# Table of Contents

**Board of Editors.** .............................................................. 2

**Sponsorship** ...................................................................... 3

**Research Articles**

A Better Leveled Playing Field for Assessing Satisfactory Job Performance of Superintendents on the Basis of High-Stakes Testing Outcomes. ................................. 4  
*I. Philip Young, PhD; Edward P. Cox, EdD; and David G. Buckman, MEd*

The Influence of Inclusion on the Academic Performance of General Education Students on the New Jersey Assessment of Skills and Knowledge in Grades 6, 7, and 8 ..................... 16  
*Christie M. Robinson, EdD and Gerard Babo, EdD*

The Principal Evaluation Process and Its Relationship to Student Achievement ...................... 34  
*Marcia McMahon, MS; Michelle L. Peters, EdD; and Gary Schumacher, PhD*

**Book Reviews**

*The Principal: Three Keys to Maximizing Impact* by Michael Fullan ................................. 49  
Reviewed by D. Cameron Hauseman, MEd

*Automate This: How Algorithms Took Over Our Markets, Our Jobs and the World*  
by Christopher Steiner ................................................................. 52  
Reviewed by Art Stellar, PhD

**Mission and Scope, Copyright, Privacy, Ethics, Upcoming Themes,**  
**Author Guidelines & Publication Timeline.** .......................... 56

**AASA Resources** ................................................................. 60
Board of Editors

_AASA Journal of Scholarship and Practice_
2012-2014

Editor
Christopher H. Tienken, _Seton Hall University_

Associate Editors
Barbara Dean, _AASA, The School Superintendents Association_
Ken Mitchell, Superintendent, _South Orangetown Central School District, NY_

Editorial Review Board
Albert T. Azinger, _Illinois State University_
Sidney Brown, _Auburn University, Montgomery_
Brad Colwell, _Bowling Green University_
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_
John Decman, _University of Houston, Clear Lake_
David Dunaway, _University of North Carolina, Charlotte_
Daniel Gutmore, _Seton Hall University_
Gregory Hauser, _Roosevelt University, Chicago_
Jane Irons, _Lamar University_
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_
Sue Mutchler, _Texas Women’s University_
Margaret Orr, _Bank Street College_
David J. Parks, _Virginia Polytechnic Institute and State University_
George E. Pawlas, _University of Central Florida_
Dereck H. Rhoads, _Beaufort County School District_
Paul M. Terry, _University of South Florida_
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

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 would like to offer special thanks to Seton Hall University, in particular the College of Education and Human Services.

We appreciate this unique relationship between research and practice, 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 Seton Hall University, the *AASA Journal of Scholarship and Practice* would not be possible.
A Better Leveled Playing Field for Assessing Satisfactory Job Performance of Superintendents on the Basis of High-Stakes Testing Outcomes

I. Phillip Young, PhD
Professor and Chair
Department of Educational Leadership and Policies
University of South Carolina
Columbia, SC

Edward P. Cox, EdD
Associate Professor
Department of Educational Leadership and Policies
University of South Carolina
Columbia, SC

David G. Buckman, MEd
Graduate Student
Department of Educational Leadership and Policies
University of South Carolina
Columbia, SC

Abstract

To assess satisfactory job performance of superintendents on the basis of school districts’ high-stakes testing outcomes, existing teacher models were reviewed and critiqued as potential options for retrofit. For these models, specific problems were identified relative to the choice of referent groups. An alternate referent group (statewide population), that addressed these shortfalls, was proposed and tested via regression procedure. The results indicate that 72% of the variance in student outcomes is beyond their control because of certain student and organizational characteristics. We demonstrated how this information can be used to identify like-type comparisons yielding a better leveled playing field.

Key Words
superintendent job satisfaction, student achievement, high stakes testing
Since the advent of high-stakes testing outcomes (HSTOs) for students, a standard empirical metric has emerged for assessing the job performance of both teachers and school administrators in the public school setting (Nichols & Berliner, 2008). Although the assessment of high-stakes testing outcomes were mandated by federal legislation (NCLB, n.d.) for all states, the means and methods for assessing HSTOs were delegated to individual states and vested with their departments of education. As a result, the metric for HSTOs is state specific and comparisons among states are inappropriate.

However, within a state, a common metric exists and has been used as a benchmark criterion for several types of comparisons. HSTOs have been used to construct report cards (Cupertino Unified School District n.d.) illuminating similarities and differences within a school district as well as among school districts. Most concerning, HSTOs are beginning to be used for making important human resource decisions for employees (Baker et al. 2010), including public school superintendents.

The two major human resource decisions made on the basis of HSTOs for superintendents are pay increases and continuation of employment (Young & Fawcett, 2013). Although these human resource decisions are related, they differ in an important way often overlooked in the published literature. That is, the benchmark criteria used to define job performance on the basis of HSTOs is different.

For pay decisions, exceptional HSTOs for a public school district are used. On the other hand, for continuation of employment, satisfactory HSTOs for a public school district are utilized. Between these two types of human resource decisions (pay increase vs. employment continuation) and between these different benchmark criteria (exceptional HSTOs vs. satisfactory HSTOs), we focus on continuation of employment decisions for public school superintendents as defined by a satisfactory level of HSTOs at the district level.

We do so in several ways. First, we review existing models developed for teachers and point out specific flaws and voids of these models if retrofitted for superintendents. Second, we address these shortcomings by proposing a new referent source involving “like-type” comparisons and by using a descriptive multiple regression approach to define a satisfactory level of HSTOs within a particular state.

**Literature Review**

At present, within the published literature, several models exist for defining satisfactory job performance on the basis of HSTOs for teachers. These basic models have been identified by Baker et al. (2010), as follows: (a) status comparison model, (b) change comparison model, (c) growth comparison model, and (d) value added model. Collectively, these models are similar in approach but have been operationalized differently. They are similar in that all models define satisfactory job performance on the basis of HSTOs by using a referent group as a source of comparison.

However, these models differ as to how the referent group is constituted (see below). Consequently, satisfactory job performance of superintendents as defined by HSTOs is model specific because different referent groups are used.
Existing Models for Defining Satisfactory Job Performance

Status comparison model (SCM)
Purported peer superintendents would be used to define the referent group for defining satisfactory job performance on the basis of HSTOs within the SCM. To the extent that the school district’s HSTOs for a specific superintendent either equals or exceeds those district level HSTOs obtained by the peer group of superintendents severing as the referent source, satisfactory job performance on the basis of HSTOs is so defined and would warrant a continuing contract for the targeted superintendent. However, the SCM suffers from several shortfalls when retrofitted for defining satisfactory job performance of superintendents on the basis of HSTOs. These shortfalls are rooted both in the way that the peer superintendents are chosen and the actual number of superintendents used to constitute a peer referent group. In general, the peer group of superintendents is chosen from a policy as opposed to an empirical perspective and is comprised of a small number, i.e., usually less than 10 (Young & Fawcett, 2013). As a result, the school districts of peer superintendents can differ in many ways likely influencing HSTOs (to be discussed).

This scenario is based on a small N, and, as such, would yield a large standard error of estimate (SEE). Also, because of a small N, a single change in membership of the peer group, could alter the baseline used to assess satisfactory job performance when defined on the basis of HSTOs.

Change comparison model (CCM)
The referent source, used by the CCM to define satisfactory job performance of superintendents relative to HSTOs, is their school district’s prior year’s HSTOs. That is, a school district’s past performance on HSTOs serves as the benchmark criterion for assessing current satisfactory job performance on HSTOs. Although the CCM controls more precisely for district variables than the SCM, likely influencing HSTOs by focusing only on a single school district, the CCM has another major shortfall when assessing the job performance of superintendents on the basis of HSTOs: criterion for hiring newly appointed superintendents.

According to the CCM for superintendents, previous HSTOs under their tutelage should be used as the benchmark criterion for assessing satisfactory job performance. However, newly appointed superintendents, by definition, fail to have any history of HSTOs within the school district. Therefore, they do not have the benchmark criterion which is to be used for assessing satisfactory job performance on the basis of HSTOs.

Growth comparison model (GCM)
Rather than focusing either on a small peer group (SCM) of purportedly similar school superintendents or on the prior year’s school district history of HSTOs (CCM) for superintendents, a different referent source is used by the GCM: individual student gains on HSTOs. That is, individual student gains are defined within the GCM from a student as opposed to a district level (see CCM) change.

Within the GCM, using students as the unit of analysis, the academic performance of individual students is tracked across academic years. The net within student differences on HSTOs are used to define satisfactory job performance of superintendents. The major advantage of the GCM over the CCM is that it includes only those students who have an instructional history within the particular school district. On the other hand, there is a
major void with the GCM. That is, no consideration is given to specific characteristics of students reported to influence HSTOs. These characteristics can either inflate or deflate the assessment of satisfactory job performance of superintendents when made on the basis of HSTOs.

Value added model (VAM)
In practice, the VAM is an extension of the GCM model in that it uses the same basic paradigm (i.e., gain scores for individual students) but it does so in a more precise way. The VAM considers certain student characteristics suggested, as well as found to, influence the HSTOs of students. These student characteristics include poverty conditions, minority status, and English language deficiencies.

Although the VAM represents an improvement over the GCM and has been noted by Baker et al. (2010) to be the best model to date for using HSTOs for defining satisfactory job performance on the basis of HSTOs, concerns remain over the use of growth scores. According to Koedel and Betts (2009), the VAM uses gain scores to define satisfactory job performance on the basis of HSTOs and fails to address either cellar or ceiling effects. More specifically, gains on HSTOs are far easier to come by in low performing school districts than in high performing school districts.

Advancements
Foremost, we propose a different referent group that can be used to define satisfactory job performance of superintendents when assessed, at least in part, on the basis of HSTOs. The referent group advocated within our study is the entire population of superintendents located within a particular state. Thus, we capitalize on the usage of all statewide data and not a subset of data to assess the job performance of public school superintendents as measured by HSTOs.

By using a statewide population of superintendents, we are able to address all issues and shortfalls associated with existing models (SCM, CCM, GCM, and VAM) reviewed. We are also able to provide a better leveled playing field for decision making. With respect to the SCM model using a small number of peer superintendents as the referent group, the entire statewide population of superintendents serves as the referent group in our study. As a result, subjectivity in the choice of a peer group is eliminated and a much smaller standard error of estimate will be obtained when making decisions about satisfactory job performance of superintendents on the basis of HSTOs.

Basic to the CCM is that satisfactory job performance of superintendents is assessed relative to their prior year’s HSTOs. However, for newly appointed superintendents, these data fail to exist. By using statewide superintendents as the referent source employed within the same academic year, concerns about prior year performance on HSTOs when defining satisfactory performance of superintendents on the basis of HSTOs is addressed.

A major shortcoming of the GCM is the failure to consider student characteristics found (Baker et al. 2013; Baker et al. 2010; Nichols & Berliner, 2008; Vang, 2008) to influence HSTOs when assessing satisfactory job performance of employees (e.g., either teachers or administrators). Although this void is addressed by the VAM, both the GCM and the VAM share a common practical, as well as a basic, flaw when using gain scores to define satisfactory job performance on the basis of HSTOs. That is, the failure to address adequately either cellar or ceiling effects.
whereby superintendents in underperforming school districts (i.e., cellar districts) are likely to post greater gains than superintendents in high performing school districts (i.e., ceiling effects).

