific multiple regression equations to generate predictions for current year test scores, taking into account up to three years of prior tests scores, along with various demographic and other factors. Recently, the reliability of these predictions was called into question in the form of a legal challenge.

In August 2015, oral arguments were heard in a case brought against former NYS Education Commissioner John King by a fourth grade teacher from Long Island. The teacher sought a remedy to the arbitrary and capricious nature of her SPGS, which dropped from 14/20 in 2012-13 to 1/20 in 2013-14. Part of the case centered on the influence a single student's test score had on the teacher's score.

One student received a perfect score on the state test prior to entering the plaintiff's classroom, and the growth model predicted another perfect score in 4th grade. The student ended up getting a total of two questions wrong, which lowered the teacher's score into the ineffective range. The student's score was higher than 99% of all 4th graders state-wide, but the teacher was rated in the bottom 6% in part due to this "failure" (B. Lederman, personal communication, August 12, 2015).

While a decision is pending in this case, the New York State Education Department

made a significant policy change in September 2015, creating a process by which, under certain conditions, teachers and principals can appeal their SPGS and have it thrown out (New York State Education Department, 2015b). On its face, this change intimates concern by the Education Department about the ability of the model to produce meaningful results.

The Study

Description

This study aimed to explore the question, *To what extent does the New York Growth Model for Educator Evaluation provide meaningful student level growth data to inform educator practice and gauge effectiveness?* To answer this question, the study relied on an analysis of a region-level (16 school districts), and district-level (1 district) dataset based on the 2015 New York State English language arts (ELA) and mathematics tests.

Each dataset included de-identified student-level data for: test name, current and prior-year (2014) scale score, current year predicted scale score, current and prior-year (2014) performance level, and current and prior-year (2014) percentile rank. Calculated variables included change in performance level (2015-2014) and change in percentile rank (2015-2014), categorized into deciles (0-10 = 10, 11-20 = 20, etc.). The district-level data set also included student growth percentiles, where available, back to 2011-12. It is important to note that growth percentiles are first generated for students in grade four, as that is the first possible year in which students have a prior-year test score, the most important independent (predictor) variable in the growth model.

Accordingly, grade eight is the last year in which student growth percentiles are

calculated. The question was answered using descriptive statistics (frequencies, means) and one-way analysis of variance.

Limitations

The study had several limitations. First, the region-level dataset did not contain SGPs from years prior to 2015, preventing a broader examination of the year-to-year stability of SGPs in the larger sample.

Second, these results, while based on a large N-size of over 4,300 students, may not be generalizable to populations in other parts of the state, in part because the percentage of students on free or reduced lunch is much lower than the state-wide average.

Third, more complex statistical analyses need to be done to further explore the correlations between SGPs year-to-year.

Findings

The analysis began with an examination of ELA results. Descriptive statistics related to the 16-district dataset are presented in Table 1.

The proportion of English language learners (ELL) and students with disabilities is similar to the state-wide average (3% and 8%, respectively; New York State Education Department, 2015a), but the free or reduced lunch percent-age falls 10% shy of the state average. The mean SGP ranges from 47.0 (grade 4) to 51.2 (grade 5), with an overall value of 49.6, very close to the expected mean of the state-wide distribution.

This suggests that sample population, overall, exhibits characteristic student growth behavior, hovering at the mean.

Table 1

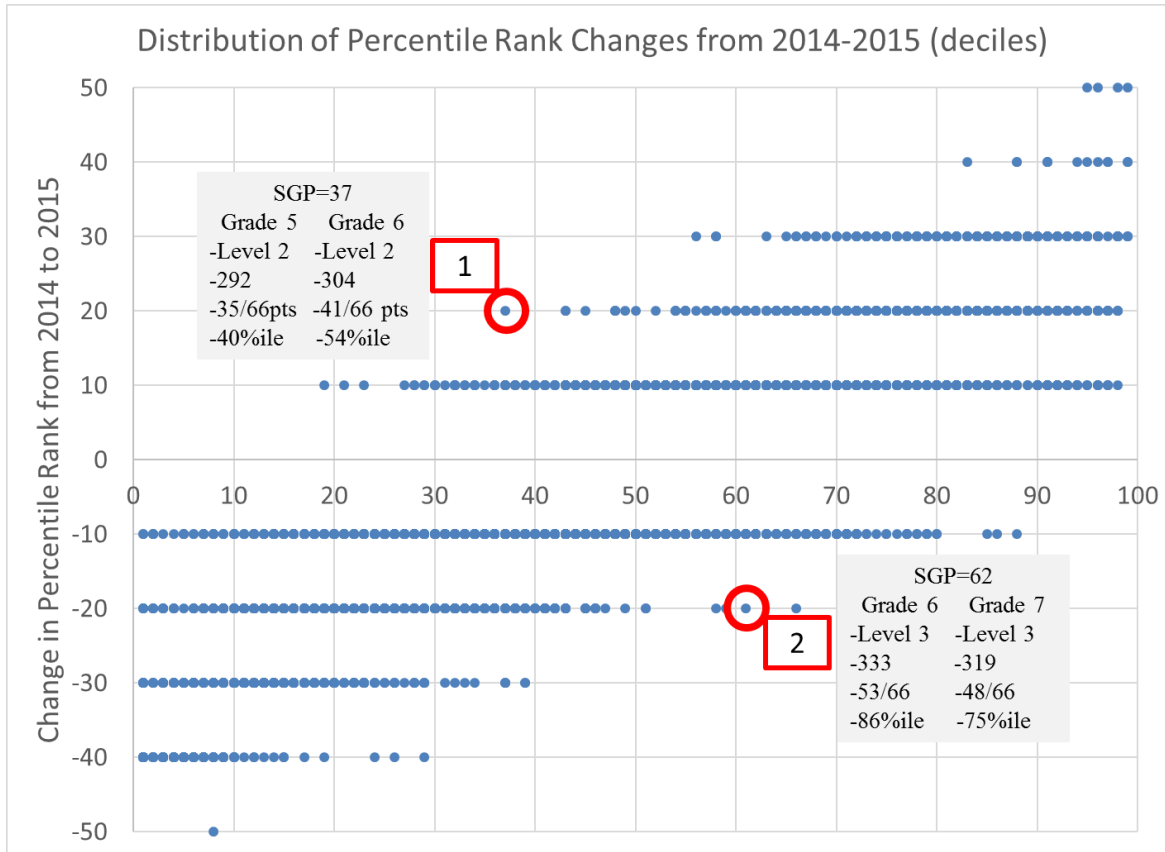
Descriptive Statistics for 2015 Region-Level Student ELA Data

Grade	N	Free or Reduced Lunch	Students with Disabilities	English Language Learners	Mean ELA Student Growth Percentile
4	889	145	75	24	47.0
Female	422	64	24	11	47.6
Male	467	81	51	13	46.4
5	905	138	92	25	51.2
Female	428	66	34	6	53.2
Male	477	72	58	19	49.5
6	898	122	86	23	49.5
Female	431	54	38	11	51.9
Male	467	68	48	12	47.2
7	863	116	88	30	50.9
Female	417	46	31	16	53.4
Male	446	70	57	14	48.5
8	812	124	81	21	49.5
Female	425	71	38	13	51.1
Male	387	53	43	8	47.6
Total	4367	645 (15%)	422 (10%)	123 (3%)	49.6

Figure 1 (below) shows more detail by illustrating the distribution of the ELA SGPs grouped according to changes in overall achievement percentile rank for students in the dataset. The percentile rank for a student represents the overall percentage of students state-wide which that student outperformed on the same test in the same year. The change in this achievement measure was calculated by

subtracting the 2014 rank from the 2015 rank, and returning a value. These values were then clustered into ranges in order to make the graph easier to interpret. Thus, a -10 value means the student's achievement percentile rank was between 0-10 points lower in 2015 than in 2014. While every decile and SGP is displayed on the graph, more than one student can be represented by each point plotted.

Figure 1. Distribution of percentile rank changes, in deciles, by SGP.



Bivariate analysis between SGP and change in percentile rank showed a strong, significant positive correlation ($r = .851, p < .01$), indicating that as the SGP increases, so does the change in percentile rank. However, figure 1 tells a more complicated story about individual students, as it uncovers a more nuanced relationship between calculated SGP and achievement (percentile rank). For example, the student labelled with the number 1 in the graph performed at the same level (2) in consecutive years, but exhibited an improvement in percentile rank from 40 to 54,

and answered more questions correctly (41 versus 35).

However, despite this student's improvement in standing relative to his peers, his SGP is only 37. Incidentally, this is a dangerously low contribution toward a teacher's mean growth percentile used to determine effectiveness. By contrast, the student labelled with the number 2 has an SGP of 62, yet shows a downward achievement trend as measured by raw scores, scale score and percentile rank. However, because the

MGP is above 50, the measure suggests this student exhibited substantial growth (attributed to the teacher).

Table 2 also illustrates data from the 16-district data set. The four students represented in this table have consistently achieved at the highest performance level on the math tests (level 4). However, their most recent SGPs range from a low of 11 to a high of 95. The gray boxes highlight the “outlier” test scaled scores that drive the 2015 SGP. The first student out-performed his “typical” scoring pattern in 2014 (note 2012 reflects an older,

pre-Common Core test scale), resulting in higher-than-typical predicted score of 373.4. The student was unable to reach that prediction, and the low SGP reveals that fact. The second student, row 2, shows a high degree of consistency. This student’s scoring pattern falls comfortably along the line predicted by the growth model, and the student has an SGP right smack in the middle- 50. This student’s test taking pattern is as predicted. The third row shows a student who exceeded prediction slightly, and the fourth row is the converse of the first row—this student’s “good” year is in 2015.

Table 2

Examples of Prior-year Tests Influencing Predictions and Growth Scores

	2015 SGP	2015	2014	2013	2013	Predicted Score
Grade 6 Math	11	350	377	347	725	373.4
Grade 8 Math	50	357	349	345	726	357.2
Grade 6 Math	70	376	360	354	725	365
Grade 7 Math	95	374	341	353	742	344.1

Continued examination reveals a pattern that shows what it takes to get low, close-to-mean, or high SGPs. When a student substantially exceeds a predicted score in a single year, and then performs closer to the longer-term average in the subsequent year, his/her SGP reflects a big drop, resulting in a low SGP. This suggests the phenomenon of *regression to the mean* (Healy & Goldstein, 1978). In other words, repeated measures of SGP over time for an individual student with one year of an outlier score will experience an SGP closer to the mean of 50 over the course of multiple testing experiences. Meanwhile, there

is a “good year” to be this student’s teacher, and a “bad year” (like 2015).

While Table 2 focused on high-performing test takers, Table 3 illustrates two lower performing students. Both exhibit above-average SGPs for the 2015 school year. However, when you examine the test score history, the grade 8 student is persistently low-performing, while the grade 7 student appears to be on a downward trajectory. Common sense would suggest these students are not heading in the right direction, but their SGPs suggest they are.

Table 3

Examples of Strong-SGP Students with Low Performance Levels

	2015 SGP	2015	2014	2013	2013	Predicted Score
Grade 8 Math	73	263	261	231	658	250.5
		Level 1	Level 1	Level 1	Level 2	
Grade 7 Math	53	290	284	294	703	288.8
		Level 1	Level 2	Level 2	Level 3	

Mean SGP Variability

Moving to the district-level data set, descriptive statistics are again presented (Table 4, below). The population exactly matches state-wide averages for free or reduced lunch and special education, and exceeds the state average for ELLs. The overall mean SGP is 46.5, lower than in the larger dataset, and mean SGPs range from 30.5 in grade 8 to 59.3 in grade 5.

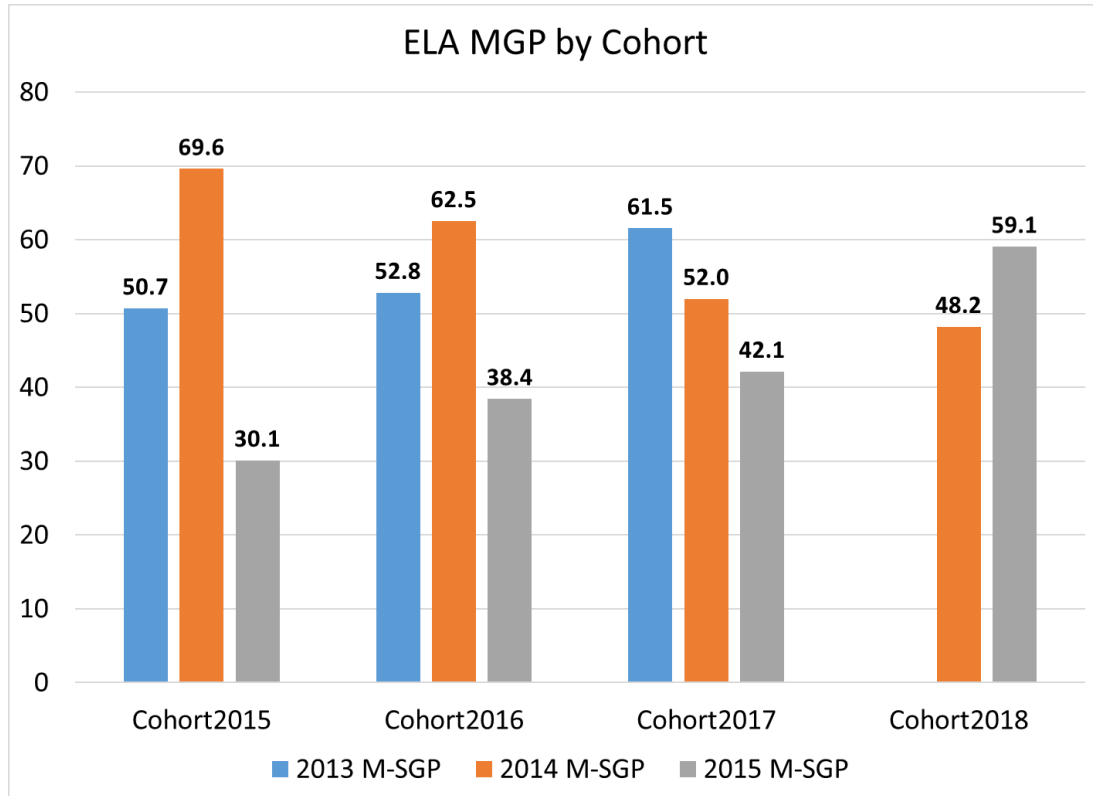
Overall, grades 6, 7 and 8 have decreasing SGPs. This is illustrated more clearly in Figure 2, below. Organized by cohort (i.e., Cohort 2015 are 9th graders in 2015), there are substantial fluctuations in the mean of the SGPs for this cohort over the three years represented. For example, cohort 2015 shows an SGP increase of nearly 20 percentile points, followed by a nearly 40 percentile point drop.

Table 4

Descriptive Statistics for 2015 District-Level Student ELA Data

Grade	N	Free or Reduced Lunch	English Language Learners	Students with Disabilities	Mean Student Growth Percentile
4	288	79	27	28	58.2
Female	136	40	12	5	61.8
Male	152	39	15	23	55.1
5	312	80	27	35	59.3
Female	145	35	8	11	61.2
Male	167	45	19	24	57.5
6	266	54	13	19	42.4
Female	138	27	7	6	43.3
Male	128	27	6	13	41.5
7	277	66	15	21	38.9
Female	126	33	7	5	41.0
Male	151	33	8	16	37.2
8	265	52	2	15	30.5
Female	134	24	1	3	32.3
Male	131	28	1	12	28.6
Total	1408	331 24%	84 6%	118 8%	46.5

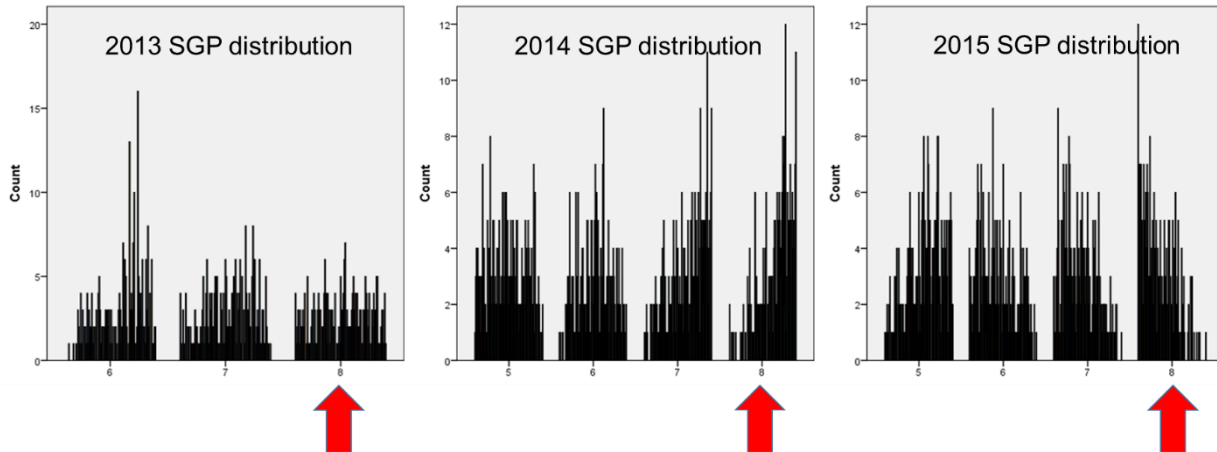
Figure 2. ELA Mean Growth Percentiles by Cohort, 2013-2015.



The graphs in Figure 3 illustrate in greater detail the distribution of growth scores for the same cohorts depicted in the previous figure. Cohort 2015 is identified by the red arrows, and the left-most graph shows the SGP distribution for 2013, the middle for 2014, and

right-most graph 2015. The bars represent the number of students at each SGP value received. While there is a roughly-normal distribution in 2013, this is extremely skewed toward higher SGPs in 2014, and shifts even more drastically in 2015 to mostly lower growth scores.

Figure 3. ELA student growth percentile frequency distributions by year and cohort.



SGPs by Performance Level

Finally, Table 5 (below) reports an analysis of the mean SGP for all students scoring at performance level 1, 2, 3 or 4 in 2015 (16 districts) for ELA and math. There is a nearly 30-point difference in SGP from level 1 performers to level 4 performers in both subjects. All things being equal, the growth model should produce a normal distribution of growth scores across similar student groups, but when results are translated into

performance level ranges, it appears that lower performers are systematically receiving lower growth scores than higher performers. Table 6 (below) reports the results of a one-way ANOVA with Bonferroni post-hoc analysis. We see that the large differences between these means is statistically significant ($p = .05$), meaning there is at least a 95% likelihood that these differences are due to something other than chance alone.

Table 5.

Mean SGP by Performance Level for 2015 ELA and Math

Performance Level	N	Mean ELA SGP	N	Mean Math SGP
1	605	36.1	520	32.6
2	1412	43.8	1001	43.1
3	1509	51.8	1378	49.4
4	841	65.2	1250	60.0
Total	4367	49.6	4149	49.0

Table 6.

SGP One-Way ANOVA results with Bonferroni Post-Hoc Analysis, 2015 ELA and Math Performance Levels

2015 ELA Performance Level		Mean Difference	Std. Error	Sig.	2015 Math Performance Level		Mean Difference	Std. Error	Sig.
1	2	-7.7320*	1.1946	.000	1	2	-10.5037*	1.3651	.000
	3	-15.6685*	1.1830	.000		3	-16.7921*	1.2997	.000
	4	-29.0767*	1.3106	.000		4	-27.4334*	1.3178	.000
2	3	-7.9365*	.9103	.000	2	3	-6.2883*	1.0488	.000
	4	-21.3447*	1.0708	.000		4	-16.9297*	1.0711	.000
3	4	-13.4082*	1.0579	.000	3	4	-10.6414*	.9864	.000

*The mean difference is significant at the 0.05 level.

Discussion

The main purpose of this study was to address the question, *To what extent does the New York Growth Model for Educator Evaluation provide meaningful student level growth data to inform educator practice and effectiveness?* Analysis of the two data sets, both containing student-level data, raises questions about the meaning of individual SGPs and their potential to influence MGPs (and therefore teacher growth scores) in a manner that can be discordant with evidence of achievement. The following observations based on the above analysis summarize these concerns:

- The relationship between SGP and achievement as measured by percentile rank exhibits a strong positive correlation, but large numbers of individuals exhibit information that can be viewed as contradictory to teachers trying to use this information to determine whether a student has had indeed made meaningful learning gains

over the course of the year (figure 1). As the Lederman case has demonstrated, even one missed target (reasonable or not) can negatively influence a teacher rating.

- Year-to-year fluctuations with individual SGPs exhibit regression to the mean over time. This effect is especially evident when students substantially exceed or fail to meet predictions in a given year (tables 2 & 3). Students who far exceed a prediction receive a high SGP in that year, but are likely destined for an equally low SGP in the subsequent year. Non-random assignment of students to teachers can therefore pose a potential threat to the SPGS of teachers who get a disproportionate number of students receiving high SGPs in a given year.
- Regression to the mean also has the potential to occur for entire cohorts of

students in a school (figures 2 & 3). In the single district dataset, large swings in mean SGP resulted in high, then low, teacher evaluation scores in ELA. When a large group of students beats their respective predicted scores, low teacher and principal ratings and scores are likely to follow in the subsequent year.

- Statistically significant differences exist between student growth scores at each of the four student performance levels reported (tables 5 & 6). These differences are substantial, and would cause any reasonable person to recognize the disincentive this could create against wanting to teach a class of low performers.

Policy Implications

New York State's teacher and principal evaluation law, as written, explicitly and implicitly articulates a theory of action that, arguably, communicates the following set of beliefs:

1. Changes in student achievement from one year to the next are an indication of teacher and principal effectiveness.
2. Teacher and principal effectiveness can be differentiated through an analysis of observed student growth on state assessments.
3. Observed differences on these measures allows for identification of bad teachers and principals.
4. Bad teachers and principals will be motivated by their ratings to improve, or to get out of the profession.

5. Better teaching and leadership, or new and better teachers and principals, resulting from this policy will improve student achievement.

The problems outlined in this study paint a confused picture of SGPs as derived from New York's Growth Model for Educator Evaluation. The instability of SGPs experienced by both individual and cohorts of students, coupled with large differences by performance level, raise serious doubts about the ability of this particular model to aid in accomplishing any of the steps in the theory chain above. This study provides enough evidence to warrant a fuller exploration of the model to include an analysis of state-wide SGP trends and patterns, and their implications for the stability of corresponding state-provided growth scores (SPGS).

In the meantime, the State Education Department should consider a moratorium on the use of this model until such a time as a more complete analysis can be done, inclusive of multiple years of SGP data from all districts in the state.

It is particularly important that this occur prior to widespread implementation of the most recent educator evaluation law, which promises to increase the influence of this portion of the evaluation system from 20% to nearly 50% of the overall score. Furthermore, serious effort should be made toward helping teachers and principals make meaning of the confusing, often contradictory measures of student learning based on the state testing program, including SGPs, percentile rankings, scale scores and performance levels. Until this happens, the link between teacher practice and

student achievement on state tests will remain obscured by the confusing output of the growth model.

Finally, it would be prudent for all states and districts using VAMs around the

country to carefully examine their use in light of these results. Education leaders and policy makers should establish a mechanism to gauge the degree to which their respective VAMs are meeting the intended policy objectives through empirical studies.

Author Biography

Drew Patrick is a doctoral candidate in the educational program at Manhattanville College. He serves as assistant superintendent for curriculum and instruction in the Bedford Central School District, Westchester County, NY. E-mail: apatrick1258@bcsdny.org

References

- Baker, B. D., Oluwole, J., & Green, P. C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the race-to-the-top era. *Education Evaluation and Policy Analysis Archives*, 21, 1-71.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers. EPI briefing paper# 278. *Economic Policy Institute*,
- Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, 44(2), 77-86. doi:10.3102/0013189X15574904
- Betebenner, D. W. (2011). A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories. *The National Center for the Improvement of Educational Assessment*. Retrieved from [Http://Www.Gadoe.Org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Sgp_technical_overview.Pdf](http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Sgp_technical_overview.Pdf), 439-450.
- Braun, H. (2015). The value in value added depends on the ecology. *Educational Researcher*, 44(2), 127-131. doi:10.3102/0013189X15576341
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593-2632.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633-79.
- Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, 44(2), 132-137. doi:10.3102/0013189X15575346
- Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher*, 44(2), 87-95. doi:10.3102/0013189X15574905
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, 100(2), 267-271.
- Healy, M., & Goldstein, H. (1978). Regression to the mean. *Annals of Human Biology*, 5(3), 277-280.

- Hoffer, T. B., Hedberg, E. C., Brown, K. L., Halverson, M. L., Reid-Brossard, P., Ho, A. D., & Furgol, K. (2011). Final report on the evaluation of the growth model pilot project. *US Department of Education*. Retrieved from <https://www2.ed.gov/rschstat/eval/disadv/growth-model-pilot/gmpp-final.pdf>
- Holloway-Libell, J., & Amrein-Beardsley, A. (2015). “Truths” devoid of empirical proof: Underlying assumptions surrounding value-added models in teacher evaluation. *Teachers College Record*, June 29
- McCaffrey, D. F., Lockwood, J., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability. monograph*. ERIC.
- New York State Education Department (2012). Guidance on New York State’s annual professional performance review for teachers and principals to implement Education Law §3012-c and the Commissioner’s regulations. Retrieved from <http://www.engageny.org/resource/guidance-on-new-york-s-annual-professional-performance-review-law-and-regulations>
- New York State Education Department (2015a). New York State Public School Enrollment (2014-15). Retrieved from <http://data.nysed.gov/enrollment.php?state=yes&year=2015&grades%5B%5D=03&grades%5B%5D=04&grades%5B%5D=05&grades%5B%5D=06&grades%5B%5D=07&grades%5B%5D=08>
- New York State Education Department (2015b). Amendment of Subpart 30-2 and Addition of a New Subpart 30-3 to the Rules of the Board of Regents and Section 100.2(o) of the Commissioner’s Regulations, Relating to Annual Professional Performance Reviews of Classroom Teachers and Building Principals to Implement Subparts D and E of Part EE of Chapter 56 of the Laws of 2015. Retrieved from <https://www.regents.nysed.gov/common/regents/files/meetings/Sep%202015/915p12hea1revised.pdf>
- Rockoff, J. E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *The American Economic Review*, 100(2), 261-266.
doi:<http://dx.doi.org.librda.mville.edu:2048/10.1257/aer.100.2.261>
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1), 175-214.
doi:10.1162/qjec.2010.125.1.175
- Strong, M., Gargani, J., & Hacifazlıoğlu, Ö. (2011). Do we know a successful teacher when we see one? experiments in the identification of effective teachers. *Journal of Teacher Education*, 64(4), 367-382. doi: 10.1177/0022487110390221

Tyler, J. H., Taylor, E. S., Kane, T. J., & Wooten, A. L. (2010). Using student performance data to identify effective classroom practices. *The American Economic Review*, *100*(2), 256-260.
doi:<http://dx.doi.org.librda.mville.edu:2048/10.1257/aer.100.2.256>

Doctoral Research in Educational Leadership: Expectations for Those Thinking About An Advanced Degree

David J. Parks, PhD
Professor Emeritus
Department of Educational Leadership, Counseling, and Research
Virginia Tech
Blacksburg, VA

Abstract

The tallest hurdle in completing a doctoral degree is the dissertation, which continues to be the primary capstone experience for the degree. Dissertation research is a mystery to many considering an advanced degree and can be intimidating to those who are unfamiliar with the nature of universities and doctoral research. In this report, the author removes some of the mystery by reviewing criteria applied by faculty in major universities and reporting the results of a questionnaire administered to faculty in Virginia. Both process and product criteria applied to doctoral research by the respondents in the study are reported.