Beyond addressing problems associated with existing models (SCM, CCM, GCM, and VAM), we expand current knowledge about using HSTOs for defining satisfactory performance of superintendents in another way. In light of Branch, Hanushek, and Rivkin’s (2013) statement “the fundamental challenge to measuring the impact of school leaders [on HSTOs] is separating their contributions from the many other factors that drive student achievement.” Based on Branch et al. challenge, we address student characteristics like the VAM but expand this repertoire to include organizational variables beyond the control of superintendents found to influence district level HSTOs.

Like the VAM, we include in our model student characteristics that are found to influence HSTOs (Baker et al. 2013; Baker et al. 2010; Nichols & Berliner, 2008; Vang, 2008) beyond the control of superintendents. Student characteristics included in our model are language differences resulting in English Language Learner (ELL) classification and student poverty as measured by free/reduced lunch entitlement. We used only these two student characteristics because recent research (Young & Fawcett, 2013) indicates when minority status, ELL classification, and poverty of students were considered simultaneously, via a regression analysis, only the latter two were found to account for significant variance in HSTOs at the school building level.

Organizational characteristics included in our study are school district enrollments and per pupil expenditures (PPE). An analysis of data for superintendents taking part in a national study (Kowalski et al. 2011) indicated that those in larger school districts have more specialists to support their instructional programs than those in smaller school districts. With respect to PPE, Verstegen and King (1998) reported an increase of .05 points on high stakes testing outcomes for each dollar increase spent on instruction.

To assess the utility of our proposed advancements with a statewide population of public school superintendents, we conducted an empirical study addressing the following research questions:

Research Question 1: Can a linear combination of student and/or organizational characteristics beyond the control of superintendents account for a substantial amount of systematic variance in school district level HSTOs?

Research Question 2: Can this information be used to level the playing field when assessing satisfactory job performance of superintendents on the basis of school district level HSTOs?

Methodology

The focal population in our study is all public school superintendents (N = 68) employed within a specific southeastern state (i.e., South Carolina) reporting complete information for all those variables considered in our study. We focused only on a specific state because Nichols and Berliner (2008) indicated instruments used to assess HSTOs are state specific. Our population was limited to 68 public school superintendents. Even though South Carolina has 79 public school districts, only 68 of these districts reported all required information to the State Department of Education.
Procedure
An archival database was used to collect all data in our study. This database is maintained by the South Carolina Department of Education and is available to the public. Contained within this database is information about biographical and experiential characteristics of superintendents, HSTOs for each public school district within the state, student characteristics pertaining to ELL, as well as to poverty, school district enrollments, and school district PPE (see Table 1).

Table 1
Descriptive Statistics for Superintendents and School Districts

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superintendents’ Sexa</td>
<td>.37</td>
<td>.49</td>
<td>68</td>
</tr>
<tr>
<td>Superintendents’ Educationb</td>
<td>.22</td>
<td>.42</td>
<td>68</td>
</tr>
<tr>
<td>Superintendents’ Years of Experience</td>
<td>5.00</td>
<td>4.28</td>
<td>68</td>
</tr>
<tr>
<td>School District’s FTE Enrollments</td>
<td>9743.97</td>
<td>1175.00</td>
<td>68</td>
</tr>
<tr>
<td>School Districts’ % of Poverty Students</td>
<td>64.52</td>
<td>15.25</td>
<td>68</td>
</tr>
<tr>
<td>School Districts’ % of ELL Students</td>
<td>4.49</td>
<td>4.11</td>
<td>68</td>
</tr>
<tr>
<td>School Districts’ Per Pupil Expenditure</td>
<td>9439.91</td>
<td>1515.00</td>
<td>68</td>
</tr>
<tr>
<td>School Districts’ Composite HSTOs</td>
<td>623.55</td>
<td>17.14</td>
<td>68</td>
</tr>
</tbody>
</table>

Notea: Males scored as “0” and females scored as “1.”
Noteb: Ph.Ds. scored “0” and others scored as “1.”

Variables of interest
School districts’ HSTOs were assessed by the Palmetto Assessment of State Standards (PASS) test. Specially measured by the PASS are English and math using a multiple choice format whereby scores can range potentially from a low of 300 to a high of 900. Importantly, results from this test are used to comply with both state and federal mandates concerning HSTOs.

Characteristics of students enrolled in public school districts served as predictor variables considered in our analysis. Based on recent research (Young & Fawcett, 2013), we considered only those student characteristics addressing language difficulties as defined by an ELL classification and students of poverty as defined by free/reduced meals. Student characteristics were operationalized on these variables according to their percentage of representation within each school district and could range from “0” to 100% (see Table 1).

Organizational variables noted within our review of the literature to impact HSTOs were school district enrollments (Kowalski et al. 2011) and PPE (Verstegen & King, 1998). School district enrollments serve as a proxy for additional instructional support personnel, while PPE serves as a proxy for financial contribution to the instructional program. These
organizational variables were defined by FTE student enrollments and by actual dollars amounts, respectively.

**Statistical Analysis**

To provide empirical data addressing our first research question, we conducted a descriptive multiple regression procedure because our data represents population parameters rather than sample estimates. Within this regression analysis, the overarching criterion/dependent variable is a school district’s composite HSTOs reflecting the joint contributions of math and English scores on the PASS. Reliability of this overarching composite score was assessed by Chronbach’s *Alpha* and was found to be .94.

District level HSTOs of students were regressed on student characteristics (i.e., % of ELL students and % of poverty students) and on organizational variables (i.e., FTE student enrollments and dollar amount for PPE) purported to influence HSTOs and which are beyond the control of superintendents. These predictor variables were entered within the regression equation using a simultaneous block-wise procedure. Results of this regression analysis addressing our first research question are found in Table 2.

Table 2

*Regression Model considering Student and Organizational Characteristics for HSTOs*

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>602.19</td>
</tr>
<tr>
<td></td>
<td>% of Poverty Students</td>
<td>-.93</td>
</tr>
<tr>
<td></td>
<td>% of ELL Students</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>FTE Student Enrollment</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>Per Pupil Expenditures</td>
<td>-.01</td>
</tr>
</tbody>
</table>

*Dependent Variable: District Level Composite HSTOs.*

Collectively, $R^2$ was found to be .72 (i.e., omnibus model SEE = 9.15) indicating that the vast majority of the systematic variance in school district level HSTOs can be accounted for by variables beyond the control of superintendents. Standardized regression coefficients (i.e., *Beta* $[\beta]$) reflect the relative importance of those variables beyond the control of superintendents and account for systematic variance associated with district level HSTOs. They are reported in Table 2, as follows: (a) % of poverty students $[\beta = .83]$, (b) % of ELL students $[\beta = .01]$, (c) FTE student enrollments $[\beta = -.07]$, and (d) school district PPE $[\beta = -.09]$. As can be observed in Table 2, the most influential variable, by far, is the percentage of poverty students enrolled in a school district where $\beta = -.83$, and the unstandardized regression coefficient equals $-93$ indicating that for each .93% decrease in students of poverty a 1.0% increase in HSTOs could be expected.
With respect to our second research question concerning if these data can be used to level the playing field when defining satisfactory job performance of superintendents on the basis of HSTOs, unstandardized regression coefficients (B) are used to create two different scenarios focusing on those variables beyond the control of superintendents (see Table 2).

In one scenario, a superintendent is depicted as being employed by a school district having favorable attributes on variables beyond their control. In the other scenario, a superintendent is depicted as being employed by a school district having less favorable attributes on these same variables. Degrees of favorability between these two scenarios are defined according to either a plus or minus 1 SD from the statewide population mean (see Table 1) on each predictor variable (see Table 2) with consideration being given to signage. More specifically, a favorable work environment for a superintendent was defined as having a low percentage of poverty students (i.e., -1 SD = 49%), a low percentage of ELL students (i.e., -1 SD = .38%), a high PPE expenditure (i.e., + 1 SD = $10,955), and a large FTE school district enrollment (i.e., + 1 SD = 10,919). In contrast to the favorable work attribute levels for superintendents, the parameter values for a less favorable work environment are as follows: (a) a higher percentage of poverty students (i.e., + 1 SD = 80%), (b) a higher percentage of ELL students (i.e., + 1 SD = 9%), (c) a lower PPE (i.e., -1 SD = $7,925), and (d) a smaller FTE school district enrollment (i.e., -1 SD = 8,769).

With respect to these different work environments, separate regressions were calculated using the same intercept value (i.e., 602.19), as well as all unstandardized regression coefficients reported in Table 2.

These separate regression analyses revealed that distinctly different benchmark criteria should be used when defining satisfactory job performance of superintendents on the basis of their school districts’ HSTOs if a level playing field is obtained (see research question 2). For the favorable work environment scenario, the benchmark criterion for defining satisfactory HSTOs is 566 points on the PASS test.

On the other hand, for the less favorable work environment scenario, the benchmark criterion is 535 points on the PASS test. Most importantly, the difference between these two benchmark criteria is 1.81 SD on HSTOs, and this difference is extremely important when defining a satisfactory level of job performance on the basis of HSTOs for superintendents.

**Conclusion**
Results from our study addressing the use of HSTOs for assessing satisfactory job performance of public school superintendents are timely because “policymakers throughout the country are increasingly embedding score-based approaches within the educational evaluation and accountability systems” (Amreim-Beardsley, Collins, Polasky, & Sloat, 2013, p. 5). In so doing, policymakers have focused largely only on the use of HSTOs as one of the many job facets or criteria to be considered within the total job evaluation process but have been silent with respect to how satisfactory or meritorious job performance can be defined on the basis of HSTOs.

To define a satisfactory level of job performance on the basis of HSTOs for superintendents, Baker et al. (2010) identified four potential models (SCM, CCM, GCM, and VAM) that have been used for teachers. In all of these models, a satisfactory level of job
performance on the basis of HSTOs is defined by different benchmarks involving a comparison referent source. Consequently, the definition of satisfactory job performance on the basis of HSTOs is model specific because different referent sources are used to establish benchmarks.

As such, we reviewed these models and identified specific methodological shortcomings, if they are retrofitted for public school superintendents, when defining their job performance on the basis of district level HSTOs.

These shortcomings include the following:
(a) the magnitude of the SEE for the SCM which involves a small number of comparisons,
(b) the inability of the CCM to rate newly appointed superintendents because they lack a prior history of HSTOs for students under their watch,
(c) the failure of GCM to address variables influencing HSTOs beyond their control, and
(d) the insensitivity of the VAM to address either cellar or ceiling effects by focusing only on gain scores.

To address all of these shortfalls within a single model, we proposed a different referent group that can be used to better assess a satisfactory level of job performance for public school superintendents on the basis of district level HSTOs. This referent group would utilize a statewide population involving all public school superintendents.

Because the SEE is a function of the number of comparisons used to establish a benchmark criterion and because the N for our statewide population will always be larger than the N used in the SCM, our calculated SEE will be much smaller. From an applied perspective within the field setting, a smaller SEE affords more confidence on any decisions made about satisfactory job performance of superintendents on the basis of HSTOs than a larger SEE, hence a better leveled playing field for decision-making.

Satisfactory job performance on the basis of HSTOs is defined within the CCM by a self-referent benchmark calculated on the basis of superintendents’ prior year’s district level HSTOs as compared to their current year’s district level HSTOs. The CCM assumes that all superintendents have a prior year history within their district. This assumption creates an unworkable situation in many school districts charged with using HSTOs to determine satisfactory job performance on the basis of HSTOs, especially given the high turnover rate for superintendents (Cooper, et al. 2000).