Key Words

Dissertations, Education, Leadership

Graduate-student research comprises a large portion of the research completed in educational leadership. Most of this research is done as dissertations. In 2011-2012, 3857 students received either an EdD or PhD degree in some area of PK-12 educational leadership (United States Department of Education, National Center for Education Statistics, 2013). About two-thirds (63.4%) of the recipients were women; about a third (36.6%) were men. In Virginia, 565 doctoral students completed degrees in 2012-2013 (State Council of Higher Education for Virginia, n. d.). Nearly 70% (69.9%) were women; a little less than 30% (29.6%) were men (The gender of three graduates was not recorded.).

Most education degrees culminate with a dissertation, which is the largest stumbling block to completion of the degree. Comparatively, passing courses is easy. The dissertation, however, is a major piece of semi-independent research requiring persistence, knowledge of the subject, skill in planning and conducting research, and finesse in interpersonal relations.

Upon completion of the research, a report is prepared as a dissertation and is reviewed by a committee comprised of university faculty and, sometimes, practitioners, all of whom have been through the process for their own dissertations or through the review of the dissertations of others. What do they look for when they review these dissertations? What criteria do they apply? What standards guide their evaluations?

In this paper I review standards for measuring the quality of doctoral research by some major universities and report criteria that faculty from across Virginia believe to be

“essential (critical or indispensable)” in assessing the quality of that research. These criteria may be useful to faculty in educational leadership programs and to school practitioners as they attempt to assess, interpret, and apply research in their teaching and administrative roles. The criteria may be of interest, as well, to those who plan on pursuing the doctorate in educational leadership.

A Review of Standards for Dissertation Research

Standards for dissertation research vary by institution, faculty chair, and committee composition. Graduate schools across universities promulgate criteria for evaluating dissertations. Faculty chairs have their own views on what comprises an acceptable dissertation. And, committee members hold their own standards, which may differ from those of the faculty chair. In the end, the quality of dissertation research is assessed by the votes of the committee members and chair.

Passing or failing is largely a political decision. As is well known, those decisions are overwhelmingly positive (de-Miguel, 2010). There are few failures at the defense stage of the dissertation, regardless of the quality of the work. Despite this fact, there are standards that are promulgated by universities to maintain an acceptable level of dissertation quality and to guide dissertation advisors and committees. As with any policy or regulation, effectiveness of standards is determined by application and enforcement at the point where action is taken.

The dissertation standards of five top-rated (U. S. News and World Reports, 2014) programs in education policy were reviewed. These programs are at Stanford, Harvard, the University of Wisconsin--Madison, Vanderbilt

(Peabody), and Teachers College, Columbia. Each has process and product standards that are applied to maintain the quality of dissertation research.

Process standards involve how the dissertation is produced and evaluated. Product standards are applied to the quality of the content of the dissertation.

Standards from other notable universities in the United States and internationally were reviewed to supplement those of the United States. There are some similarities and many differences, with the international universities tending to be more conscious of averting the potential effects of friendships and political behavior of chairs and committee members.

Process standards for dissertation research

Process standards include selecting dissertation chairs and committee members, constituting dissertation committees, openly sharing the work of doctoral students with the full academic community, separating the dissertation advisor from the summative evaluation process, requiring external reviewers, and conducting multiple levels of evaluation.

Selecting dissertation chairs and committee members

Chairs and members are selected in various ways across universities. In all cases, chairs and members must meet the requirements of the governing bodies of the university.

At Stanford, the chair and committee represent the university, school, or department and verify that the standards of these bodies have been met (Stanford University, n. d.^a). Chairs of dissertation reading committees must be members of the Academic Council Professoriate, which consists of tenure-line and non-tenure-line teaching faculty at all ranks, non-tenure-line research faculty at all ranks, and senior fellows at policy centers and institutes. A co-advisor, who is a member of the Academic Council Professoriate, is required when an emeritus Academic Council member (after two years in emeritus status), a non-Academic Council member, or a former Academic Council member is appointed as chair. The co-advisor assures that someone directly connected to the department represents the student (Stanford University, n. d.^a).

At the University of Wisconsin—Madison (2013), dissertation review committees have two parts: a reading committee of three members and an oral examination committee of five members. At least three of the five members on the oral examination committee must be from the Department of Educational Leadership and Policy Analysis. At least four members of the committee must have Graduate Faculty status at the University of Wisconsin—Madison. At least one member of the oral examining committee must be from outside the student's department (University of Wisconsin—Madison, 2013).

Vanderbilt's three-year EdD program in the Department of Leadership, Policy, and Organizations has a capstone project rather than a dissertation (Vanderbilt University, 2015^b). The capstone project is designed with a partner organization that has an interest in making a change or implementing a program. Past partners are the Montgomery County, Maryland,

Public Schools and the Metropolitan Nashville Public Schools. The project is guided by the EdD faculty of the Peabody Department of Leadership, Policy, and Organizations. No standards for assigning faculty to supervise the students, other than the interests and competence of the faculty, were found on the Vanderbilt website.

Constituting doctoral committees

At Stanford University (n. d.^a), dissertation reading committees have a minimum of three and not more than five members, including the chair. One member must be from the student's department; the remaining members are appointed from the Academic Council Professoriate, from emeritus members of the Council, or from non-Academic Council members with special competence in some aspect of the dissertation. Only one of the three readers may be a non-Academic Council member. If more than three readers (but not more than five) are on the committee, a majority must be Academic Council or emeritus Academic Council members (Stanford University, n. d.^a).

At the University of Wisconsin—Madison (2013), chairs and co-chairs of dissertation committees must be members of the graduate faculty. Graduate faculty members are those with the rank of professor, associate professor, assistant professor, or instructor in any graduate degree-granting department in the university. Retirees and others who leave the university hold graduate-faculty status for one year. Thereafter, they may serve as co-chairs or other non-graduate faculty members on committees.

Information on how doctoral committees are constituted at George Peabody College of Vanderbilt University was not readily available. Committees that guide the capstone project appear to be constituted around faculty competence and interests.

At Teachers College, Columbia, doctoral committees have two (or more) members: a sponsor, usually the student's major advisor, and one or more other faculty members (Teachers College, Columbia, 2014). Any member of Teachers College with professional rank may serve on committees. Oral examination committees are comprised of two (or more) members and at least one other external examiner selected by the Office of Doctoral Studies. The oral examination committee is chaired by someone other than the student's sponsor.

Openly sharing dissertation work of doctoral students

Milestone examinations during the development and defense of a dissertation may or may not be public events, open to all faculty and practitioners. These events may be advertised widely throughout the academic and practice arenas. Invitations may be issued to individuals who may have an interest in the topic. No requirement for openly sharing or advertising the dissertation or defense was found for Stanford University, the University of Wisconsin--Madison, or George Peabody College at Vanderbilt University. The University of Oxford in Great Britain opens PhD examinations to all faculty members who may attend if they are in academic dress. The examination is published in the *University Gazette*, the official university newspaper (University of Oxford, 2014). Harvard Graduate School of Education requires a public airing of the capstone projects of its students in its Doctor of Education Leadership program (Leddy, 2014).

Separating dissertation advisor from summative evaluation process

A potential conflict of interest occurs when the dissertation advisor and advisory committee are the evaluators of the dissertation. Failure of the dissertation to meet acceptable criteria is a failure of the student, the dissertation advisor, and the dissertation committee. It is unlikely that the dissertation advisor and committee will evaluate the work of the student negatively. To avoid this conflict of interest, some institutions require that the chair or at least one member of the final defense examining committee is an impartial outsider. This is the case at Syracuse University, where the chair of the six-member oral defense committee is appointed by the Graduate School from faculty in other departments (Syracuse University, 2011).

Requiring external reviewers

Stanford University appoints an outside chair to its five-member oral defense committees. The chair is selected from faculty in other departments recommended by the student's department (Stanford University, n. d.^b). In both cases, the chair's responsibility as a voting member is to assure that departmental and graduate school rules and policies governing doctoral study are followed and to protect the academic integrity of the examination and dissertation. At the University of Wisconsin—Madison, one member of the examining committee must be from outside the student's department (University of Wisconsin—Madison, 2014). At the University of Oxford, there are two examiners, both appointed from recommendations submitted by the student and his or her supervisor. One is internal to the student's department, and the other is external to the department. The advisor may attend the *viva voce* (oral defense) (University of Oxford, 2014).

Conducting multiple levels of evaluation

Three levels of evaluation are proposed by de-Miguel (2010) to increase the quality of dissertations: pre-public review by peers, committee review in a public setting, and post-acceptance review by the field. The pre-public review by peers occurs when the student distributes his or her work to peers in the field for review and comment on the quality of the content. These reviews, much like the reviews for refereed articles in journals, may be used to make revisions in the dissertation prior to submission of the document to a committee for review in a public setting. The official committee review is publicly advertised and open for attendance by anyone in the academic or general community. The post-acceptance evaluation occurs when the degree recipient publishes the work through whatever channels and receives feedback on the effect of the work on the development of theory, research, or practice. Although de-Miguel wrote about the process in Spain, his work is applicable to any cultural setting. His three levels of evaluation, if taken seriously, have the potential for improving the quality of doctoral dissertations in any field. The three levels of evaluation, as a whole, were not found at the institutions reviewed for this paper.

Product standards for dissertation research

Product standards run from extremely general to quite specific across the universities reviewed. For example, at the most general level, Stanford makes this statement:

The doctoral dissertation is expected to be an original contribution to scholarship or scientific knowledge, to exemplify the highest standards of the discipline, and to be of lasting value to the intellectual community. (Stanford University, n. d.^a, Rationale, para. 1)

At a slightly more specific level, the University of Wisconsin at Madison and Peabody College at Vanderbilt University make the following statements:

The PhD degree is a research degree and is granted on evidence of general proficiency, distinctive attainment in a special field, and particularly on ability for independent investigation as demonstrated in a dissertation presenting original research or creative scholarship with a high degree of literary skill. (University of Wisconsin—Madison, 2014, Degrees, Minors, Certificates section, para. 5)

Peabody believes the capstone, rather than the traditional dissertation, brings to bear the analytic abilities, professional understanding, contextual knowledge and teamwork skills that are accrued throughout the EdD program, and more closely mirror the challenges of contemporary education practice. (Vanderbilt University, 2015^a, EdD Capstone Experience section, para. 1)

At the most specific level, the Penn Graduate School of Education has the following standards for EdD dissertations:

1. The topic is stated clearly and relevant background literature reviewed and evaluated.
2. The research question(s) are stated clearly.
3. The contribution and importance of the research question(s) with respect to relevant literature, theory, policy, and/or practice are articulated in a convincing manner.
4. The research plan and methods are appropriate and adequate to study the research question(s) posed, and are explicitly described.
5. The research plan and methods are implemented effectively.
6. The research produced trustworthy evidence that bears on the research question(s).
7. The conclusions follow convincingly from the evidence and its interpretation.
8. The dissertation manuscript is coherent, well structured, clearly written and is in accordance with the specifications of a standard style manual regarding grammar, punctuation, spelling, etc.
9. With appropriate revisions, the dissertation is of sufficient quality to be publishable in an academic or practice-oriented journal that is peer reviewed. (Penn Graduate School of Education, 2015, Standards for the Dissertation section, para. 1)

Standards of Quality for Doctoral Research Recommended by Professors of Educational Leadership in Virginia

A survey of 75 faculty members in educational leadership programs in 13 Virginia colleges and universities was conducted. All of the Virginia doctoral-granting institutions were included, and some institutions with faculty

known to have served on doctoral committees were added. Twenty-eight faculty members responded, and 21 responses were useable.

Description of respondents

Four questions were asked about the experience of respondents in education and in supervising dissertation students or serving on dissertation committees. The data are in Table 1.

Table 1

Experience of Respondents in Educational Settings and in Supervising Dissertations, N=21

	Years of experience in public or private education	Years of experience in colleges or universities	Number of dissertations chaired	Number of dissertation committee memberships
<i>M</i>	24.52	15.24	28.10	41.24
Median	27	12	15	30
<i>SD</i>	11.69	11.57	36.79	48.88
Min	4	3	0	0
Max	43	46	140	200

Respondents had much experience in the practice of education. The median for years of experience in public schools, private schools, or other positions associated with education, such as a consultant, was 27 years. The median for years of experience in higher education was 12. The variance is large for both experience groups, with a standard deviation of over 11 years.

The respondents ranged widely in the number of dissertations chaired (0 to 140) and the number of dissertation committees on which they served (0 to 200). The medians of

15 and 30, respectively, for these two variables, indicate that the distribution is heavy on the lower end.

Some faculty members who have been in higher education for many years have served as chair or a member on large numbers of dissertation committees. The Pearson correlation coefficient between years in higher education and number of dissertations chaired was .944. The correlation between years in higher education and the number of committees was .841. Such correlations are to be expected in research-oriented institutions, where faculty

members in earlier years were hired in educational leadership with fewer years of PK-12 experience and a direct interest in research or university teaching. In more recent years, experienced school leaders who may have retired from the PK-12 system have joined the faculties in school administration.

Data collection

Qualtrics survey software was used to distribute a five-item questionnaire. The primary item was “Please write THREE criteria

(you may write more if your wish) that you believe are ESSENTIAL (CRITICAL OR INDISPENSABLE) in assessing the quality of doctoral dissertations IN EDUCATIONAL LEADERSHIP.”

The other four questions requested information on years of experience in universities and public schools and the number of dissertations chaired and committees on which the respondent served. The data are reported here without disaggregation by experience.

Data analysis and findings

The Maykut and Morehouse (1994) constant comparative method was applied in the analysis of the data. Raw data matrices were prepared to summarize the data within categories and subcategories of criteria.

Four large categories of criteria were identified in the data. These were labeled:

1. conditionals,
2. conceptual adequacy,
3. technical adequacy, and
4. advisement adequacy.

Conditionals were statements by respondents about the nature of the dissertation or dissertation research that may affect the criteria that they proposed. Conceptual adequacy contained criteria on the purpose, grounding, and value of a study. Technical adequacy contained criteria on the research methods and presentation of the dissertation. Advisement adequacy had statements about the competence of the advisor. The numbers appearing at the ends of quotations are the identification numbers assigned to the respondents.

Category of criteria #1: conditionals

Respondents made several observations about the nature of doctoral degrees and the research associated with those degrees. Distinctions were made between EdD and PhD degrees, masters’ degrees and doctoral degrees, and degrees with capstone projects and degrees with traditional dissertations. Qualifications about the nature of doctoral research were presented by two respondents:

Distinctions in degrees and related dissertations

Distinctions were made between the EdD and the PhD and the nature of the research appropriate for each. One respondent defined the difference between the EdD and the PhD

when he or she wrote, “Does the dissertation address a significant problem of practice (EdD) or a significant theoretical/methodological issue (PhD)?” (11). Another was concerned that doctoral work was something more than that required for the master’s degree (14). A third had chaired two capstone projects, but made no distinction in criteria for evaluating capstone projects and dissertations. The criteria presented by this respondent could be applied to either capstone projects or traditional dissertations. This person wrote, “Well organized and understand[s] the interconnection of the various chapter[s] of the dissertation” (7).

Qualifications on the nature of doctoral research

Respondents were concerned that the expectations for dissertation research should be reasonable, yet they expected high-quality, verifiable work. One wrote that the dissertation was the first, last, and only piece of research that most EdD students would do (1). This same respondent asserted that dissertation research was semi-independent work and that the quality of the work was the responsibility of the student, the faculty chair, and the committee (1). A second respondent raised the specter of potential misbehavior. Did the student actually do the work? His or her criterion was, “Presentation of data that assures the reader that the work has been done and leads to findings that would be apparent to the reader of those findings, based on the presented data” (9).

Category of criteria #2: conceptual adequacy

There were four sub-categories of conceptual adequacy in the data: originality, grounding, value, and generalizability.

Originality

Originality implies that the dissertation topic is novel; that the student has conceptualized an educational problem in a new, creative, and interesting way; or that the methods of collecting and analyzing data have the potential to contribute to the field in ways that have not been used by prior researchers. Although originality is identified by some universities (for example, Stanford University, n. d.^a, and the University of Wisconsin—Madison, 2013) as a criterion for evaluating the quality of doctoral student research, only one of the respondents listed originality as a criterion. This person thought that the dissertation should offer “something new that augments what is already known” (4).

Grounding

Grounding is situating the dissertation clearly within the area of leadership, basing the dissertation on a framework or on research questions that have been carefully derived from the literature or from practice, and identifying a clear purpose for the work. Grounding had more criteria (18 criteria) than any other subcategory in the conceptual-adequacy category. This is apparently a critical area when faculty members review dissertations.

A focus on leadership. Respondents expected students in educational leadership to do research on leadership. Two of the respondents specifically listed “educational leadership” (3, 26) as the

focus of the dissertation, but another respondent was willing to accept studies that examined leadership more broadly. This respondent wrote, “The study has a component that clearly connects to leadership at some level” (2).

A problem, research questions, or conceptual framework derived from literature or practice. Respondents expected the student to do a thorough review of the literature and create a problem statement, research questions, or a conceptual framework from that review. This theme came through strongly in the criteria. Respondents used such phrases as “a comprehensive review” (2), “conceptual framework or other chain of logic to the topic” (9), “research questions tied in with a conceptual framework” (10), “a thorough awareness of the extant literature” (11), “[a] command of the literature” (18), and “grounding in existing research” (28). It is clear that dissertation chairs and committee members would not look kindly on a dissertation that did not explicitly connect the research questions, the problem statement, the purpose, and the conceptual framework to the research and theory within the field of study. One person wanted the problem studied to be grounded in practice and the research literature (8).

A clear purpose. Anyone who begins a dissertation should have a clear end in mind. The purpose of the work should be clear to the student, and the student should be able to articulate that purpose to his chair, committee, or anyone else who may ask. Purpose is often confused by students with the “what” of their studies. Purpose is about the “why” of their studies. The student must state explicitly “why” he or she is doing the work. Clarity of purpose was offered as a criterion by one of the respondents (15).

Value of the research

Value of the research was the second largest component of conceptual adequacy. Thirteen respondents listed criteria related to the value of the dissertation. They thought that value rested in the extent to which the dissertation might lead to further studies (3); contribute to the development or extension of theory (4); contribute to the field (8, 14, 28) by solving a problem (14), addressing a research need or issue (6, 14, 18), improving practice, generally (4, 11, 18, 19), or improving practice for the individual, specifically (8); or addressing a methodological issue (PhD) (11). One respondent added the general qualifier that the dissertation should have “substance” (20).

Generalizability

Generalizability is a criterion for large scale studies in which samples are taken from a population and statistical techniques are applied to determine whether inferences can be made from the sample statistics to the population parameters. Generalizability is not an applicable criterion in most small-scale, qualitative studies. Only one respondent listed generalizability as a criterion. This respondent expected the dissertation to have “implications beyond the local school division” (3). These implications could be what is meant by “transferability” (Colorado State University, (1993-2014). Transferability exists when the findings of a study that is conducted in one setting are applied to or “transferred” to another setting with similar

characteristics. For example, findings in a study of a fifth-grade classroom in School A in City A are applied to a fifth-grade classroom in School B in City B. The conditions of the two settings may be similar enough for some transfer of findings to be possible.

Value of research

Value of the research was the second largest component of conceptual adequacy. Thirteen respondents listed criteria related to the value of the dissertation. They thought that value rested in the extent to which the dissertation might lead to further studies (3); contribute to the development or extension of theory (4); contribute to the field (8, 14, 28) by solving a problem (14), addressing a research need or issue (6, 14, 18), improving practice, generally (4, 11, 18, 19), or improving practice for the individual, specifically (8); or addressing a methodological issue (PhD) (11). One respondent added the general qualifier that the dissertation should have “substance” (20).

Category of criteria #3: technical adequacy

Technical adequacy had two components: methods adequacy and presentation adequacy. Methods adequacy was the larger of the two and is concerned with whether the overall design and the specific scientific process applied in collecting, analyzing, and interpreting the data are sufficient to answer the research questions and achieve the purpose(s) of the study. Presentation adequacy is concerned with how well the report of the study is written and shared with the community of scholars and practitioners who may be interested in the findings.

Methods adequacy

Methods adequacy had four subcategories: clarity and alignment of research questions, an overall design that is expected to provide data to answer the research questions, trustworthiness in the findings, and alignment across the design components.

Clarity and alignment of the research questions. Respondents expected dissertations to have “clearly defined research questions” (2) that “emanate from the conceptual framework” (13), and are aligned with “a methodology that promises to answer the questions” (2).

An overall design that is expected to provide data to answer the research questions. The overall design of the dissertation research was a critical area of concern for the respondents. Eleven respondents provided criteria for assessing the quality of the design. They thought the design should be “appropriate for the research problem” (8, 17, 19, 24), “aligned to the research questions” (13), “replicable” (9), “clearly stated and rigorously followed” (26), and “defensible” (28).

Trustworthiness. Trustworthiness is the idea that a reader can rely on what the author reports as the results of the study. It is the reader’s assessment of the “truthfulness” and “dependability” of the researcher and his or her findings. One respondent focused directly on trustworthiness by writing “the research [is] carried out in a trustworthy fashion” (4). Another wrote, “[a]

presentation of data that assures the reader that the work has been done and leads to findings that would be apparent to the reader of those findings, based on the presented data” (9). The idea that the methods should have “rigor” ran through several of the criteria. “Quality—the study was well done from a technical standpoint” (4), “sound methodology” (17), “scholarship” (12), “intellectual rigor” (12), “conducted competently ...” (8), and “reflect[s] ...the difference between high-quality and non-rigorous research” (11) were included in the list of criteria. The respondents were concerned that methods were valid (14), the data were appropriately interpreted (17), the findings were appropriately summarized (24) and answered the research questions (26), and the conclusions were appropriately drawn from the findings (24).

Alignment of elements of dissertation. This is the idea that the purpose, research questions, conceptual framework with related literature, population and samples, data collection methods, data analysis methods, findings, and conclusions should be consistent. All of these parts of a dissertation should have the same focus. One respondent focused squarely on this idea when he or she listed “alignment across the entire dissertation” (17) as a criterion.

Presentation adequacy

Presentation adequacy is concerned with how the final report of the dissertation is constructed and presented to the research and practice communities. Dissertation writers should know that the report of their work remains on the World Wide Web for eternity; thus, it must be carefully prepared to avoid embarrassment to their chairs, committee members, and, above all, themselves. Six respondents listed criteria for assessing the presentation of the report. Two listed the “quality of writing” (6, 19) as a criterion. The others wrote that the study should be “well-constructed and easy to follow” (15), and the writing should be “clear” (10, 12), “logical” (10), “organized” (10), and “effective” (20).

Category of criteria #4: advisement adequacy

One respondent addressed the qualifications of the dissertation advisor. Those qualifications were classified into content (conceptual) competence, research (technical) competence, dissertation (process) competence, and personal competence. The respondent thought that an advisor should demonstrate content competence by “currently teaching classes in educational leadership” (7). They should demonstrate research competence by being able to “assist graduate students ... [with] data collection strategies” (7) and by “hav[ing] some working knowledge of research” or by constructing dissertation committees with “at least one member who is strong in statistical design and data analysis” (7). Chairs should demonstrate dissertation-process competence by showing that they “understand the interconnection of the various chapter[s]of the dissertation” (7). Finally, they must demonstrate personal competence by being “well organized” (7).

Conclusions

The first, and primary, conclusion that can be drawn from the data is that faculty members across Virginia believe that there are criteria

that should be applied to the assessment of doctoral dissertations in education. Further, these criteria are associated with the conceptual, technical, and advisory adequacy of

the dissertation. In their minds, conceptual adequacy is concerned with the originality, grounding, generalizability, and value of the dissertation. Technical adequacy is concerned with the methods applied in doing the study and the literary skills with which the study is presented to the public. Advisory adequacy is associated with the competence of the chair of the dissertation committee.

A second conclusion is that the adequacy of advisement should be assessed with criteria that focus on the content, research, dissertation-process, and personal competence of the advisor. This conclusion is based on the responses of only one person; however, this person has raised a major issue in the evaluation of dissertations. The quality of a dissertation is dependent upon the quality of the inputs, and one of the critical inputs is the quality of the advisement received by the student.

A third conclusion is that the doctorate in educational leadership is in somewhat of a muddle. Some respondents made clear that a distinction should be made between the EdD and the PhD, between the master's degree and the doctorate, and between dissertations and capstone projects. The criteria reported by nearly all respondents are appropriate for the traditional research dissertation and may not be appropriate for the variety of dissertation types that are being developed in educational leadership.

Discussion of the Findings

Doctoral research in educational leadership appears to be stuck in the past. The Carnegie Foundation has been promoting reform in doctoral programs with its Carnegie Initiative

on the Doctorate (CID) since 2001 (Golde, Walker, & Associates, 2006).

Educational leadership is one of the areas targeted for reform, and one of the reforms is to reconstruct the preparation of leaders at the doctoral level. Reconstruction would focus attention on preparing educators for practice.