Our benchmark criterion for assessing satisfactory job performance on the basis of HSTOs makes no assumptions about a prior year history within a district but defines the benchmark criterion according to other superintendents employed in a “like-type” school district within the same academic year.

The referent group used by the GCM to assess satisfactory job performance of superintendents is student gains. The benchmark criterion is defined by a net difference using a repeated measures design yoking students across academic years. Failed to be considered by the GCM are any additional variables found to influence students HSTOs. Our proposed model addresses this omission by including student characteristics as well as organizational variables reported to influence HSTOs of students (see Table 2). Thus, we follow the lead of Branch et al. (2013) when they suggested “the fundamental
challenge to measuring the impact of school leaders [on HSTOs] is separating their contributions from the many other factors that drive student achievement.”

Although the VAM addresses, like our proposed model, those variables beyond the control of school administrators, as suggested by Branch et al., we propose a distinctly different referent group for defining the benchmark criteria used to define satisfactory job performance of superintendents on the basis of their districts’ HSTOs.

VAM uses a norm reference criterion where the benchmark criterion is net gains of students on HSTOs. This fails to address adequately either cellar or ceiling effects (Koedel & Betts, 2009) unique to a particular school district. More specifically, school districts posting a prior low (cellar effects) performance on HSTOs have a far greater growth opportunity to post a higher subsequent performance on HSTOs than a school district posting a prior higher (ceiling effect) performance due to a restricted growth range for improvement.

To address cellar and ceiling effects, our referent source is other superintendents employed in “like-type” school districts as defined by those variables found to influence HSTOs beyond their control. As such, we advocate using a criterion referent source to define benchmarks when defining satisfactory job performance of superintendents on the basis of HSTOs. This advancement, as noted in our Statistical Analysis section of this manuscript, illustrates how different benchmark criteria should be used to address cellar and ceiling effects for low and high performing school districts.

Clearly, our findings should be welcomed by superintendents for several reasons. First, our results indicate much of the variance in HSTOS at the school district level is beyond their control. Second, the playing field for assessing satisfactory job performance on the basis of HSTOs is at a better level by our model than retrofitted models (SCM, GCM, CCM, and VAM) because we use empirically defined like-type comparisons to define this benchmark criterion.

Although our study, like all studies, suffers from certain limitations, these important limitations are acknowledged. Foremost, we focused only on a single state using a specific test (PASS) for assessing HSTOs at the school district level. However, the basic methodology espoused and illustrated is generalized across states. Although we were able to account for 72% of the systematic variance in HSTOs beyond the control of superintendents, by no means should it be interpreted that the remaining 28% of the variance in HSTOs be attributed solely to superintendents because the inclusion of other variables beyond the control of superintendents may well account for additional systematic variance.
Author Biographies

I. Phillip Young is a professor and the chairman of the University of South Carolina Department of Educational Leadership and Policies in the College of Education. E-mail: ipyoung@mailbox.sc.edu

Edward Cox is an associate professor at the University of South Carolina College of Education, Department of Educational Leadership and Policies. E-mail: coxep@mailbox.sc.edu

David Buckman is a graduate student at the University of South Carolina Department of Educational Leadership and Policies in the College of Education.
References


The Influence of Inclusion on the Academic Performance of General Education Students on the New Jersey Assessment of Skills and Knowledge in Grades 6, 7, and 8

Christie M. Robinson, EdD
Director of Special Services
Little Silver Public Schools
Monmouth County, NJ

Gerard Babo, EdD
Assistant Professor
Department of Education Leadership, Management, and Policy
Seton Hall University
South Orange, NJ

Abstract

This study examined the influence of student demographic variables (i.e., SES, race, attendance, and gender) and the school variable of placement in an inclusion setting on the academic achievement of general education students in grades 6, 7, and 8 (n=1200) in an urban school district as measured by the 2010—2011 NJASK, the state’s annual standardized performance assessment. Analyses were conducted using hierarchical multiple regression models with results suggesting that placement in an inclusion classroom did have a statistically significant impact on the NJASK scores of non-disabled students in one of the two schools, implying variation of implementation at the school level. These results suggest that there are school-level factors at work in determining attainment and achievement in schools with similar levels of inclusivity. These results indicate that further research needs to be conducted in the area of inclusion to determine why inclusion might negatively impact the academic achievement of non-disabled students.

Key Words

special education, inclusion, general education, academic achievement
Introduction

The federally mandated Individuals with Disabilities Education Act (IDEA) requires that each district board of education provides a free and appropriate public education for all students with disabilities between the ages of 3-21 where students with disabilities are educated in the least restrictive environment.

Accordingly, students with disabilities are not exempt from meeting typical benchmarks and must be tested on grade level with their non-disabled peers with appropriate modifications to level the playing field. In many states already, and potentially many more to follow, scores rendered from mandated state assessments will be tied to a teacher and the school principal in determining an individual’s success and/or failure as an educational leader/professional.

As state standardized tests continue to be the primary indicator to measure student achievement, teacher quality, and principal effectiveness, school leaders will continue to struggle to raise scores to meet student achievement benchmarks.

Furthermore, as inclusive classrooms are formed to accommodate learners with special needs, where is the best academic placement for general education students in terms of them meeting academic performance targets on state mandated tests? Is it in a traditional classroom structure without the inclusion of students with special needs or an inclusive setting containing learners with special needs?

Without a solid research base to consider and their jobs as school principals at stake, many school leaders continue to question the efficacy of the inclusive environment, specifically with how it might influence general education student academic performance.

Review of Literature

An assortment of empirical studies document the benefits of education in an inclusive environment, both academically and socially, for students with disabilities. The results from many studies point out that children in inclusive programs generally do at least as well as special education students placed in specialized programs (Odom et al. 2011, Idol, 2006; Obrusnikova et al. 2003). Additional qualitative findings strongly support the practice of including students with special education challenges in general education programs as it provides opportunities for typical peer interaction and modeling of academic standards and rigor resulting in an enhancement of their individualized education program (Dyson, et al. 2004).

On the contrary, review of the empirical literature on the influence of the special education inclusive classroom on the academic achievement of students without disabilities have yielded a variety of mixed results and conclusions (Huber, Rosenfeld, & Fiorello, 2001; Kalambouka et al. 2007; Saint-Laurent et al. 2002; Dyson et al. 2004; Dessemontet & Bless, 2013). While some studies found that there was no significant difference in the progress of the low, average or high achieving general education students and concluded that there were no adverse effects on pupils without disabilities in mainstream schools, others found there was a negative impact on specific groups of general education students.

It is important to expand the literature review to incorporate and review relevant research addressing grouping and interactions.
of students in homogeneous or heterogeneous classes as inclusion creates a heterogeneous classroom of learning needs. A common argument among researchers is that within class ability grouping generally favors higher achieving students whereas low achieving students receive less appropriate instruction (Aydin & Tugal, 2005).

Additionally, advocates of homogeneous grouping by ability propose four basic rationales for doing so which include the following: 1) allows the teacher more efficient planning, 2) high ability students learn more than low ability students, 3) low ability students do not get frustrated by the progress of high ability students, 4) it is easier to teach therefore less discipline problems occur in homogeneous groups (Aydin & Tugal, 2005).

That said, researchers have found that grouping students by ability or performance has drawbacks which may offset any advantages especially when grouping students who are not in top tracks to second class instruction and depriving students of the examples and stimulation provided by heterogeneous classes (Slavin, 2008; Hong et al. 2011). The overall effect size of ability grouping, as per the results from studies conducted since 1962, suggest an absolutely no positive effect on student achievement (Aydin & Tugal, 2005; Mosteller, Light, and Sachs, 1996).

Additional criticisms of ability grouping include: 1) quality of learning as students in low-ability groups are often exposed to lower quality instruction, 2) achievement is generally lower and less rigorous, exposure to students with only low level skills, segregation, self-esteem and feelings of inferiority, 3) and in some cases delinquency and dropout (Burris et al. 2006, Aydin & Tugal, 2005).

Problem

One problem rests with the lack of empirical quantitative evidence explaining the influence of inclusion on the student achievement of general education students. In examining the body of empirical literature addressing this topic, there are a few caveats that cannot be overlooked.

First, the published literature is limited to small-scale class experiments with small sample sizes and it cannot be assumed that if these studies were replicated they would yield similar, comparable results.

Also, in examining the conclusions made by researchers, there is little to no commentary addressing inclusion at the school level, classroom level makeup of inclusive settings, nor the relationship between inclusion and achievement of students across the large scale (Dyson et al. 2004).

In other words, policymakers have continued to focus on the benefits of inclusion for students with disabilities and neglected to consider that the general education population of students is not homogeneous in ability; therefore, these policies and placements in an inclusion setting could influence their academic achievement.

This study attempted to yield additional insight into the effects of placement in an inclusive setting on the academic achievement of specific subgroups within the general education population, thus fostering new knowledge in determining the best placement for students within the general student population and aiding school leaders in making data based decisions about the structure and design of inclusive classrooms to help them reach annual benchmarks for student growth percentiles.
**Purpose and Question**

The purpose of this study was to determine whether placement in an inclusive setting would affect the academic achievement of general education students on the Language Arts Literacy and mathematics section of the New Jersey Assessment of Skills and Knowledge (NJ ASK), Grades 6, 7, and 8.

This study aimed to examine specific models including the independent variables of inclusive setting, non-inclusive setting, student attendance, and eligibility for free and reduced lunch that, paired with placement in an inclusive/non-inclusive setting, may result in an effect on the dependent variable of student achievement on the NJ ASK Grades 6-8.

The overarching research question for this study was: What is the influence of placement in the inclusive setting on the performance of non-disabled students in the area of language arts and mathematics as measured by the NJ ASK when controlling for student mutable variables at Grades 6, 7, and 8?

**Population and Sample**

The participants from this study were selected from an urban, lower middle class, PreK-12 school district located in central New Jersey. This Title I city district houses eight Pre-K-5 elementary schools, two Grade 6-8 middle schools, and one Grade 9-12 high school, with an approximate enrollment of about 6,000 students.

The sample population was limited to the two Grade 6-8 middle schools in the district. School A has approximately 710 students, 260 in Grade 6, 240 in Grade 7, and 210 in Grade 8 and is a diverse school with approximately 34% white, 26% African-American, 25.8% Hispanic, and approximately 15% Haitian, Portuguese and Polish. School B has approximately 700 students, 49 in Grade 5 (excluded for this study), 197 in Grade 6, 239 in Grade 7, and 216 in Grade 8 and continues to be diverse with approximately 21% white, 36% African-American, 21% Hispanic, 7% Polish, 5% Haitian, and 8% Portuguese.

Student data that met the following criteria were selected for the study: (1) each student in the sample had valid overall and cluster scores in language arts/literacy and mathematics on the NJ ASK, (2) each student in the sample completed both previous grade levels in the same district and school (as indicated by obtaining two years of NJ ASK scores 2009-2010, 2010-2011, (3) each student in the sample was in grades 6-8 during the time of the study, (4) each student was considered a general education student and not deemed eligible for special education services.

Student participants were assigned to classrooms, both inclusive and non-inclusive, prior to the onset of this investigation by school district administration.

Although this study was unable to control for class placement, pre-achievement scores could be used to get an overall achievement level for each group. Archival data were collected from student files. Achievement test scores were retrieved 2010-2011 via the district student management software package. Non-disabled students were coded by grade level based on placement in Comparison Group 1 (non-disabled students assigned to inclusive placements) or Comparison Group 2 (non-disabled students assigned to non-inclusive placements).

In this school district, an inclusive classroom was defined as an academic setting where general education students and students in special education learn academics (for this study specifically math and language arts) in
the same classroom environment. These students are taught in a classroom containing two certified teachers, one content area expert and one special education teacher. It is important to note that students in special education are not pulled from the general education setting for small group instruction at any time during “inclusion.” Both the general education students and special education students are exposed to the same curriculum and materials as peers placed in non-inclusive settings.

**Results**

We conducted a series of hierarchical multiple regression analyses to explore the scheduling of non-disabled students to inclusive or non-inclusive classrooms and what influence, if any, that might have on NJ ASK LAL and Math 2011 performance when controlling for student mutable variables.

Independent/predictor variables in the models included the primary variable of interest, classroom setting (inclusive and non-inclusive) along with the control variables gender, race, student attendance, student academic aptitude (as measured by past performance on standardized assessments), and SES (eligibility for free and reduced lunch) in an effort to determine the amount of variance accounted for on the dependent/predictor variable of student achievement as measured by the NJ ASK 6-8 LAL and math assessments.

The hierarchical multiple regression models incorporated a deliberate block entry model where the variable of interest, inclusive/non-inclusive setting, was entered in Step 1; student mutable variables were entered in Step 2; and student past academic performance was entered in Step 3. This was done for the purpose of better controlling for the influence of the predictor variables on the influence of the variable of interest.

Green (1991), as cited in Field (2009), recommends the following equation for determining minimum sample size for hierarchical multiple regression, $104 + k$, where “$k$” represents the number of predictor variables to be entered into the overall model. In this case, since there are 6 predictor variables, minimum sample size would be 110. Both regression analyses more than met this minimum standard with an “$n$” of 486 for School A and an “$n$” of 535 for School B.

The following sections will report and discuss the findings for both School A and B on the NJ ASK 2011 LAL and math standardized assessment results, respectively.

**School A**

Results for the Hierarchical Multiple Regression analysis for School A performance on the NJ ASK 2011 LAL appears in Table 1.
Table 1

Hierarchical Multiple Regression Analysis for School A NJ ASK 2011 Language Arts

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>R²</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC LA</td>
<td>-27.68</td>
<td>1.77</td>
<td>-.58***</td>
<td>.34</td>
<td>.34</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC LA</td>
<td>-27.49</td>
<td>1.76</td>
<td>-.56***</td>
<td>.37</td>
<td>.36</td>
</tr>
<tr>
<td>SES</td>
<td>-3.98</td>
<td>1.80</td>
<td>-.08*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>4.02</td>
<td>1.65</td>
<td>.09*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>-5.56</td>
<td>2.21</td>
<td>-.10**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>-.20</td>
<td>.16</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC LA</td>
<td>-15.66</td>
<td>1.52</td>
<td>-.33***</td>
<td>.61</td>
<td>.60</td>
</tr>
<tr>
<td>SES</td>
<td>-2.08</td>
<td>1.42</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>2.52</td>
<td>1.30</td>
<td>.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>-2.14</td>
<td>1.74</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>-.12</td>
<td>.13</td>
<td>-.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAL 2010</td>
<td>.54</td>
<td>.03</td>
<td>.55***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>102.03</td>
<td>6.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05; **p<.01; ***p<.001

When INC LA was entered alone in Step 1 of the regression analysis it significantly predicted LAL achievement, \( F(1, 484) = 245.75, p < .001 \), adjusted \( R^2 = .34 \), indicating that 34% of the variance in School A LAL performance can be explained by the Step 1 model. The negative beta (\( \beta = -.58 \)) for INC LA indicates that non-disabled students that are not in inclusion classroom are performing better than non-disabled students assigned to an inclusion classroom accounting for all of the variance in this model.

When the student mutable demographic variables SES, race, gender and attendance are entered in Step 2 of the model, prediction is only minimally improved as evidenced by an \( R^2 \) change = .03, \( F(4, 480) = 5.38, p < .001 \) with INC LA accounting for most of the variance, 31%, favoring non-disabled students that are not in inclusion classrooms. Additionally, the variables SES, race and gender are also found to be significant predictors in the Step 2 model. SES contributes .6% of the variance in favor of students not on free and reduced lunch, gender contributes .8% of the variance in favor of females and race predicts 1% of the variance. Although these variables are significant their predictive contributions to the overall model is minimal.

Step 3 added the variable LAL 2010 to the model in order to determine what amount of the variance in student academic performance could be explained by student past academic performance. When LAL 2010 is included, predicted student LAL performance as measured by the NJASK 2011 is significantly improved, \( R^2 \) change = .24, \( F(1, 479) = 295.377, p < .001 \) with LAL 2010 accounting for 30% of the variance in this model. Step 3 identifies the strongest predictive model (\( F(6,479) = 123.479; p <.001 \)) accounting for 60% of the variance in student performance on the LAL 2011, as is indicated by the adjusted \( R^2 \).
The Step 3 model indicates that both INC LA and LAL 2010 are significant predictors whereas student mutable demographic factors are not found to be significant contributors to the Step 3 model. INC LA contributes 11% of variance to the model in favor of non-disabled students that are not in inclusion classrooms and student past academic performance as accounted for by LAL 2010 scores contributes 30% of the variance to student LAL 2011 performance.

Table 2

Hierarchical Multiple Regression Analysis for School A NJ ASK 2011 Mathematics

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>R²</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC Math</td>
<td>-48.31</td>
<td>2.36</td>
<td>-.68***</td>
<td>.46</td>
<td>.46</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC Math</td>
<td>-46.80</td>
<td>2.34</td>
<td>-.66***</td>
<td>.49</td>
<td>.48</td>
</tr>
<tr>
<td>SES</td>
<td>-1.65</td>
<td>2.37</td>
<td>-.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-6.35</td>
<td>2.17</td>
<td>-10**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>-8.74</td>
<td>2.90</td>
<td>-10**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>-.41</td>
<td>.22</td>
<td>-.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC Math</td>
<td>-22.85</td>
<td>2.12</td>
<td>-.32***</td>
<td>.71</td>
<td>.70</td>
</tr>
<tr>
<td>SES</td>
<td>-60</td>
<td>1.80</td>
<td>-.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-3.59</td>
<td>1.67</td>
<td>-.05*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>-74</td>
<td>2.25</td>
<td>-.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>-15</td>
<td>.17</td>
<td>-.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math 2010</td>
<td>57</td>
<td>.03</td>
<td>.59***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>101.72</td>
<td>7.74</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05; **p<.01; ***p<.001

Table 2 displays the hierarchical multiple regression analysis on NJASK Math 2011 student assessment results for School A. When INC Math is entered alone in Step 1 of the regression analysis it significantly predicted Math achievement, F(1, 464) = 419.56, p < .001, adjusted R² = .46, indicating that 46% of the variance in School A Math performance can be explained by the Step 1 model. The negative beta (β = -.68) for INC Math indicates that non-disabled students that are not in inclusion classroom are performing better than non-disabled students assigned to an inclusion classroom accounting for all of the variance in this model.

When the student mutable demographic variables SES, race, gender and attendance are entered in Step 2 of the model, prediction is only minimally improved as evidenced by an R² change = .03, F(4, 484) = 5.99, p > .001 with INC Math accounting for most of the variance, 44%, favoring non-disabled students that are not in inclusion classrooms. Additionally, the variables gender and race are also found to be significant predictors in the Step 2 model. Gender contributes 1% of the variance in favor of male students and race predicts 1% of the variance. Although these variables are significant their predictive contributions to the overall model are minimal.
Step 3 added the variable Math 2010 to the model in order to determine what amount of the variance in student academic performance could be explained by student past academic performance in math.

When Math 2010 is included, predicted student math performance as measured by the NJASK 2011 is significantly improved, \( R^2 \) change = .22, \( F(1, 479) = 349.14, p > .001 \) with Math 2010 accounting for 35% of the variance in this model.

Step 3 identifies the strongest predictive model (\( F(6,479) = 190.70; p < .001 \)) accounting for 70% of the variance in student performance on the Math 2011, as is indicated by the adjusted \( R^2 \).

The Step 3 model indicates that both INC Math and Math 2010 are significant predictors whereas gender is found to be the only student mutable demographic factor to be a significant contributor to the Step 3 model accounting for .3% of the variance in favor of males. INC Math contributes 10% of variance to the model in favor of non-disabled students that are not in inclusion classroom and student past academic performance, as accounted for by Math 2010 scores, contributes 35% of the variance to student Math 2011 performance.

School B
Results of the Hierarchical Multiple Regression analysis for School B performance on the NJ ASK 2011 LAL appears in Table 3
Table 3

Hierarchical Multiple Regression Analysis for School B NJ ASK 2011 Language Arts

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>R²</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC LA</td>
<td>-15.75</td>
<td>2.434</td>
<td>-.27***</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC LA</td>
<td>-13.78</td>
<td>2.38</td>
<td>-.24***</td>
<td>.14</td>
<td>.13</td>
</tr>
<tr>
<td>SES</td>
<td>-3.83</td>
<td>1.89</td>
<td>-.09*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>6.11</td>
<td>1.78</td>
<td>.14**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>-7.84</td>
<td>1.97</td>
<td>-.17***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>.188</td>
<td>.20</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC LA</td>
<td>-1.30</td>
<td>1.86</td>
<td>-.02</td>
<td>.53</td>
<td>.53</td>
</tr>
<tr>
<td>SES</td>
<td>-1.24</td>
<td>1.40</td>
<td>-.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1.86</td>
<td>1.33</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>-4.10</td>
<td>1.47</td>
<td>-.09**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>-.17</td>
<td>.15</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAL 2010</td>
<td>.645</td>
<td>.03</td>
<td>.69***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>72.64</td>
<td>6.41</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05; **p<.01; ***p<.001

When INC LA was entered alone in Step 1 of the regression analysis it significantly predicted LAL achievement in School B, $F(1, 534) = 41.885, p < .001$, adjusted $R^2 = .07$, indicating that 7% of the variance in School B LAL performance can be explained by the Step 1 model. The negative beta ($β = -.27$) for INC LA indicates that non-disabled students that are not in inclusion classroom are performing better than non-disabled students assigned to an inclusion classroom accounting for all of the variance in this model.

When the student mutable demographic variables SES, race, gender and attendance are entered in Step 2 of the model, prediction is improved as evidenced by an $R^2$ change $= .07$, $F(4, 529) = 10.08, p < .001$ with INC LA accounting for 6% of the variance in the model favoring non-disabled students that are not in inclusion classrooms. Additionally, the variables SES, gender and race are also found to be significant predictors in the Step 2 model.

SES contributes .8% of the variance favoring students not on free and reduced lunch, gender contributes 2% of the variance in favor of female students and race predicts 3% of the variance.

Step 3 added the variable LAL 2010 to the model in order to determine what amount of the variance in student academic performance could be explained by student past academic performance. When LAL 2010 is included, predicted student LAL performance as measured by the NJASK 2011 is significantly improved, $R^2$ change $= .39, F(1, 528) = 443.53, p < .001$ with LAL 2010 accounting for 48% of the variance in this model. Step 3 identifies the strongest predictive model ($F (6,528) = 99.96; p <.001$) accounting for 53% of the variance in student performance on the LAL 2011, as is indicated by the adjusted $R^2$.