Preparation for practice would include a capstone project rather than a traditional dissertation. A few institutions (Harvard and Peabody College at Vanderbilt for two) have moved in this direction. Most, however, continue to use the traditional methods of preparation, including a research-based dissertation, in their programs. To date, the Carnegie initiative seems to be ignored by or not visible to most faculties in educational leadership. The result is the continuation of the past, and the questions asked and the criteria presented in this brief piece of research in Virginia reflect this orientation.

The fact that capstone projects were raised by one person in this study shows a small crack in the monolithic approach to doctoral education. That small crack may be a sign that university faculty should begin a serious discussion of the nature of doctoral education in educational leadership and the processes that we use to stimulate and further that education.

The result of these discussions may be a reconstruction of how we do our work and what we ask of our students. This does not mean that we must have uniform programs.

What it does mean is that we must have "thoughtful" programs for preparing our school

leaders at the doctoral level. With respect to criteria for assessing the quality of doctoral research, it means that we should have multiple sets of criteria, depending upon the nature of that research. For those school leaders contemplating taking an advanced degree in educational leadership, the findings of this study have several implications.

First, they should anticipate some turmoil as university faculties come to grips with the nature of doctoral education and research.

Second, they should do some thinking about the kind of advanced education they want and pursue a seat in those universities that have programs that match their preferences.

Third, those practitioners who enter the university following employment in the schools should express their views and exert their influence in departments of educational leadership to bring change in how educational leaders are prepared and how the research in educational leadership is conducted.

Author Biography

David Parks has authored or coauthored numerous publications on education and leadership, presented scores of papers at conferences, and consulted with many educational organizations. He has directed the doctoral research of over 100 students. E-mail: parks@vt.edu

References

- Colorado State University. (1993-2014). *Generalizability and transferability*. Writing@CSU. Colorado State University. Available at <http://writing.colostate.edu/guides/guide.cfm?guideid=65>
- de-Miguel, M. (2010). The evaluation of doctoral thesis. A model proposal. *RELIEVE: e-Journal of Educational Research, Assessment and Evaluation*, 16(1), 1-17. Retrieved from http://www.uv.es/RELIEVE/v16n1/RELIEVEv16n1_4eng.pdf
- Golde, C. M., Walker, G. E., & Associates. (2006). *Envisioning the future of doctoral education: Preparing stewards of the discipline*. San Francisco: Jossey-Bass.
- Leddy, C. (2014, April 18). A capstone to learning. *Harvard Gazette*. Retrieved from <http://news.harvard.edu/gazette/story/2014/04/a-capstone-to-learning/>
- Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. Washington, DC: The Falmer Press.
- Penn Graduate School of Education, University of Pennsylvania. (2015). *Degree requirements: Doctor of education (EdD)*. Retrieved from <http://www.gse.upenn.edu/edd#Dissertation>
- Stanford University. (n. d.^a). *Doctoral degrees: Dissertations and dissertation reading committees (GAP 4.8)*. Retrieved from <http://gap.stanford.edu/4-8.html>
- Stanford University. (n. d.^b). *Doctoral degrees: University oral examinations and committees (GAP 4.7)*. Retrieved from <http://gap.stanford.edu/4-7.html>
- State Council of Higher Education for Virginia. (n. d.). *Degrees awarded--C01A2: Completion, program detail*. Retrieved from http://research.schev.edu/Completions/C1Level2_Report.asp
- Syracuse University. (2011). *What you need to graduate*. Retrieved from http://www.syr.edu/gradschool/em/current_whatyouneed.html
- United States Department of Education, National Center for Education Statistics. (2013). *Table 318.30. Bachelor's, master's, and doctor's degrees conferred by postsecondary institutions, by sex of student and discipline division: 2011-12. Digest of Education Statistics*. Retrieved from http://nces.ed.gov/programs/digest/d13/tables/dt13_318.30.asp
- University of Oxford. (2014). *Research examinations*. Retrieved from <http://www.ox.ac.uk/students/academic/exams/research>

University of Wisconsin—Madison. (2013). *Academic policies and procedures*. Retrieved from <https://grad.wisc.edu/acadpolicy/>

University of Wisconsin—Madison. (2014). *Examinations*. Retrieved from <https://elpa.education.wisc.edu/elpa/academics/PhDDegreeRequirements/examinations>

University of Wisconsin—Madison. (2014). *Graduate school catalog, 2014-2016*. Retrieved from <http://grad.wisc.edu/catalog/degrees.htm>

Vanderbilt University, Peabody College. (2015^a). *Ed.D. capstone experience*. Retrieved from http://peabody.vanderbilt.edu/departments/lpo/graduate_and_professional_programs/edd/capstone_experience/index.php

Vanderbilt University, Peabody College. (2015^b). *Ed.D. program*. Retrieved from http://peabody.vanderbilt.edu/departments/lpo/graduate_and_professional_programs/edd/

The Glass Maze and Predictors for Successful Navigation to the Top Seat to the Superintendency

Denise DiCanio
 Doctoral Alumna
 Dowling College
 Oakdale, NY

Laura Schilling
 Doctoral Student
 Dowling College
 Oakdale, NY

Antonio Ferrantino, EdD
 Project Manager
 Office of Diversity and Affirmative Action
 Stony Brook University
 Stony Brook, NY

Gretchen Cotton Rodney
 Doctoral Student
 Dowling College
 Oakdale, NY

Tanesha N Hunter, EdD
 CSE/CPSE District Administrator
 Department of Special Education
 Rocky Point Union Free School District
 Rocky Point, NY

Elsa-Sofia Morote, EdD
 Professor
 Department of Educational Administration,
 Leadership and Technology
 Dowling College
 Oakdale, NY

Stephanie Tatum, PhD
 Associate Professor
 Department of Educational Administration,
 Leadership and Technology
 Dowling College
 Oakdale, NY

Abstract

A predictive model of assistant superintendents willingness to become superintendent was created using three factors: personal (age, gender, marital status, and parenthood), professional (district size, district needs, and being mentored), and volition (willingness to appear for multiple interviews, give up their current position, be interviewed by search firms, build alliances within the community, and the desire to lead a district). One hundred and forty-nine assistant superintendents in diverse areas participated in a survey distributed in New York, 70 females and 79 males. The results showed the most influential variables in the assistant superintendent's willingness to become a superintendent are district size, type of mentorship, and volition for both females and males but to differing degrees.

Key Words

Superintendent ascendancy, assistant superintendents, gender

Introduction

Ella Flagg Young was the first female to hold the position of superintendent of Chicago Public Schools, superintendent in any major U.S. city, and president of the National Education Association. In 1909 she stated, “In the near future we will have more women than men in executive charge of the vast educational system” (Blount, 1998, p.1). Although women made significant gains in school district leadership over the next several decades, the end of World War II brought the beginning of a steady decline in the number of women occupying the top position. From 1945 to 1970 the number of female superintendents declined (Blount, 1998), which continued into the 21st Century.

An analysis of the demographic trends in school administration from the early 1920s to 2010 revealed that gender inequity existed in the position of the superintendent of schools. The percentage of female superintendents was not proportionate to the percentage of females in the field of education or to the general population of the United States. A 2010 survey of superintendents conducted by the American Association of School Administrators (AASA) indicated women account for less than a quarter of the nation’s superintendents; yet they make up 75% of the teaching force (Kowalski, McCord, Petersen, Young, & Ellerson, 2011).

With women outnumbering men in school administration graduate programs, why do they continue to lag so far behind men in the acquisition of a superintendent position? There are several theories explored in the literature including unfavorable working conditions and gender bias (Harris, Lowery, Hopson, & Marshall, 2004; and Glass & Franceschini, 2007). Whitaker and Lane (1990) found that

gender “determines the role an individual will be assigned in education.” Wesson (1998) noted that organizations fill positions in upper management with candidates that fit the organization’s existing schema. Men are seen as being better at handling discipline, working with school boards, and navigating the politics of the superintendency (Logan, 1998).

Organizations often see women as less favorable candidates for leadership positions, and when they do occupy leadership roles, displaying traditional leadership behaviors is seen negatively (Eagly & Karau 2002). Eagly and Karau go on to explain that societal beliefs hold that gender roles ascribed to women are in direct contradiction to traits required for successful leadership. In a study conducted by Elsesser and Lever (2011), however, there found to be an improvement in the perception of women in leadership roles.

This study examines the data collected from a survey developed by Hunter (2012) that was administered to 200 assistant superintendents in Nassau, Suffolk, and Westchester counties in New York with 149 responding. The instrument was originally designed to measure how the willingness to compete for a superintendent position was affected by internal motivators, external motivators, internal barriers, and external barriers. This study realigned the survey items to create a new variable, volition.

The purpose of this paper was to investigate if personal variables (gender, marital status, and parenthood), professional variables (district size, current position within the district, and being mentored), or volition predict the level of willingness an assistant superintendent has for pursuing the role of

superintendent. The research question guiding this study was as follows: was willingness to pursue a superintendent position influenced by personal variables, professional variables, and volition in females and males?

Literature Review

Volition

Vogel (1985) developed a four-factor model that delineated the lack of volition for women to become superintendents. The first factor, woman's place, identified women as caretakers and men as leaders. The second factor, discrimination, found that men were promoted over women based on gender, and school boards advanced men over women. The third factor, meritocracy, the implementation of advancement based upon intellectual talent, deemed men were more intelligent. The fourth factor, economic, indicated women worked for lower pay and the few leadership positions commanded a higher pay.

Cooper, Fusarelli, and Carella (2000) used the *Superintendents' Professional Expectations and Advancement Review* (SPEAR™) survey and found that superintendents were leaving education due to lack of proper preparation for the position. This unpreparedness resulted in many school boards filling superintendents' positions with retirees, decreasing opportunities for women and other traditionally disenfranchised groups to become a superintendent (Wolverton & Macdonald, 2001). Glass (2000), Wolverton and Macdonald (2001) suggested that volition to become a superintendent arose from opportunities afforded to the individual.

Different factors affect women's volition to pursue the role of superintendent. Leadership resilience, or the ability to bounce

back from adversity, enables women to take risks regardless of criticism and challenges (Patterson, Goens, & Reed, 2009). The key factor toward advancing to the role of superintendent involved stamina to sustain challenges rather than abilities or experience. MacTavish (2010) found cumulative education, experience, and endorsement from mentors the most salient factors contributing to a feeling of readiness to ascend to the position of superintendent.

Gender

The perception of gender differences originated from the time of Aristotle where he viewed women as defective (Jones & Montenegro, 1982), lowering women's contribution to society. The important attributes for a superintendent such as competitiveness, assertiveness, and aggressiveness were perceived negatively in women (Marshall, 1986). The societal schemata of women and work historically emphasized child caretaking (Patton & McMahan, 2006). Caceres-Rodriguez (2011) echoed this societal perception as a cultural norm deeply ingrained in organizational structures. This perception may have prevented many women from attaining higher leadership positions.

Hegemonic perceptions about the creation of organizations and valued experiences of members in organizations were based on males (Meyerson & Fletcher, 2000). Skrla, Reyes, and Scheurich (2000) viewed gender inequity as the primary reason women do not advance in the executive suite, which prevents many from attaining their professional goals. Although women make up a larger portion of the teaching profession, men were 40 times more likely to become superintendents as compared to women.

Teaching, then, became a *feminine role* while administration became a *masculine role* (Tyack & Strober, 1981; Kowalski & Brunner, 2011). According to Poll (1978), women constituted 85% of elementary school teachers, 20% of elementary school principals and 1% of superintendents.

The most recent available figures indicated that approximately 18% of superintendents in the USA are female (NCES, 2003). The 2007-2008 Schools and Staffing Survey supported Poll's (1978) finding that women were not proportionately represented in the position of superintendent (Shakeshaft, 2011). In 2009, the North Carolina Department of Public Instruction indicated that 80% of the teachers were female while 82% of the superintendents were male (Shakeshaft, 2011). With this trend, women would not hold the position of superintendent at the rate of their male contemporaries for 77 years.

Grove and Montgomery (1999) indicated, "one reason so few women are hired for educational administrative positions is due to the gender gap". They discussed three theories on why women have not dominated leadership positions in the education field. One theory is psychological and tied to power. The way women use power to empower others may be viewed by others as not desiring power for themselves (Grove & Montgomery, 1999). Gupton and Slick (1996) cautioned women about creating their own glass ceiling by doubting themselves and their potential to succeed in leadership positions. Other theories regard limitations placed on women through structure within the educational system and social norm discriminatory practices (Grove & Montgomery, 1999).

For the last twenty years, there was an increase in gender equity issues in the leadership of public education (Blount, 1998; Glass, 2000). Women continued to receive inequitable treatment in terms of pay, promotions, and authority (Eagly & Carli, 2003). Fernandez (2007) reported public policies to change gender inequity provided a limited effect.

In 2010, Congress, through the Government Accountability Office (GAO), investigated women's representation in management positions and pay differences. Their investigation determined a need for additional information about the challenges women face in advancing their careers (Sherrill, 2010). Although structural barriers impede women from advancing to the position of superintendent, researchers noted internal barriers might contribute to the willingness of some women to advance to the position of superintendent.

For example, Grove and Montgomery (1999) noted in addition to the gender inequity embedded in the infrastructure of many educational systems, some women use power to empower others and not necessarily themselves. Gupton and Slick (1996) identified some women might have self-doubt regarding their potential and choose not to seek the position of superintendent, which perpetuates the normalization of social norm discriminatory practices (Grove & Montgomery, 1999).

When female leaders advance in their organizations, they tend to "emphasize empowerment, affirm relationships, seek ways to strengthen human bonds, simplify communications and give means an equal value

with ends” (Helgesen, 1990, p.52). These characteristics highlight the development of shared values, traditions, and ideas administrators tend to focus on as they serve as the catalyst to create a learning community (Sergiovanni, 1992, 1996).

Cultural Fit

Fifty-nine percent of the respondents on Cubiks international survey on job and cultural fit (<http://www.cubiks.com/survey/Pages/default.aspx> 2015), indicated that they would be in favor of dismissing a high potential candidate if they were out of step with the organizations culture. Chatman (1991) stated that organizations devote resources in maintaining a good fit for their employees and organization because they assume some employees are better suited to perform certain jobs compared to other employees. As Rivera (2015) argues that cultural fit, or organizational fit can be positive, it can dilute the organization and create feedback loops that exclude highly qualified candidates who may not meet what the expected culture of the organization or leadership of the said organization.

Hewlett, Leader-Chivée, and Sumberg (2012), stated that sponsorship and development of pipelines is important with moving up within organizations and grooming leaders through sponsorship within the organization. While Rooth (2010) stated that individuals members of organizations that hold gatekeeper roles, such as recruiters, may have an unconscious association bias, which adversely impacts people not in the proscribed norm.

Rooth (2010) observed that negative stereotypes create bias that discriminates against potential candidates. This form of

implicit bias is due to perception of organizational fit and creates an adverse impact on people who do not seem to fit within the norm (Kayes, 2006). Promotional opportunity is prevented when decision-makers dilute the individual’s accomplishment by not taking into account the individual’s merit, but rather the perception of their merit through the lens of a stereotypical bias of the observer (Kayes, 2006).

Cubik (2013) used the International Survey on Job and Cultural Fit and found 59% of the respondents indicated that they would be in favor of dismissing a high potential candidate if they were out of step with the organizational culture. Rivera (2015) argues that although cultural fit, or organizational fit, can be positive, it can dilute the organization and create feedback loops that exclude highly qualified candidates who might not meet the norms of the expected organizational culture as it pertains to leadership practices within the organization.

Hewlett, Leader-Chivée, and Sumberg (2012) stated that sponsorship and development of pipelines is important to moving up within organizations and grooming leaders through sponsorship within the organization.

Rooth (2010) stated that individual members of organizations who hold gatekeeper roles, such as recruiters, might have an unconscious association bias, which adversely affects people not in the prescribed norm. Rooth observed that negative stereotypes create bias that discriminates against potential candidates. This form of implicit bias is due in part to perception of organizational fit and creates an adverse impact on people who do not seem to fit within the norm (Kayes, 2006).

Mentorship

Grove and Montgomery (1999) stated that for women to succeed in attaining administrative positions in education, mentoring must occur. In the early twentieth century, organized efforts to mentor and advocate for men in administrative positions included The Male Teachers' Association of New York City (Blount, 1998).

Women, however, needed more education and more experience as compared to men for the same administrative position (MacTavish, 2010; Weatherly, 2011). Negative views about the position of superintendent as an *old boys' network* contributed to women's belief that the position of superintendent was unattainable (Ottino, 2009; Weatherly, 2011; Wickham, 2007).

According to Askren-Edgehouse (2008), 50% of female superintendents surveyed in Ohio reported having male mentors who helped them attain the position of superintendent. Ottino (2009) found that 18% of women pursuing the position of superintendent perceived mentors and networking as affecting their chances of achieving the position of superintendent. However, women did not feel empowered to change the *old boys' network* and would prefer to keep their less stressful job, which supports what Ceniga (2008) identified as mentors and networking are seen as infrastructural barriers.

MacTavish's (2010) mixed-method study illustrated how superintendents used mentors, sponsors, and networks. Mentors were seen as "one who helps teach and aspirant the job responsibilities and norms of the superintendency and who helps the aspirant

grow personally and professionally in pursuit of that position" (MacTavish, 2010, p.8). Sponsors are defined as "one who actively champions and make contacts on behalf of an aspirant in order to gain a desired position" (MacTavish, 2010, p.8).

Findings indicated mentors included their own district superintendent, outside district superintendents and university professors. Three sources of sponsorship for women aspiring to become a superintendent were their own superintendent, a board member, or a professional colleague. Zachry (2009) found that it was important for female superintendents to target and encourage potential female educational leaders through mentorship, networking, sponsorship, and advocacy.

Women were less likely to seek a sponsor because of possible challenges associated with a male sponsor; there is greater scrutiny of the sponsorship relationship due to issues surrounding sexual harassment (Hill & Ragland, 1995; Hewlett, Peraino, Sherbin, & Sumberg, 2010; MacTavish, 2010). Hewlett *et al.* (2012) concluded that beyond mentors women needed sponsors, advocates who create a pipeline to senior leadership positions.

Wickham (2007) found that perceptions of success differed in high school and elementary school administrative positions. Administrators at the elementary level who aspired to become a superintendent involved acquiring a doctoral degree and exhibiting high-level curriculum vitae while administrators at the high school level who aspired to be a superintendent employed the use of a mentor.

Shore, Coyle-Shapiro, Chen and Tetric (2009) found hidden issues for advancement such as the lack of mentors and networking. The absence of mentors and networking was a significant factor contributing to the lack of women ascending to the position of superintendent (Weatherly, 2011). Weatherly examined female superintendents' perception of importance of 11 types of mentoring functions in Texas. Eighty-eight out of 140 women responded to an online 5 point Likert scale survey.

The results indicated the following mentoring functions were important to attaining the position of superintendent: sponsoring, coaching, challenging assignments, exposure and visibility, friendship, role model, and acceptance. The intersection of networking, mentorship, and sponsorship forms a complex synergistic effect that promotes one becoming a superintendent.

District Size

Grounded in motivation environmental theory, Laramore (2010) studied factors that positively influenced superintendents and non-superintendents in applying for the position of superintendent.

In terms of district size, male superintendents from large districts were more satisfied than females. In comparison, female superintendents from small districts were more satisfied than their male colleagues. Conversely, for non-superintendents, large districts appealed to females while small districts appealed to males.

Bolla (2010) found the size of the district affected how female superintendents

approached the role of superintendent more than male superintendents. Differences in the size of the district impacted public relations as well. In smaller districts female superintendents spent less time on politics than female superintendents in large districts.

Consequently, aspiring female superintendents needed to be aware of district size differences to determine their best option (Bolla, 2010).

Methodology

Design

The study examined both male and female assistant superintendents and their willingness to move up to the superintendent position. Using SPSS version 19 for statistical analysis, a binary logistic regression was conducted after the data file was split by gender to find the best model to predict willingness for assistant superintendents' ascension to the position of superintendent.

The dataset came from a larger study conducted by Hunter (2012) who examined barriers and motivators that men and women encountered in route to the position of school district superintendent. One hundred forty nine female and male assistant superintendents within Suffolk, Nassau, and Westchester counties in New York responded to the survey.

In order to examine the willingness to be superintendent a predictive model was created using three factors: personal (age, gender, marital status, and parenthood), professional (district size, district needs, and being mentored), and volition (willingness to appear for multiple interviews, give up their current position, be interviewed by search

firms, build alliances within the community, and the desire to lead a district).

Those factors were chosen because other variables in the dataset were found to be non-significant in the prediction of willingness. A factor analysis was conducted to establish the construct validity of the instrument (Hunter, 2012).

The willingness to pursue the superintendent position was taken into consideration within the survey (Hunter, 2012). A variable called volition was generated from the following items:

- q62 How willing are you to appear for multiple interviews with the board of education?
- q61 How willing are you to give up your current position?
- q63 How willing are you to be interviewed by search firms?
- q65 How willing are you to build alliances within the community for the schools?
- q33R Lack of desire to lead a district

Volition in this study has been defined by the willingness of assistant superintendent's to appear for multiple interviews, give up their current position, be interviewed by search firms, build alliances within the community, and the desire to lead a district. The dependent variable chosen was item *q60: How willing are you to pursue a job as a superintendent?*

The high and low levels were established by recoding the 5-point Likert

scale. The low level was a combination of the Likert choices of not willing at all and a little willing (1 & 2). The high level was a combination of the Likert choices of willing and very willing (4 & 5). Volition had a Cronbach Alpha Reliability of 80%. (Note *q33* was a reverse question and was recoded (shown as *q33R*).

Participants

Participants for the study were holding a position as an assistant superintendent within Suffolk, Nassau, and Westchester Counties in New York from a pool of 125 school districts; specifically 69 from Suffolk, 56 from Nassau, and 47 from Westchester respectively. Two hundred assistant superintendents were invited to participate and complete the survey; 149 participants returned completed surveys for a 75% response rate.

Of the completed surveys, 60 respondents (40%) came from Suffolk, 57 respondents (38%) came from Nassau, and 32 respondents (22%) came from Westchester. Of the participating 149 assistant superintendents, 55 (36.9%) reported their current positions as the Assistant Superintendent of Curriculum and Instruction, 53 (35.6%) as the Assistant Superintendent of Business and Finance, 15 (10.1%) as the Assistant Superintendent of Human Resources, 11 (7.4%) as the Assistant Superintendent of Personnel, 2 (1.3%) as the Assistant Superintendent of Operations, 11 (7.4%) as the Assistant Superintendent of Special Education, and 2 (1.3%) reported their assignment as other. Table 1.1 provides a breakdown of the district size for the assistant superintendents within this study.

Table 1.1

District size (number of students enrolled in district)

	Total Frequency	Female Frequency	Male Frequency	Percent
1,000 – 2,999	40	22	18	26.8
3,000 – 4,999	56	24	32	37.6
5,000 – 9,999	44	21	23	29.5
10,000 +	8	2	6	5.4
No response	1	1	0	0.7
Total	149	70	79	100.0

The district type of a majority of the respondents was suburban (89.3%). The remainder were from rural (3.4%), small town (4.0%), and urban (3.4%). The district needs levels were categorized as 28 high needs (18.8%), 57 moderate needs (38.2%), and 64 low needs (43.0%).

The respondents' genders were 79 male (53%) and 70 female (47%). From the 149 respondents, 136 self-identified as *White* (91%), 5 self-identified as *Black* (3%), 5 self-identified as *Hispanic or Latino* (3%), 1 person self-identified as *Asian* (>1%), and 2 self-identified as *other* (1%).

One hundred twenty two (82%) of the respondents self-identified as *married*, 9 (6%) respondents self-identified as *single* (never married), 15 (10%) self-identified as

divorced/separated, and 3 (2%) self-identified as *widowed*.

The age range of the respondents was from 33 to 69. The age distribution of the respondents: 14.8% of respondents were ages 33 to 41; 27.5% of respondents were ages 42 to 50; 38.9% of respondents were ages 51 to 59; and 18.8% of respondents were ages 60 and 69.

Table 1.2 revealed that 46% of the respondents reported having a mentor. Twenty-three respondents reported that their mentor was a superintendent in their district, 8 respondents reported that their mentor was a superintendent in another district, 37 respondents reported that their mentor was someone who was not a superintendent. Fifty-four of the respondents reported that they did not have a mentor.

Table 1.2

Mentor

	Total Frequency	Female Frequency	Male Frequency	Percent
Yes, the Superintendent in my district mentored me	23	11	12	15.4
Yes, the Superintendent in another district mentored me	8	5	3	5.4
Yes, someone who was not a Superintendent mentored me	37	15	22	24.8
No, I did not have a mentor	81	39	42	54.4
Total	149	70	79	100.0

Results

The initial logistic regressions included a predictive model of the willingness to be superintendent based on three factors: personal, professional, and volition. The data were split by gender, and the result or best-fit model only held onto the volition variable, the district size, and the mentor types. The other variables, such as marital status, age, district type, and district needs level showed no

significance in the prediction of willingness to become a superintendent. The dependent variable was high-low willingness. The independent variables (volition, district size, mentorship—see Table 2) are significant predictors of willingness to advance into the position of superintendent, with a large effect size of approximately 60%.

Table 2

Variables in the equation

Gender			B	S.E.	Wald	df	Sig.	Exp(B)
Female	Step 1 ^a	Volition	.98	.32	9.39	1	.002	2.67
		District Size	1.07	.94	1.28	1	.258	2.91
		Mentor			3.46	3	.325	
		Mentor(Supt in district)	3.37	1.90	3.17	1	.075	29.17
		Mentor(Supt out of district)	-19.41	15706.88	.00	1	.999	.00
		Mentor(not a Supt)	2.02	1.56	1.67	1	.196	7.54
		Constant	-22.79	7.93	8.26	1	.004	.00
Male	Step 1 ^a	Volition	.69	.18	14.57	1	.000	1.99
		District Size	.75	.51	2.15	1	.143	2.11
		Mentor			2.39	3	.496	
		Mentor(Supt in district)	1.88	1.26	2.20	1	.138	6.53
		Mentor(Supt out of district)	-19.49	25038.53	.00	1	.999	.00
		Mentor(not a Supt)	-.02	1.26	.00	1	.991	.99
		Constant	-14.69	4.26	11.90	1	.001	.00

a. Variable(s) entered on step 1: Volition, District Size, and Mentor.