The Step 3 model indicates that both race and LAL 2010 are significant predictors.
whereas the variable INC LA is not found to be a significant predictor in Step 3, contradictory to the results found for School A. Additionally, with the exception of race, student mutable demographic factors are not found to be significant contributors to the Step 3 model. In Step 3, race contributes .8% of variance to the model and student past academic performance as accounted for by LAL 2010 scores contributes 48% of the variance to student scores on the NJ ASK LAL 2011. Most, if not all, of the variance in LAL 2011 student performance at School B can be explained by LAL 2010 performance. Regardless of whether a student is scheduled in an inclusive or non-inclusive classroom does not seem to have an influence on student LAL performance at School B, which is contradictory to the findings at School A.

Table 4

Hierarchical Multiple Regression Analysis for School B NJ ASK 2011 Mathematics

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>R²</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC Math</td>
<td>-34.37</td>
<td>3.92</td>
<td>-.36***</td>
<td>.13</td>
<td>.13</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC Math</td>
<td>-29.89</td>
<td>3.87</td>
<td>-.31***</td>
<td>.18</td>
<td>.17</td>
</tr>
<tr>
<td>SES</td>
<td>-6.12</td>
<td>2.99</td>
<td>-.08*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-2.05</td>
<td>2.68</td>
<td>-.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>-14.34</td>
<td>2.91</td>
<td>-.20***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>-.25</td>
<td>.30</td>
<td>-.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC Math</td>
<td>-1.60</td>
<td>2.60</td>
<td>-.02</td>
<td>.69</td>
<td>.68</td>
</tr>
<tr>
<td>SES</td>
<td>.58</td>
<td>1.89</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1.40</td>
<td>1.67</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>-2.72</td>
<td>1.85</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>.03</td>
<td>.19</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math 2010</td>
<td>.796</td>
<td>.03</td>
<td>.81***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>45.06</td>
<td>6.87</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05; **p<.01; ***p<.001

Table 4 displays the hierarchical multiple regression analysis on NJASK Math 2011 student assessment results for School B. When INC Math is entered alone in Step 1 of the regression analysis it significantly predicts math achievement, \(F(1, 533) = 77.02, p < .001\), adjusted \(R^2 = .13\), indicating that 13% of the variance in School B math performance can be explained by the Step 1 model. The negative beta (\(\beta = -.36\)) for INC Math indicates that non-disabled students that are not in inclusion classroom are performing better than non-disabled students assigned to an inclusion classroom accounting for all of the variance in this model.

When the student mutable demographic variables SES, race, gender and attendance are entered in Step 2 of the model, prediction is improved as evidenced by an \(R^2\) change = .06,
\[ F(4, 529) = 8.91, \ p > .001 \] with INC Math accounting for 10% of the variance favoring non-disabled students that are not in inclusion classrooms. Additionally, the variables SES and race are also found to be significant predictors in the Step 2 model. SES contributes .6% of the variance in favor of students not on free and reduced lunch and race predicts 4% of the variance to the Step 2 model.

Step 3 added the variable Math 2010 to the model in order to determine what amount of the variance in student academic performance could be explained by student past academic performance in math. When Math 2010 is included, predicted student math performance as measured by the NJASK 2011 is significantly improved, \( R^2 \) change = .50, \( F(1, 528) = 843.97, \ p > .001 \) with Math 2010 accounting for 67% of the variance in this model. Step 3 identifies the strongest predictive model (\( F(6, 528) = 191.34; \ p < .001 \)) accounting for 68% of the variance in student performance on the Math 2011, as is indicated by the adjusted \( R^2 \).

The Step 3 model indicates that the variable Math 2010 is the sole, significant predictor variable for NJ ASK Math 2011 student performance. Other variables found to be significant in previous Steps of this model (INC Math, SES, Race) failed to remain significant predictors when Math 2010 performance was entered into the model. As was the case with student performance on the NJ ASK LAL 2011 at School B, variance in Math 2011 student performance can best be explained by previous student performance in this subject area. At School B, whether a student is scheduled in an inclusive or non-inclusive classroom does not influence their respective performance on the NJ ASK Math 2011 assessment.

Due to the dichotomous hierarchical regression model results by school further analyses using comparative quantitative statistics was warranted. A Factorial ANCOVA was performed in order to compare group performance on both the NJ ASK LAL 2011 and Math 2011 based on classroom inclusion status while controlling for student past performance to determine if interaction between the schools and student inclusion status was significant.

Table 5 shows that a significant interaction does exist between school designation and classroom inclusion status on student performance on the NJ ASK LAL 2011 (\( n = 1,021 \)) while controlling for student past performance on NJ ASK LAL 2010, \( F(1, 1016) = 27.76, \ p < .001, \) partial eta \( ^2 = .03 \). Based on this finding, 3% of the variance in student performance on the NJ ASK LAL 2011 can be explained based on what school a student who is scheduled in an inclusive or non-inclusive classroom attends.
Table 5

Factorial ANCOVA for NJ ASK LAL 2011 as a Function of School and Student Inclusionary Status

<table>
<thead>
<tr>
<th>Variable and source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAL 2010</td>
<td>1</td>
<td>178564.22</td>
<td>810.46</td>
<td>.000</td>
<td>.44</td>
</tr>
<tr>
<td>INC LAL</td>
<td>1</td>
<td>11573.77</td>
<td>52.53</td>
<td>.000</td>
<td>.05</td>
</tr>
<tr>
<td>School Code</td>
<td>1</td>
<td>52.15</td>
<td>.237</td>
<td>.627</td>
<td>.00</td>
</tr>
<tr>
<td>INC LAL*School Code</td>
<td>1</td>
<td>6115.67</td>
<td>27.76</td>
<td>.000</td>
<td>.03</td>
</tr>
<tr>
<td>Error</td>
<td>1016</td>
<td>220.34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 clearly illustrates that general education students assigned to inclusive classrooms in School A perform significantly lower than their counterparts not assigned to inclusive classrooms as compared to the general education students assigned to inclusive classrooms in School B when compared to their counterparts on NJ ASK LAL 2011 performance.

Figure 1

Estimated Marginal Means of LAL 2011

Covariates appearing in the model are evaluated at the following values: LAL 2010 = 198.82
Table 6 displays a similar analysis based on student performance on the NJ ASK Math 2011 scores (n = 1,021), $F (1, 1016) = 14.37, p < .001$, partial $\eta^2 = .01$. Based on this finding, 1% of the variance in student performance on the NJ ASK Math 2011 can be explained based on what school a student who is scheduled in an inclusive or non-inclusive classroom attends.

Table 6

Factorial ANCOVA for NJ ASK Math 2011 as a Function of School and Student Inclusionary Status

<table>
<thead>
<tr>
<th>Variable and source</th>
<th>df</th>
<th>MS</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 2010</td>
<td>1</td>
<td>458692.68</td>
<td>1272.55</td>
<td>.000</td>
<td>.57</td>
</tr>
<tr>
<td>School Code</td>
<td>1</td>
<td>164.20</td>
<td>.456</td>
<td>.500</td>
<td>.00</td>
</tr>
<tr>
<td>INC Math</td>
<td>1</td>
<td>15678.79</td>
<td>43.50</td>
<td>.000</td>
<td>.04</td>
</tr>
<tr>
<td>School Code*INC Math</td>
<td>1</td>
<td>5163.74</td>
<td>14.37</td>
<td>.000</td>
<td>.01</td>
</tr>
<tr>
<td>Error</td>
<td>1016</td>
<td>360.45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 clearly illustrates that general education students assigned to inclusive classrooms in School A perform significantly lower than their counterparts not assigned to inclusive classrooms as compared to the general education students assigned to inclusive classrooms in School B when compared to their counterparts on NJ ASK 2011 math performance.
Conclusions, Discussion & Recommendations

Results of this study indicate that in School A, inclusion status was found to be a significant predictor of achievement for general education students placed in an inclusive environment. These students scored significantly lower than their general education peers who were not placed in inclusive classrooms. Conversely, in School B, inclusion status was not found to be a significant predictor of general education student achievement, which implies that placement within an inclusive environment did not influence student performance in this school. The Factorial ANCOVA figures 1 & 2 serve as the visual confirmation of these dichotomous results.

Consequently, we asked ourselves, what is happening in School A that is causing inclusion to have a significant impact on the achievement of general education students in inclusive classrooms? Synthesizing the results of this study and the current empirical evidence existing in both the area of inclusion and grouping, it is clear that there are school based effects and/or practices that are impacting the effectiveness of inclusion. One of these school based factors could be that School A violated what is considered to be the creation of inclusive classrooms that have a natural proportion of the overall student population.

However, without further empirical evidence to investigate this possible explanation it can only be speculated as to why the differences between the schools are actually occurring. Suffice to say, this study strongly supports the conjecture that individual school factors and/or practice do influence the effects of inclusion on general education students and as such, need to be considered by respective...
building administrators in order to optimize inclusionary practices.

One possibility speculated by us is that administrators in School A per chance created inclusive classrooms under the misconception that having mixed-ability groups would lower expectations and standards for the typical general education student. Empirical findings supports the notion that grouping students by ability or performance has drawbacks which may offset any advantages especially when grouping students who are not in top tracks to second class instruction and depriving students of the examples and stimulation provided by heterogenous classes (Slavin, 2008; Zaharias, Achillies & Cain, 1995). However, we have no clear cut evidence that this is the case in School.

Regardless, administrators designing inclusive classrooms must consider heterogeneously grouping students to avoid “tracking” and lower academic expectations by utilizing random assignment since research supports this practice (Zaharias, Achillies, & Cain, 1995). The nature of homogeneous grouping results in a lack of peer models (Burris et al, 2006). One of the most important aspects of creating an inclusive classroom is to emulate an environment of natural student proportion based on the theoretical context promoted by Bandura, which suggests that students learn a great deal simply by observing other students and how they conduct themselves both academically and socially (Ormrod, 2008).

In order to ensure that there is a natural proportion and appropriate academic and social peer models, it is recommended that school administrators ensure that the 10-30-60 rule is not violated when designing inclusive classrooms. Research by Burke & Sass (2008) found that in order for a classroom to be successful, there needs to be quality academic role models (60%) available, an even distribution of general education students in both inclusive and non-inclusive classrooms, and previous achievement needs to be considered and examined when making final decisions on the classrooms in which students are placed.

For example, in a class of 20 students, 2 should be special education (10%), 6 should be low to average scoring general education students (30%), and 12 should be average to above average scoring students. Any structure/deviation from this formula could render results where the academic performance of the general education students is significantly compromised.

The challenge for school administrators when creating an inclusion class can also be one of financial concern and limitation. In order to maintain a fewer number of students with disabilities in a classroom there becomes a need for more inclusion classrooms in the school. Consequently, more special education teachers are needed to provide support in additional inclusion classes. Although Special Education Code 6A:14-4.6 allows a maximum of eight special education students per inclusion classroom, this policy actually violates the 10-30-60 rule.

Administrators need to consider putting fewer students with disabilities in each inclusion class even though the law allows up to eight students. This will solidify the structural aspects of inclusion and support research based practice on classroom make up and ability grouping ensuring that the variable of inclusion has no influence on the academic achievement of non-disabled students placed in an inclusion classroom.
In summary, it is important that school administrators consider the relationship between the process variable inputs and student outcomes (Greenwald, Hedges & Lane, 1996).

Although an administrator has no control over predetermined factors such as socio-economic status, gender, or race, the factor of placement in an inclusive classroom can be controlled to render different and more positive academic outcomes. Curriculum, teacher quality and experience, heterogeneous grouping, natural proportion and peer modeling must all be considered when designing inclusive classrooms especially now that student test scores will be directly tied to the evaluation of both school principals and classroom teachers.