Table 2 is the final result of the logistic regression for females and males' willingness to apply for the superintendency. The volition, mentorship, and district size variables contribute to the predictive model. The volition and district size did add to the predictive model, but varied less across gender than mentorship. The mentor variable reveals that a mentor who is a superintendent from another district does not impact the willingness to move up for either gender. A mentor who is not a superintendent has some influence on both genders.

However, the impact is 7.5 times larger for females compared to males. The most influential type of mentor in this predictive model is a superintendent within the assistant superintendent's district. Yet again the power of this type of mentor is much larger for females; in fact they are 29 times more likely to increase the willingness for advancement. The males are affected by this mentor-type, but are only 6.5 times more likely to increase their willingness for advancement.

Limitation and Delimitations

The geographic location of the participants within this study from Nassau, Suffolk, and Westchester Counties in New York are regarded as relatively affluent in comparison to other regions thus somewhat a limitation.

Discussion and Implications

This study examined the personal and professional variables, including volition, gender, marital status, age, district type, district needs' level, district size, and the presence of a mentor, that contribute to an assistant superintendent's willingness to pursue the position of superintendent. It shows that the motivating factors for both men and women are

similar and include district size and mentor-type. Although the size of the district in which the assistant superintendent is currently assigned contributes the most to the predictive model generated in this study, there is no significant difference between its effect on females and males willingness to pursue the position of superintendent. This study found that regardless of gender, the individual level of volition affects both female or male assistant superintendents' professional perseverance and level of aspiration.

Volition and investment of mentorship to sponsorship support an idea of cultural fit within district leadership positions. Whereas, individual volition is promoted by feeling that the goal to the top seat is attainable and deserving by either one's own volition or mentorship support.

The results uncover the importance of close proximity of support to increase volition. Moorosi (2010) indicated that professional and family support positively impacted overall job satisfaction of South African female principals. A mentor or sponsor within the district would play a critical role in supporting and increasing the volition of female assistant superintendents to aspire for that top seat.

The type of mentorship is a significantly stronger indicator for women who have mentors within their school district. This finding supports MacTavish (2010) who found that female superintendents reported mentors and sponsors were most often superintendents from within their district. Hunter (2012) does not delineate the difference between a mentor and a sponsor. However, it is inferred that these mentors within their school district were actually sponsors who helped increase assistant

superintendents' willingness to become a superintendent, whereas, women who have mentors outside of the district did not have the same drive to move forward to the top position. Mentors outside of the district may not have the same access to influential people in the organization to provide anticipatory socialization, as well as navigate the political landscape of particular school districts.

This is an important observation because it describes how the perceived glass ceiling may actually be a glass maze and without sponsorship, women may become frustrated with navigating what Ottino (2009) describes as an *old boys' network*. A formal support system, well-developed networks and mentor/sponsor, are critical to undertake the necessary steps to move into the position of superintendent.

The absence of a mentor impedes advancement as shown in research conducted by Shore *et al.* (2009) and Weatherly (2011). Women who have mentors within the school district are 29 times more likely to pursue the role of superintendent, while males are 6.5 times more likely. This finding echoes Shakeshaft's (1979, 2011) work, which indicated that support and encouragement is necessary for women to move toward the position of superintendent.

Boards of education, superintendents, and other stakeholders should endorse formal mentorship programs in the district, as this might promote the idea of cultural fit as conceptualized by Hewlett *et al.* (2012) who indicated sponsorship and development of pipelines are important when preparing aspiring superintendents for the role. Sponsorship can assist aspiring superintendents with developing

leadership practices conducive to the growth and development of all members in the organization.

These normalized practices shape the organizational culture in that females aspiring to become a superintendent have more opportunities to enter the pipeline in the district size they choose. For this to occur, there has to be a shift in decision makers thinking regarding the knowledge, skills, and professional disposition that females contribute to the organization as well.

Appreciating the accomplishments and merit in performance requires viewing the accomplishments through a lens not rooted in stereotypical perception of those in the pipeline that were in the teaching profession (Kayes, 2006). That is, if there is an assumption, due to implicit and association bias (Rooth, 2010), whether the denial of access to the position is intentional or unintentional, the outcome is still the same: females are not in the role of superintendent to the same degree as their male contemporaries even though they make up the majority of the teaching profession.

There is a chance of a missed opportunity to recognize these pedagogical practices can inform leadership practices that focus on doing what is in the best interest of student engagement and learning, which are critical elements in the schooling process. Reducing and ultimately eliminating the navigation of the glass maze might provide a straight ascension to the position of superintendent, particularly for females with the volition to take on the role of superintendent, as the removal of structural barriers can provide a clear pathway for

qualified candidates and have more females in the pipeline.

Another consideration regarding the elimination of structural barriers is the recruiting firm. In order for traditionally disenfranchised groups to have an opportunity to participate in the interview process for the position, those providing the pool of qualified candidates to school boards of education must recognize the formal qualification, as well as appreciate the knowledge, skills, and professional dispositions females bring to the organization.

If the perception is contextualized in deficit thinking regarding females, including recruiters who share group membership, the opportunity for females to advance is stagnated. Thus, recruiters should have training that target implicit and association bias to become aware of what cultural fit in an organization should include—various lived experiences that inform leadership practices (Cubik, 2013).

Hewlett et al. (2010) indicated that without sponsorship and advocacy, qualified women would not have the support, opportunity, nor be inspired to advance. Many women have to deal with the precarious situation of being assertive and aggressive to find the right sponsor. In order to ascend to a top leadership position they cannot sit around and wait for acknowledgement of a job well-done (Hewlett et al., 2010). This concurs with Oritz's (1980) finding that females' silence about their aspirations and accomplishments perpetuated limited opportunities.

The educational landscape is an environment that requires leaders to be proactive and move forward with intention

toward student's educational attainment, as they should enter the workforce with the knowledge and skills required to become contributing members of the complete social structure. Those wishing to become superintendent to assist students in the process through their leadership, especially women, must position themselves in a way that garners sponsorship to expedite their journey through the glass maze of top-level leadership in order to acquire the position.

To some degree, what Growe and Montgomery (1999) discussed in the context of the gender gap, reasons why women have not dominated leadership positions in the field of education, should be a consideration when developing mentor and sponsorship programming.

Becoming aware of these nuanced differences might encourage more females to choose the position of superintendent by recognizing their contribution to the role of superintendent has value. Specifically, they noted the way women use power to empower others might be perceived by others as not desiring power for themselves. Awareness of this perception might allow women to leverage this aspect of a transformational leadership practice, empowerment, in ways that produce a more favorable outcome for them—securing the position of superintendent.

Suggestions for future studies are to investigate whom the mentors are within the school district and determine their influence, organizational knowledge, and gender. The exploration of the process to form successful mentor relationships should occur. Hewlett *et al.* (2010) indicated that sponsorship is more important than mentorship. Further research

should include examining the following: gaining access to sponsors in a school district; reasons for the sponsorship; and how gender affects potential decisions to sponsor a woman or man within the school district. An exploration of the confluence of issues in

networking, mentorship, and sponsorship, as well as their complex synergistic effects, will provide insight into changing deeply held tenets and propel women aspirants through the glass maze to the top leadership position.

Author Biographies

Denise DiCano holds a doctorate from Dowling College in Oakdale, NY.

Laura Schilling is a doctoral student at Dowling College in Oakdale, NY.

Antonio Ferrantino is project manager of the office of diversity and affirmative action at Stony Brook University in Stony Brook, NY.

Gretchen Rodney is a doctoral student at Dowling College in Oakdale, NY.

Tanesha Hunter is the CSE/CPSE district administrator in the department of special education at Rocky point Union Free School District in Rocky Point, NY.

Elsa-Sofia Morote is a profesor in the department of educational administration, leadership and technology at Dowling College in Oakdale, NY

Stephanie Tatum is associate professor in the department of education administration, leadership and technology at Dowling College in Oakdale, NY.

References

- Askren-Edgehouse, M. (2008). *Characteristics and career path barriers of women superintendents in Ohio*. Bowling Green State University. *ProQuest Dissertations*
- Blount, J. M. (1998). *Destined to rule the schools: Women and the superintendency, 1873-1995*.
- Bolla, J. (2010). *The perception of the role of the superintendent in small, medium and large schools*. Saint Louis University. ProQuest Dissertations and Theses, 126-n/a. Retrieved from <http://0search.proquest.com.library.dowling.edu/docview/507900896?accountid=10549>
- Caceres-Rodriguez, R. (2011). The glass ceiling revisited: Moving beyond discrimination in the study of gender in public organizations. *Administration & Society*. Doi: 10.1177/0095399711429104
- Ceniga, B. (2008). *Women's paths to the superintendency in Oregon*. Lewis and Clark College. *ProQuest Dissertations and Theses*, 125-n/a. Retrieved from <http://0search.proquest.com.library.dowling.edu/docview/304377552?accountid=1054>
- Chatman, J. A. (1989, August). Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly*, 36, pp. 459-484.
- Cooper, B., Fusarelli, L., & Carella, V. (2000). *Career crisis in the superintendency?* Arlington, VA: AASA.
- Eagly, A. H. & Carli, L. L. (2003). Finding gender advantage and disadvantage: Systematic research integration is the solution. *The Leadership Quarterly*, 14(6), 851-859.
- Eagly, A.H. & Karau, S.J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573-598.
- Elsesser, K.M. & Lever, J. (2011). Does gender bias against female leaders persist? Quantitative and qualitative data from a large-scale survey. *Human Relations*, 64(12), 1555-1578.
- Fernandez, R. (2007). *Women, work, and culture*. NBER Working Paper No. 12888. *National Bureau of Economic Research*.
- Glass, T. & Franceschini, L. (2007). *The state of the American School Superintendency: A mid-decade study*. Arlington, VA: American Association of School Administrators.
- Glass, T. (2000). The shrinking applicant pool. *Education Week*, 20(10), 49-51.

- Grove, R., & Montgomery, P. (1999). Women and the leadership paradigm: Bridging the gender gap. Retrieved on 9.17.2015 from <http://nationalforum.com/Electronic%20Journal%20Volumes/Grove,%20Roslin%20Women%20and%20the%20Leadership%20Paradigm%20Bridging%20the%20Gender%20Gap.pdf>
- Gupton, S.L., & Slick, G.A. (1996). Highly successful women administrators: The inside stories of how they got there. Thousand Oaks, CA: Corwin.
- Harris, S., Lowery, S., Hopson, M., & Marshall, R. (2004). Superintendent perceptions of motivators and inhibitors for the superintendency. *Planning and Changing*, 35(91-2 108-119
- Helgesen, S. (1990). *The Female advantage: Women's ways of leadership*. New York: Doubleday.
- Hewlett, S.A. Leader-Chivée, L., Sumberg, K., Fredman, C. and Ho, C.(2012). *Sponsor Effect*. Center for Talent Innovation, U.K..
- Hewlett, S. A., Peraino, K., Sherbin, L., & Sumberg, K. (2010). *The sponsor effect: Breaking through the last glass ceiling*: Harvard Business Review.
- Hill, M. S., & Ragland, J. C. (1995). *Women as educational leaders: Opening windows, pushing ceilings*. Thousand Oaks, CA: Corwin Press.
<http://0-search.proquest.com.library.dowling.edu/docview/287974946?accountid=10549>
- Hunter, T. (2012). A comparison of male and female Assistant Superintendents and their descriptions of internal barriers, external barriers, motivators, stressors, and discriminatory acts they would encounter on the route to the Superintendency. (Doctoral dissertation). Retrieved from Dowling College database.
- Jones, E. & Montenegro, X. (1982). *Climbing the career ladder: A research study of women in school administration*. Washington, DC: American Association of School Administrators.
- Kayes, P. E. (2006). New Paradigms for Diversifying Faculty and Staff in Higher Education: Uncovering Cultural Biases in the Search and Hiring Process. *Multicultural Education*, 14(2), 65-69.
- Kayes, P. E. (2006). New Paradigms for Diversifying Faculty and Staff in Higher Education: Uncovering Cultural Biases in the Search and Hiring Process. *Multicultural Education*, 14(2), 65-69.
- Kim, Y. & Brunner, C. (2009). School administrators' career mobility to the Superintendency: Gender differences in career development. *Journal of Educational Administration*, 47, 75-107.

- Kowalski, T. & Brunner, C. (2011). The School Superintendent: Roles, challenges, and issues. In F.W. English (Eds.) *The SAGE Handbook of Educational Leadership: Advances in Theory, Research, and Practice* (2nd ed). Thousand Oaks, CA: SAGE.
- Kowalski, T. J., McCord, R., Petersen, G., Young, I. P., & Ellerson, N. (2011). *The American school superintendent: 2010 decennial study*. Lanham, MA: American Association of School Administrators and Rowman & Littlefield Education.
- Laramore, C. (2010). Motivating factors in applying for the superintendency (Doctoral dissertation). Retrieved from <http://hdl.handle.net/10057/3653>
- Logan, J.P. (1998). School leadership of the 90's and beyond: A window of opportunity for some educators. *Advancing Women in Leadership*, Retrieved from www.advancingwomen.com/awl/summer98/LOGAN.html
- MacTavish, N. (2010). *Mentors, sponsors, and networks: Women superintendents in Washington state*. Seattle University). *ProQuest Dissertations and Theses*, 132. Retrieved from <http://0-search.proquest.com.library.dowling.edu/docview/748283294?accountid=10549>
- Marshall, S.A. (1986). Women reach for the top spot. *School Administrator*, 43(10), 10-13.
- Meyerson, D. E. & Fletcher, J.K. (2000) "A modest manifesto for shattering the glass ceiling." *Harvard Business Review* 78.1: 126-136.
- Mitchell, D. E. (Ed.). (2006). *New foundations for knowledge in educational administration, policy, and politics: Science and sensationalism* (pp. 129-134). Mahwah, NJ: Lawrence Erlbaum Associates.
- Moorosi, P. (2010), South African women principals' career path: Understanding the gender gap in secondary school management, educational management administration and leadership, 38(5): 547-562.
- National Center for Education Statistics (NCES) (2003) *Schools and staffing public use data set*.
- Noddings, N. (1984). *Caring: A feminine approach to ethics and moral education*. Berkeley CA: University of California Press.
- Noddings, N. (2005). Identifying and responding to needs in education. *Cambridge Journal of Education*, 35, 147-159.
- Oritz, F. (1980, April). Career change and mobility for minorities and women in school administration. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. Retrieved from ERIC database. (ED186979)

- Ottino, K. L. (2009). *Diminished aspiration: Women central office administrators and the superintendency* (Order No. 3352809). Available from ProQuest Dissertations & Theses Full Text: The Humanities and Social Sciences Collection. (304955844). Retrieved from <http://0-search.proquest.com.library.dowling.edu/docview/304955844?accountid=10549>
- Owens, L. & Ennis, C. (2005). The ethic of care in teaching: An overview of supportive literature. *Quest*, 57, 392-425.
- Patterson, J.L., Goens, G.A., and Reed, D.E. (2009). *Resilient leadership for turbulent times: A guide to thriving in the face of adversity*.
- Patton, M. & McMahon, M. (2006). *Career development and systems theory: Connecting theory and practice*. Sense Publishers: The Netherlands.
- Poll, C. (1978). *No room at the top: A study of the social processes that contribute to the underrepresentation of women on the administrative levels of the New York City school system*. City University of New York. *ProQuest Dissertations and Theses*, 340. Retrieved from <http://0-search.proquest.com.library.dowling.edu/docview/302914758?accountid=10549>
- Rivera, L. (2015, May 30). Guess Who Doesn't Fit In at Work. *New York Times*, p. 3. Retrieved June 5, 2015, from http://www.nytimes.com/2015/05/31/opinion/sunday/guess-who-doesnt-fit-in-at-work.html?_r=0
- Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3), 523-534.
- Sergiovanni, T. (1992). *Moral leadership: Getting to the heart of school improvement*. San Francisco CA: Jossey Bass.
- Sergiovanni, T. (1996). *Leadership for the schoolhouse: How is it different? Why is it important?* San Francisco CA: Jossey Bass.
- Shakeshaft, C. (2011). Wild Patience. In F.W. English (Eds.) *The SAGE Handbook of Educational Leadership: Advances in Theory, Research, and Practice* (2nd ed). Thousand Oaks, CA: SAGE.
- Shakeshaft, C. S. (1979). *Dissertation research on women in educational administration: A synthesis of findings and paradigm for future research*. Texas A&M University.
- Sherrill, A. (2010). Women in management: Analysis of female managers' representation, characteristics, and pay. GAO-10-892R, Sep 20, 2010. Retrieved from: <http://www.gao.gov/products/GAO-10-892R>

- Shore, L. M., Coyle-Shapiro, J. A.-M., Chen, X.-P., & Tetrick, L. E. (2009). Social exchange in work settings: Content, process, and mixed models. *Management and Organization Review*, 5: 289–302. Doi:10.1111/j.1740-8784.200900158.x
- Skrla, L., Reyes, P., & Scheurich, J. J. (2000). Sexism, silence and solutions: Women superintendents speak up and speak out. *Educational Administration Quarterly*, 36(1), 44-75.
- Tyack, D. & Strober, M. (1981). Jobs and gender: A history of the structuring of educational employment by sex. In P.A. Schmuck, W.W. Charters, Jr., & R.O. Carlson (Eds.), *Educational policy and management, sex differentials* (pp. 131 – 152). New York: Academic Press.
- Vogel, B. E. (1985). Economic agendas and sex-typing in teaching. *Contemporary Education*, 57(1), 16-21.
- Weatherly, S. G. (2011). Examining women superintendents' perceptions of the importance of types of mentoring functions. Baylor University. ProQuest Dissertations and Theses, 111. Retrieved from <http://search.proquest.com.library.dowling.edu/docview/873461079?accountid=10549>
- Wickham, D. M. (2007). Female superintendents: Perceived barriers and successful strategies used to attain the superintendency in California. University of the Pacific. ProQuest Dissertations and Theses, 99. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.6905&rep=rep1&type=pdf>
- Wolverton, M. & Macdonald, R. (2001, November). Women in the superintendency: Barking up the wrong chain of command? Paper presented at the Annual Meeting of the University Council for Educational Administration, Cincinnati, Ohio.
- Zachry, C. A. R. (2009). Breaking the glass ceiling from the top in what ways do women county superintendents support and encourage woman in educational leadership. University of California, Davis. ProQuest Dissertations and Theses, 138. Retrieved from <http://search.proquestcom/docview/304849643>

Mission and Scope, Copyright, Privacy, Ethics, Upcoming Themes, Author Guidelines, Submissions, Publication Rates & Publication Timeline

The *AASA Journal of Scholarship and Practice* is a refereed, blind-reviewed, quarterly journal with a focus on research and evidence-based practice that advance the profession of education administration.

Mission and Scope

The **mission** of the Journal is to provide peer-reviewed, user-friendly, and methodologically sound research that practicing school and district administrations can use to take action and that higher education faculty can use to prepare future school and district administrators. The Journal publishes accepted manuscripts in the following categories: (1) Evidence-based Practice, (2) Original Research, (3) Research-informed Commentary, and (4) Book Reviews.

The **scope** for submissions focus on the intersection of five factors of school and district administration: (a) administrators, (b) teachers, (c) students, (d) subject matter, and (e) settings. The Journal encourages submissions that focus on the intersection of factors a-e. The Journal discourages submissions that focus only on personal reflections and opinions.

Copyright

Articles published by AASA, The School Superintendents Association (AASA) in the *AASA Journal of Scholarship and Practice* fall under the Creative Commons Attribution-Non-Commercial-NoDerivs 3.0 license policy (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). Please refer to the policy for rules about republishing, distribution, etc. In most cases our readers can copy, post, and distribute articles that appear in the *AASA Journal of Scholarship and Practice*, but the works must be attributed to the author(s) and the *AASA Journal of Scholarship and Practice*. Works can only be distributed for non-commercial/non-monetary purposes. Alteration to the appearance or content of any articles used is not allowed. Readers who are unsure whether their intended uses might violate the policy should get permission from the author or the editor of the *AASA Journal of Scholarship and Practice*.

Authors please note: By submitting a manuscript the author/s acknowledge that the submitted manuscript is not under review by any other publisher or society, and the manuscript represents original work completed by the authors and not previously published as per professional ethics based on APA guidelines, most recent edition. By submitting a manuscript, authors agree to transfer without charge the following rights to AASA, its publications, and especially the *AASA Journal of Scholarship and Practice* upon acceptance of the manuscript. The *AASA Journal of Scholarship and Practice* is indexed by several services and is also a member of the Directory of Open Access Journals. This means there is worldwide access to all content. Authors must agree to first worldwide serial publication rights and the right for the *AASA Journal of Scholarship and Practice* and AASA to grant permissions for use of works as the editors judge appropriate for the redistribution, repackaging, and/or marketing of all works and any metadata associated with the works in professional indexing and reference services. Any revenues received by AASA and the *AASA Journal of Scholarship and Practice* from redistribution are used to support the continued marketing, publication, and distribution of articles.

Privacy

The names and e-mail addresses entered in this journal site will be used exclusively for the stated purposes of this journal and will not be made available for any other purpose or to any other party. Please note that the journal is available, via the Internet at no cost, to audiences around the world. Authors' names and e-mail addresses are posted for each article. Authors who agree to have their manuscripts published in the *AASA Journal of Scholarship and Practice* agree to have their names and e-mail addresses posted on their articles for public viewing.

Ethics

The *AASA Journal of Scholarship and Practice* uses a double-blind peer-review process to maintain scientific integrity of its published materials. Peer-reviewed articles are one hallmark of the scientific method and the *AASA Journal of Scholarship and Practice* believes in the importance of maintaining the integrity of the scientific process in order to bring high quality literature to the education leadership community. We expect our authors to follow the same ethical guidelines. We refer readers to the latest edition of the APA Style Guide to review the ethical expectations for publication in a scholarly journal.

Upcoming Themes and Topics of Interest

Below are themes and areas of interest for publication cycles.

1. Governance, Funding, and Control of Public Education
2. Federal Education Policy and the Future of Public Education
3. Federal, State, and Local Governmental Relationships
4. Teacher Quality (e.g., hiring, assessment, evaluation, development, and compensation of teachers)
5. School Administrator Quality (e.g., hiring, preparation, assessment, evaluation, development, and compensation of principals and other school administrators)
6. Data and Information Systems (for both summative and formative evaluative purposes)
7. Charter Schools and Other Alternatives to Public Schools
8. Turning Around Low-Performing Schools and Districts
9. Large scale assessment policy and programs
10. Curriculum and instruction
11. School reform policies
12. Financial Issues

Submissions

Length of manuscripts should be as follows: Research and evidence-based practice articles between 2,800 and 4,800 words; commentaries between 1,600 and 3,800 words; book and media reviews between 400 and 800 words. Articles, commentaries, book and media reviews, citations and references are to follow the *Publication Manual of the American Psychological Association*, latest edition. Permission to use previously copyrighted materials is the responsibility of the author, not the *AASA Journal of Scholarship and Practice*.

Potential contributors should include in a cover sheet that contains (a) the title of the article, (b) contributor's name, (c) terminal degree, (d) academic rank, (e) department and affiliation (for inclusion on the title page and in the author note), (f) address, (g) telephone and fax numbers, and (h) e-mail address. Authors must also provide a 120-word abstract that conforms to APA style and a 40-word biographical sketch. The contributor must indicate whether the submission is to be considered original research, evidence-based practice article, commentary, or book or media review. The type of submission must be indicated on the cover sheet in order to be considered. Articles are to be submitted to the editor by e-mail as an electronic attachment in Microsoft Word.

Acceptance Rates

The *AASA Journal of Scholarship and Practice* maintains of record of acceptance rates for each of the quarterly issues published annually. The percentage of acceptance rates since 2010 is as follows:

- 2011: 16%
- 2012: 22%
- 2013: 15%
- 2014: 20%
- 2015: 22%

Book Review Guidelines

Book review guidelines should adhere to the author guidelines as found above. The format of the book review is to include the following:

- Full title of book
- Author
- City, state: publisher, year; page; price
- Name and affiliation of reviewer
- Contact information for reviewer: address, country, zip or postal code, e-mail address,
- telephone and fax
- Date of submission

Publication Timeline

Issue	Deadline to Submit Articles	Notification to Authors of Editorial Review Board Decisions	To AASA for Formatting and Editing	Issue Available on AASA website
Spring	October 1	January 1	February 15	April 1
Summer	February 1	April 1	May 15	July 1
Fall	May 1	July 1	August 15	October 1
Winter	August 1	October 1	November 15	January 15

Additional Information

Contributors will be notified of editorial board decisions within eight weeks of receipt of papers at the editorial office. Articles to be returned must be accompanied by a postage-paid, self-addressed envelope.

The *AASA Journal of Scholarship and Practice* reserves the right to make minor editorial changes without seeking approval from contributors.

Materials published in the *AASA Journal of Scholarship and Practice* do not constitute endorsement of the content or conclusions presented.

The Journal is listed in Cabell's Directory of Publishing Opportunities. Articles are also archived in the ERIC collection.

Editor

Kenneth Mitchell, EdD

AASA Journal of Scholarship and Practice

Submit articles electronically: kenneth.mitchell@mville.edu

To contact by postal mail:

Dr. Ken Mitchell
Associate Professor
School of Education
Manhattanville College
2900 Purchase Street
Purchase, NY 10577

AASA Resources

- ❖ Learn about **AASA's books program** where new titles and special discounts are available to AASA members. The AASA publications catalog may be downloaded at www.aasa.org/books.aspx.

- ❖ **Join AASA** and discover a number of resources reserved exclusively for members. Visit www.aasa.org/Join.aspx. Questions? Contact C.J. Reid at creid@aasa.org.

- ❖ **Upcoming AASA Events**
 - ✓ 2016 National Conference, Feb. 11-13, 2016, Phoenix, Ariz., Phoenix Convention Center. For direct, easy access about the conference, go to www.aasa.org/NCE
 - ✓ 2016 AASA & MASA Women's Leadership Conference, Feb. 26, 2015, Southern Oaks House & Gardens, Hattiesburg, MS
 - ✓ 2016 AASA, NJASA, FEA Women's Leadership Conference, March 9, 2016, FEA Conference Center, Monroe, NJ
 - ✓ 2016 AASA Advocacy Conference, July 12-14, Marriott Metro Center, Washington, DC
 - ✓ 2017 National Conference, March 2-4, 2017, New Orleans, LA, Ernest N. Morial Convention Center