The implementation of new evaluation systems across the country with student performance serving as a primary criterion will continue to increase the pressure and importance of student performance on annual statewide standardized assessments. Ultimately, these results could mean the difference between contract renewal or non-renewal for both teachers and school administrators alike.

Author Biographies

Christie Robinson earned her doctorate from Seton Hall University in New Jersey. Her career includes varied experiences throughout Pre-K-12 education, special education and administration. She is currently the director of special services in the Little Silver School District, Monmouth County, NJ. E-mail: christieschutz@yahoo.com

Gerard Babo is an assistant professor at Seton Hall University Department of Education Leadership, Management and Policy. Previously, he spent 30 years in the New Jersey public schools as a teacher and administrator. His research interests include teacher and principal preparation and evaluation/supervision. E-mail: gerard.babo@shu.edu
References


The Principal Evaluation Process and Its Relationship to Student Achievement

Marcia McMahon, MS
Assistant Superintendent for Curriculum
La Marque Independent School District
Friendswood, Texas

Michelle L. Peters, EdD
Assistant Professor
Research and Applied Statistics
School of Education
University of Houston, Clear Lake
Houston, Texas

Gary Schumacher, PhD
Associate Professor
Educational Leadership
School of Education
University of Houston, Clear Lake
Houston, Texas

Abstract
This mixed methods study examined the relationship between school districts’ principal evaluation practices and their impact on student achievement in mathematics and reading. School achievement data were collected from a sample of 41 campuses representing 27 Texas school districts. Data were also solicited from a convenience sample of campus principals who were administered the Principal Evaluation Attributes, Feedback and Perception survey to measure their perceptions concerning the evaluation process in the areas of accountability, increased student achievement, professional growth, incentives for improvement, adherence to policy, and fostering school climate. Findings indicated no correlation between principal evaluation instruments and student achievement. Survey results indicated that evaluation processes within districts are perceived as merely a checklist for complying with district policy and have little impact on principal professional growth or student achievement.

Key Words
educational leadership, mixed methods, principal evaluation, student achievement
Introduction

Effective teaching, as measured by high levels of student achievement on state mandated standardized test, has received significant attention since the enactment of the *No Child Left Behind* Act in 2001. The Race to the Top (The White House, 2012) initiative began to provide significant financial incentives to school districts that incorporated student achievement data into their teacher appraisal systems. A report from the Center for Public Education (2013) found:

- 41 states require or recommend that teachers be evaluated using more than one measure of performance, which might include student test scores, classroom observations, student surveys, lesson plan reviews and teacher self-assessments;
- 38 states require teacher evaluations based on the teacher’s impact on student achievement; eight more states recommend this practice;
- 23 states require or recommend that student achievement indicators such as standardized test scores, student portfolios or learning goals for students make up at least half of a teacher’s evaluation.

Just as there are increased demands for accountability in teaching, there are increased demands for principals to document evidence of high levels of teaching performance and also to use student achievement in evaluating teacher performance. Their efforts in this regard can have a significant impact on student achievement on their campus. “No great accomplishments ever happened without great leadership, and transforming public education is no exception” (New Leaders, 2013). Research has shown that the school principal accounts for nearly 25% of the total impact on student achievement (Marzano, Walters, & McNulty, 2005, p. 10) and has been reported, on average, to impact student performance for approximately 500 students annually (New Leaders). Over the last 30 years, researchers examined the role principals play in shaping school improvement and developing a foundation for student learning (Clifford & Ross, 2012). An integral part of the landscape of contemporary educational reform has been focused on evaluating principal effectiveness. An effective principal is defined as one whose students make greater than average gains than similar students in other schools (Center for Public Education, 2012).

One influence on student achievement, as measured by state mandated state tests, is principal performance (Jacobson, 2011; Seashore Louis, Dretzke, & Wahlstrom; 2010) Some research findings support the quality of teaching and learning aligns to the campus principal’s expertise in shaping school culture and influencing people (Leithwood, Seashore Louis, Anderson, & Wahlstrom, 2004).

Thus, the principal as the instructional leader plays a role in indirectly influencing student academic success (Lazardious & Iordanides, 2011; Leithwood, Seashore Louis, Anderson, & Wahlstrom; Robinson, Lloyd, & Rowe, 2008; Wahlstrom, Seashore Louis, Leithwood, & Anderson, 2010). In an effort to understand how principal leadership influences student achievement Leithwood, Patten, and Jantzi (2009) found that principals had the
largest influence on student achievement in the areas of protecting instructional time and developing professional learning communities within the school. Dean (2012) concluded that a single principal, due to his or her ability to influence a large number of students could affect student achievement more than four times the influence of a highly effective teacher.

In an effort to ensure effective principals are leading our nation’s schools, states and local school districts may need to implement systems that align to high levels of student performance in the principal evaluation process. When principal evaluations are aligned with the Interstate School Leadership Consortium (ISLLC) Standards for School Leaders, student achievement should theoretically increase (Babo, 2010).

Current criticisms of principal evaluation systems are their lack of clear performance standards and rigor in both the design and attention to implementation (Goldring et al. 2009a, 2009b; Reeves, 2008). Additionally, few widely available principal evaluation instruments display psychometric rigor or make testing public so that validity and reliability can be examined (Goldring et al.).

In the state of Texas, the Texas Education Agency (TEA) requires all public school principals to be evaluated annually on 10 domains of performance, including effective management of teachers and consideration of student growth (TEA, 2013b). Although the Texas Administrative Code does not indicate what portion of the principal evaluation should be based on student performance, a TEA 2011 survey of school district evaluation instruments reported that only 30% of local school districts indicated their principal evaluation system included a student achievement component (TEA, 2011). Unless principal evaluation is rooted in student achievement, it will not yield best practices that support student success. It is a multi-faceted issue upon which many influences play.

Theoretical Framework

Theories related to effective leadership practices help to recognize how leaders may be influenced in ways that enhance their ability to impact student achievement. This can contribute to creating processes and procedures to measure how principals promote instruction and enhance the school culture and climate. Kahn, Wolfe, Quinn, Snoek, and Rosenthall (1964) indicated in the development of role theory that leadership practices of principals can be influenced by the expectation of others.

In development of evaluation instruments, the district defines the behaviors the principal should display when performing administrative duties. These roles are further developed in the evaluation process set by the principal’s supervisor with whom the principal interacts when performing his/her duties. These individuals develop beliefs, attitudes, and expectations about what the principal should or should not do as part of his/her role (Torrence, 2002).

While conducting this study, a review of Victor Vroom’s Expectancy Theory provided perspectives for understanding principal responses to the research question. An individual’s personal history is the basis of his or her level of expectancy and defines a belief that an act will lead to a particular outcome (Vroom, 1964). In order to achieve a high level of expectancy in the principal evaluation process, it is necessary for principals, just as it is for teachers, to grasp control over their own performance and outcomes (Schumacher, 2010).
Methods

Participants
A sample of 41 elementary schools across 27 school districts in Texas participated in this study. For federal accountability purposes, a campus that does not make AYP (Adequate Yearly Progress) in the same indicator (reading, mathematics, attendance rate, or graduation rate) for two or more consecutive years, must be identified for school improvement (TEA, 2004). Based upon longitudinal data, schools may progress through five stages of school improvement.

Utilizing the 2011 TAKS (Texas Assessment of Knowledge and Skills) scores, the elementary school at the highest level of school improvement was identified. Then using TEA’s 2011 Academic Excellence Indicator System (AEIS), a matched sample of all the elementary schools in Texas having a similar percentage of students who were Anglo, African American, Hispanic, economically disadvantaged, and limited English proficient were also selected for participation.

Additionally, a convenience sample of 27 campus principals who served at each respective campus during the 2010-11 academic year and remained in the same principal position (at the same campus) were identified and completed the survey. The selected sample provided a wide range of school structures (K-4th = 31.3%, K-5th = 56.3%, K-6th = 6.3%, 2nd-5th = 6.3%); district size (average number of schools = 28.2, range of schools = 1-73); principal longevity in a single position (average years = 5.5, range = 1-11 years); as well as total years of principal experience (average years = 9.8, range = 5-19 years).

The Texas Administrative Code, §150.1001, requires principals to be evaluated on an annual basis; 100% of the surveyed principals indicated that at a minimum they were evaluated on an annual basis. Data indicates the majority of principals were evaluated on a more frequent basis with principals reporting 37.5% as having semi-annual, 12.5% quarterly, and 6.3%, monthly evaluations. Principals also reported, 37.6%, were evaluated by a supervising director, 25.0%, by the Superintendent, 18.8%, by the Assistant Superintendent, and 6.3%, by the Deputy Superintendent.

Instrumentation

Principal evaluation rubric
Given that a rubric to evaluate/score locally developed district principal evaluations in terms of whether they were in compliance with the Texas Education Commissioner’s rules was nonexistent, one had to first be developed.

The development of the Principal Evaluation Instrument Rubric was a collaborative effort of seven professional educators from seven of the 20 Regional Education Service Centers (RESCs). The RESCs are quasi-governmental entities with the primary goal of enabling school districts within each region to operate more efficiently and economically by providing a number of services (Texas Legislative Budget Board, 2012).

All members of the expert panel have served as a campus principal and currently provide training and supervision of the principal certification process across the state of Texas. The expert panel first reviewed the Commissioner’s Rules concerning administrator appraisals in Texas and then decided on the 4-point rating scale (0 = Unsatisfactory, 3 = Exceeds Expectations), to rate the state’s nine performance domains, and 36 indicators for the rubric.
To determine the descriptor for each indicator, the members of the expert panel referenced two documents to support each statement, *Leading Learning Communities - Standards for What Principals Should Know and Be Able To Do* (NAESP, 2008) and the *New Leaders Principal Evaluation Rubric* (New Leaders, 2012).

Expert panelists provided feedback and adjustments were made to the content of the document. After each session of feedback, the researcher provided a revised version of the rubric to the panel for additional comments and ultimately, final approval.

The most challenging aspect in the development of the rubric was to determine the descriptors for the *Proficient* and *Exceeds Expectations* ratings. The *Proficient* scale rated the instrument's purpose in “evaluating the principals’ ability to develop a culture of collaboration for the campus.” This is essential for alignment to student achievement as research indicates the development of school culture and climate is critical for impacting student performance (Witziers, Bosker, & Krüger, 2003; Valentine & Prater, 2011).

For an instrument to rate *Exceeds Expectations*, the focus rests largely in seeking evidence to determine the “principal’s ability to build the capacity of others,” so campus culture and student achievement will be sustainable (Robinson, Lloyd, & Rowe, 2008).

Using the rubric, a team of experienced evaluators reviewed and scored each of the principal evaluation instruments. To reduce rater bias, the researcher altered the instruments to conceal the identity of the individual district. The inter-rater reliability was determined by calculating the interclass-correlation among the seven raters (ICC = .931).

**Principal survey**

This study also employed the *Principal Evaluation Attributes, Feedback and Perception* survey from a previous dissertation research study.

To develop the survey, Condon (2009) solicited the dissertation committee for input regarding the wording of the survey questions in relation to the survey constructs.

The question was posed: “Are these survey questions crafted so that they garner the required information about the constructs from the participants?” Questions developed for the survey could not be confusing, misleading, or biased, and peer-review of the survey questions helped to minimize these risks (Condon, 2009).

The survey was validated by piloting it with a group of respondents. The respondents were provided with the opportunity to share feedback about the integrity of the instrument and its fidelity to those constructs being measured. This was the ideal point at which to identify potential threats to internal validity, had they existed.

Finally, the survey was piloted with an independent group of respondents who were not in the survey participation group, but have like backgrounds. This validation process was valuable in determining the degree to which the survey functioned as designed, which would increase confidence in the findings.

The questionnaire consisted of 33 items broken into four subscales: (a) the extent of various methods used to evaluate the performance level of campus principals; (b) the frequency of principal evaluation; (c) the perception of the campus principal related to the evaluation process in the areas of accountability, increased student achievement,
professional growth, incentives for improvement, adherence to policy, and fostering school climate; and (d) demographics. Participants responded by either selecting the most appropriate response, by filling in the blank with the requested information, or by providing more in-depth answers to open-ended response items.

Student achievement
Mathematics and reading achievement was measured using scores obtained from the Texas Assessment of Knowledge and Skills (TAKS) test. The Texas Assessment of Knowledge and Skills assessment is a standardized test that was administered from 2003-2011 in Texas for grades 3-11 to assess students' attainment of reading, writing, math, science, and social studies skills required under Texas education standards. Developed and scored by Pearson Educational Measurement, with close supervision by TEA, TAKS assessments were designed to measure the extent to which a student has learned and is able to apply the defined knowledge and skills at each tested grade level (TEA, 2013a).

Data collection and analysis
Using a rubric developed by a panel of Texas RESC experts in the field of principal certification and evaluation, each of the 27 district principal evaluation forms were scored based on their alignment to the 10 performance domains and descriptors required in the Texas Administrative Code.

Scoring was completed by a team of central office school superintendents, assistant superintendents, and other education service center leadership staff, working independently for purposes of peer debriefing and inter-rater reliability. School achievement data, reported as the percentage of students within each of the 41 identified schools as passing the reading and mathematics TAKS examination, was retrieved from the TEA website reporting student accountability. Campus principals, who have served at the campuses since 2011, were surveyed to determine their perceptions on the quality of the district's evaluation process and its impact on student achievement at their school.

It should be noted that the self-report nature of the survey responses may serve as a limitation (Dipboye & dePontbriand, 1981). Survey indicators include methods used to evaluate principal performance, the frequency of evaluations, and impact on achieving objectives such as identification of needs for principal professional development.

Quantitative data was analyzed using percentages and Pearson product moment correlations, while an inductive coding process was used to analyze the qualitative survey data. To ensure validity of the qualitative findings, member checking and peer debriefing was employed.

Results
Student achievement
In order to examine the relationship between how each district evaluates its campus principals and the average student mathematics and reading achievement for each respective campus, each district’s principal evaluation instrument was first scored using the Principal Evaluation Instrument Rubric.

The reviewers of the expert panel assigned a composite score to each of the evaluation instruments according to their adherence to state requirements and rigor of assessment. To obtain one overall score for each evaluation, an average was taken of each
of the seven reviewer’s scores. A relationship was not found to exist between how the districts are evaluating their campus principals and their average school’s mathematics,  
\[ r = -0.212, p = .221, \text{ and reading achievement, } r = -0.271, p = .115. \]

**Principal evaluation, methods and objectives**

The principals included in this study provided responses to the *Principal Evaluation Attributes, Feedback, and Perception* survey.

In terms of the types of methods used to evaluate principals, participants reported that the three most widely utilized methods for determining a principal’s evaluation rating were:

1. the use of a check list/rating system (75.1%),
2. use of data (75.1%), and
3. supervisor observations (56.3%)

The three least used methods included:

1. perception feedback (18.8%),
2. anecdotal evidence (12.3%), and
3. peer review (0.0%).

Table 1 displays the various methods used for evaluating the participants and the degree to which they are applied by his or her respective district.

When the principals were asked “What effect, if any, has your experience with evaluation had on your beliefs about your performance?” 74.1% of the principals surveyed felt the evaluation process did not reflect their work. One principal commented, “My evaluations HAVE NOT reflected the amount of work I have done, nor the quality. I have to believe in myself and value the work I have done without receiving the compliments or adequate evaluations from my superiors.”
Table 1

*Evaluation Methods and Degree of Implementation*

<table>
<thead>
<tr>
<th>Evaluation Method</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Checklist/Rating</td>
<td>75.1</td>
</tr>
<tr>
<td>2. Data Based</td>
<td>75.1</td>
</tr>
<tr>
<td>3. Supervisor Observations</td>
<td>56.3</td>
</tr>
<tr>
<td>4. Narrative Self Evaluation</td>
<td>37.6</td>
</tr>
<tr>
<td>5. Narrative by Supervisor</td>
<td>31.3</td>
</tr>
<tr>
<td>6. Survey Data</td>
<td>25.0</td>
</tr>
<tr>
<td>7. Portfolio</td>
<td>18.8</td>
</tr>
<tr>
<td>8. Perception Feedback</td>
<td>18.8</td>
</tr>
<tr>
<td>9. Anecdotal Evidence</td>
<td>12.3</td>
</tr>
<tr>
<td>10. Peer Review</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Participants were also requested to identify what he or she thought were the primary objectives for performing principal evaluations. The three major objectives reported were to support the instructional program (68.8%); to increase student assessment scores (68.8%); and to document substandard performance (62.6%). Interestingly, the objectives least viewed by the principal included rewarding exemplary performance (31.3%); providing incentives for improvement (37.6%); and identifying personal professional growth (43.8%). Table 2 displays the various objectives for conducting observations and their degree of importance.
Table 2

*Evaluation Objective and Degree of Importance*

<table>
<thead>
<tr>
<th>Evaluation Objective</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Support the Instructional Program</td>
<td>68.8</td>
</tr>
<tr>
<td>2. Increase Assessment Scores</td>
<td>68.8</td>
</tr>
<tr>
<td>3. Document Substandard Performance</td>
<td>62.6</td>
</tr>
<tr>
<td>4. Satisfy District Requirements</td>
<td>62.5</td>
</tr>
<tr>
<td>5. Improve Pupil Achievement</td>
<td>56.3</td>
</tr>
<tr>
<td>6. Adhere to Policy and Procedures</td>
<td>50.1</td>
</tr>
<tr>
<td>7. Foster a Positive School Climate</td>
<td>50.0</td>
</tr>
<tr>
<td>8. Provide Principal Professional Growth</td>
<td>43.8</td>
</tr>
<tr>
<td>9. Identify Needs for Professional Development</td>
<td>43.8</td>
</tr>
<tr>
<td>10. Provide Incentive for Performance Improvement</td>
<td>37.6</td>
</tr>
<tr>
<td>11. Reward Exemplary Performance</td>
<td>31.3</td>
</tr>
</tbody>
</table>

When asked if the principal believed the evaluation process impacted pupil performance, approximately 75% of the participants felt the evaluation process did affect student achievement.

Interestingly the comments supporting student achievement could not determine the impact of the process regarding pupil performance.

Participants stated, “The evaluation certainly accounts for pupil performance, however, I’m not sure what the impact is on pupil performance,” and “Yes, principals don’t want bad evaluations so they will work hard to get the scores up.”

Others obtained similar results when studying principal evaluation methods and impact on student achievement (Brady, 2012,
Condon, 2009; Davis & Hensley, 1999; Johnson, 1989; Yavuz, 2010). All of the principal participants (100%) indicated the evaluation process did not develop their individual professional development.

Participants reported, “Test scores have influenced my professional development,” and principals may “like more feedback on possible professional growth opportunities that may benefit me.”

Additional studies confirmed evaluation processes are lacking in the establishment of prescriptive professional development to increase principal performance and impact student achievement (Johnson; Kimball & Pantsch, 2008; Yavuz).

Discussion
Findings appear to indicate that there is a lack of alignment between principal evaluation instruments and state performance standards. Survey data of campus principals suggests that they perceive that the evaluation process that is currently being used in their district lacks rigor and fidelity.

This is also consistent with findings from Brady (2012), which indicated that principals perceived that the locally developed evaluation instrument lacked alignment to state standards, was not researched-based, and had not been published or tested. Therefore, principals perceived the evaluation instruments lacked rigor and were unreliable.

Principals indicated the evaluation process does not provide them with adequate feedback in order to identify individual potential professional growth opportunities.

Additional studies confirmed evaluation processes lack in the identification of prescriptive professional development to improve principal performance and impact student achievement (Kimball & Pantsch, 2008; Yavuz, 2010).

The results from this study also suggest that the evaluation process does not measure the principal’s impact on student achievement.

Other studies concluded that principals consider the evaluation processes as lacking consistency and validity, and producing little impact on student achievement (Brady, 2012, Davis & Hensley, 1999; Yavuz).

Implications
To effectively evaluate principals, evaluation instruments should be research-based and aligned to state developed leadership standards or national standards such as the Interstate School Leadership Consortium (ISLLC) Standards for School Leaders. Canto and Stronge (2006) also found when evaluation instruments align to standards, principal behavior aligns to student achievement.

However, contradictory findings indicate principal standards in an evaluation instrument do not cover behaviors to ensure a rigorous curriculum or quality instruction, both a necessary component to enhance student achievement (Goldring et al. 2009a, 2009b). Emphasis should be placed on training of principal evaluators, inter-rater reliability practices, and on creating valid and reliable psychometrics measures, which provide fidelity across the district when conducting principal evaluation.

This supports the findings of Kimball and Pantsch (2008). Given the various professional demands placed on district evaluators, the districts struggled to complete principal evaluations consistently or with
fidelity. In some cases there is little confidence regarding the motives or intentions of the district office evaluators.

Principals reported evaluators typically spent little time on campus observing principals’ leadership behaviors and lacked training in effective evaluation techniques.

The development of principal evaluation systems should include an in-depth study of implementation practices, as well as instrumentation design, to produce fidelity and validity within the evaluation process. Thus, a connection between the evaluation instrumentation and processes may produce a greater impact on student achievement.

**Recommendations for Future Research**
Most research studies indicate the classroom teacher has the greatest impact on student achievement.

Given that the principal indirectly impacts teacher quality, an area of future study may include teacher perceptions regarding principal practices which support their efforts in the classroom which increase student achievement.

This information may provide possible areas to measure principal effectiveness aligned to student achievement.

Additional information regarding the preparation and implementation of principal evaluations processes may be helpful in developing evaluation methods aligned to state requirements.

Data collected from district individuals who conduct the evaluation process may indicate areas of improvement in both instrumentation and delivery, thus impacting student achievement.

**Author Biographies**

Marcia McMahon is the La Marque Independent School District Assistant Superintendent for Curriculum. Prior to this, she served for 5 years as the program coordinator for the principal certification program at the Region 4 Education Service Center. E-mail: marciamcmahon29@yahoo.com

Michelle Peters is currently a tenured-track assistant professor of research and applied statistics at the University of Houston, Clear Lake School of Education. She collaborates on research studies in teacher education, educational leadership, and program evaluation. E-mail: petersm@uhcl.edu

Gary Schumacher is an associate professor and teaches at the University of Houston, Clear Lake School of Education in the educational leadership department. His research interests include teacher appraisal systems, human resources practices, and school planning models. E-mail: schumacher@uhcl.edu
References


Using evidence gathered throughout his esteemed career, Michael Fullan argues in his most recent offering, titled *The Principal: Three Keys to Maximizing Impact* (2014), that the contemporary principalship is outmoded and a relic of a different time.

The author seeks to “reposition the role of the principal as overall instructional leader so that it maximizes the learning impact of all teachers, and in turn all students” (Fullan, 2014, p. 6). In three sections, the author presents an evidence-based framework for how principals should approach and operationalize their work. This framework contains the following three “keys:”

1. leading learning;
2. being a district and key player; and
3. becoming a change agent.

Embracing these keys and integrating them into their work will help principals maximize their impact at the school site.

Composed of six chapters, this book can be loosely grouped in three different sections. Containing the first two chapters, the first section is devoted to describing issues and concerns with how the principalship is currently constructed. These chapters are filled with a variety of evidence-based examples from the work of Fullan and other seminal scholars such as Kenneth Leithwood and Viviane Robinson. Using this approach, the author successfully demonstrates how the contemporary principalship is broken and hints at how the three keys found in the remaining chapters can solve issues that currently plague the position.

The second section includes the third to fifth chapters that describe each of the three keys. The final section explores how two pervasive contextual factors (the “unplanned digital revolution” and Common Core Curriculum Standards) will influence principals’ work in the near future.

The book concludes by reiterating how these three keys can cure issues that have been identified with the contemporary principalship, such as work intensification, and a misguided focus on faulty notions of instructional leadership.
The first key in Fullan’s (2014) framework is titled, *Leading Learning*. This key is rooted in instructional leadership, but emphasizes principals using their energy to develop the group rather than focusing on improving their own knowledge of direct instruction. It is described as the ability, “to lead the school’s teachers in a process of learning to improve their teaching while learning alongside them about what works and what doesn’t” (Fullan, 2014, p. 55).

The second key is titled, *Being a District and System Player* and is rooted in evidence that school leaders will improve their own performance, as well as that of their own school and others in their district by disseminating and mobilizing knowledge.

The third and final key of the framework explores principals’ work related to their role as a change leader. Aptly titled *Becoming a Change Agent*, this key is about developing the ability of principals to move their school in a positive direction when faced with difficult circumstances.

The author makes a compelling argument for the need to reposition both the contemporary principalship and notions of instructional leadership by presenting evidence indicating principals are overwhelmed, and struggle to deal with the many challenges they face on a daily basis.

Instructional leadership, at least how it is currently constructed and mandated, is not successful, and is a main culprit in much of the work intensification experienced by contemporary principals as it calls for unnecessary micromanaging of staff. While not quite revolutionary, an approach to instructional leadership that emphasizes developing the staff capacity would be a welcome change for many overworked principals. The evidence offered that suggests principals learn from developing connections with other schools is particularly convincing. Because of these well-reasoned arguments, the first two keys make perfect sense.

However, there is debate as to the utility and inclusion of the third key, especially in the wake of recent scholarship questioning why the notion of change is so often mentioned in the educational leadership literature. Rather than advocating change for change’s sake, the author positions a “change agent” as a principal who has the conviction necessary to stick to their core values and take calculated risks in the face of difficult circumstances.

The sense of urgency with which the author communicates problems associated with the current role and responsibilities of principals highlights a need to alter the contemporary principalship. More importantly, the author provides an evidence-based road map for principals to pursue these changes in their own practice. Though many of the examples used in this book will seem familiar to those who are acquainted with the author’s more recent work, his arguments surrounding the need to re-cast the principalship for the 21st Century are valid. As such, *The Principal: Three Keys to Maximizing Impact* holds appeal for policymakers and practitioners alike.
Reviewer Biography

D. Cameron Hauseman is a doctoral candidate studying educational administration at the University of Toronto Ontario Institute for Studies in Education in Toronto, Ontario, Canada. E-mail: cameron.hauseman@mail.utoronto.ca

Automate This: How Algorithms Took Over Our Markets, Our Jobs and the World
by Christopher Steiner

Reviewed by:
Art Stellar, PhD
Vice President
National Education Foundation and CyberLearning

Many in professional circles these days are talking about Big Data. The champions of Big Data envision it as a revolutionary advance almost without limits that is changing the world. Even skeptics have to admit that Big Data is making an impact in many fields. Like any new system or innovation, there are individuals opposed to the use of Big Data on philosophical grounds. Regardless of one’s stance, adding to one’s knowledge base by reading Automate This will upgrade the quality of the conversation.

Many of the mathematical formulas and concepts for algorithms have been in existence for centuries. However, as more and more data has been collected and as computers have added capacity, algorithms have come to life with additional utilitarian value.

One of the first fields to tap the potential of algorithms has been Wall Street. One of the first to make billions from the stock market using algorithms was Thomas Peterfly. He was not a stockbroker, but a computer programmer who wrote computer code. His first significant operation was a single IBM computer connected to a Nasdaq terminal. His system did not simply determine what to trade, so humans could execute the trade, his computer had hacked the system to eliminate the slow humans and make trades by itself. Minutes were saved in the process that later became seconds as others added their own computer systems which are now completing trades in hundreds of a second.

Fundamentally what Peterfly’s system did was to survey the market and issue bids and asks that captured the difference between the prevailing price at which buyers would buy and sellers would sell. This is called the spread. A difference of a mere 25 cents on a share times a thousand shares yielded a near riskless $250 profit.
A human being might have the same formula and data, but miss the trade because of the time it takes to find a partner or to articulate the trade. Humans often stop to say “hello” to fellow traders during these interactions. Computers skip all the niceties, never stop for lunch or a restroom break or think about anything else.

Another new book, *Flash Boys*, focuses on the latest version of Wall Street on steroids where algorithms powered by computers rule. The author makes the case that the average individual day trader investor simply can’t beat the machines and the algorithms built within their operating systems. This same author Michael Lewis wrote the classic *Money Ball*.

That book, later made into a movie of the same name starring Brad Pitt, describes how the losing Oakland Athletics baseball team with one of the lowest payrolls in baseball transformed itself by applying algorithms to various decisions. Baseball has a history of collecting data, however, much of the assembled numbers did not truly help general managers or coaches produce winning teams so management relied upon seasoned scouts with “gut feelings”—until the Oakland Athletics shocked the baseball world by going to the World Series due to mathematical algorithms.

More and more professional sports teams are incorporating advanced math and algorithms into their planning. For example, Bill Belichick coach of the New England Patriots of the National Football League modified his philosophy of punting on fourth down years ago based upon mathematical analysis. He discovered that under different scenarios the chances of “going for it” on fourth down were more favorable than punting. While the math has not always translated into a first down on the field, it has given him another tool in his coaching kit.

In 1989 IBM began working on a computer programming system with stated purpose of beating the best chess player in the world. In 1997 the chess grand master Garry Kasparov was beaten by IBM’s Deep Blue computer system. Chess is a perfect game for algorithms powered by computers. Kasparov could mentally review approximately three positions per second, while Deep Blue could examine 200 million per second. It was inevitable that as Deep Blue “learned” Kasparov’s moves, the machine’s algorithm would win.

By 2011 IBM had produced a computer labeled Watson that scored higher than all human contestants on Jeopardy! This television quiz show has random questions packed with humor, irony and other sorts of human idiosyncrasies. IBM stored 200 million pages of content running an algorithm with more than six million logic rules. The television show Jeopardy! remains on the air because watching humans is more entertaining than watching Watson who would nearly always win in the end.

Blackjack has already been conquered by algorithms as it is a relatively simple game of odds. Poker was thought to be safe from algorithms until the last few years. Thomas Sandholm, a computer science professor at Carnegie Mellon, has been working on an algorithm that now regularly beats the best professional poker players in most games. His algorithm still lags the very best poker professionals when it comes to no-limit poker with more than four top players in a game. That will come as the algorithms are learning faster than the humans.

There is a Las Vegas sports betting company that is building a super algorithm based upon a Wall Street model. It is called Midas and goes beyond any previous
mathematical formula. Of course, the house still takes a cut of all earnings which absolutely ensures a “no lose” situation for Midas.

Algorithms have moved way beyond games to life and death matters. The CIA has been using algorithms as highly skilled intelligence analysts in the battle against terrorism. The CIA did a study covering twenty years of scientist Bueno de Mesquita’s results with algorithms compared with their own experienced human analysts. The CIA found that his algorithms were right twice as often as expert CIA analysts. This is much like the television show Numbers where a college math genius helps the FBI solve crime. Law enforcement and the military are increasingly incorporating algorithms into their procedures.

Politicians have gone to the library to check out William Riker’s The Theory of Political Coalitions which applies game theory with algorithms to predict political races. Few have the background in calculus to understand or use the algorithms or to be able to make their own algorithms. The void is being filled as national political campaigns have been moving in this direction for some time.

Algorithms are no longer the sole domain of elite computer programmers or college mathematicians. Two years before Jeremy Linn burst on the professional basketball scene with the New York Knicks, Ed Wetland, a FedEx driver with a stat blog was predicting Linn would be a star in the National Basketball League. Lynn had played his college ball at Harvard and went undrafted and may never have had the chance to play except for an unusual string of injuries. Wetland used a statistical algorithm to forecast Linn’s rise to stardom which none of the basketball experts had anticipated.

Algorithms have even taken over match making with respect to love. Dating algorithms are used by over two dozen companies who claim to have a part in bringing together couples for whopping eight percent of all marriages! While this is a relatively new field to capitalize upon algorithms, experts suggest that the creators of dating algorithms will continue to perfect their systems to produce more compatible relationships.

Health care is rapidly emerging as “algorithm friendly” territory. This book describes a situation where an individual would have an all knowing “algorithm-doctor” who knew everything about that person’s medical history throughout the entire family, the current condition in every measurable way and who is capable of matching up any physical fluctuations with a database of possible diseases while prescribing treatments. There is no human doctor that can do this without making any mistakes and who is available 24/7. Small scale algorithms are presently being built and tested which will be combined with more sophisticated algorithms. The future health care systems will employ algorithms as default doctors. Computer programmers may need medical malpractice insurance.

Marketing has been using computerized algorithms for some time to analyze personality types and buying habits. Once an internet company has a bead on a person, ads for certain products and services are regularly streamed to that person whenever he/she signs on to the internet on a computer anywhere in the world. Companies that record customer conversations have those recordings reviewed by algorithms to squeeze out information useful for the organization.

Algorithms have been used to produce music for some time, although until recently the quality of an entire composition has been uneven. There are several music algorithms now capable of composing music that is indistinguishable from human production. In
fact, some music generated by algorithms has been entered in competitions for human composers with the algorithm music occasionally winning or doing well. In what is somewhat telling, judges who know that the music originates from an algorithm give it a lower score than when they do not know.

The examples in this book from the field of education are very simple matching algorithms such as lining students up with the right courses to graduate on time. The adoption of new technology by education usually follows other fields since there is no profit motive or stature to be gained. However, it is inevitable that at some point algorithms will become commonplace in education once incentives are discovered. While not included in this book, computer algorithms already assess student compositions and are incorporated in some instructional software. It will be interesting both to see how algorithms are incorporated into educational models and how educators react. While Automate This does not predict automated instruction, insightful readers can extrapolate the future impact the growth of algorithms will have on education.

Reviewer Biography

Art Stellar is vice-president of the National Education Foundation and CyberLearning with headquarters in McLean, Virginia. He served as a superintendent for 25 years for urban, rural and suburban communities in various states. A life member of AASA, he is widely published, the recipient of many honors and has held elected positions of professional organizations such as ASCD and the Horace Mann League. He is a three-time graduate of Ohio University.

Mission and Scope, Copyright, Privacy, Ethics, Upcoming Themes, Author Guidelines & 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 the 2012-2014 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.

**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

**Additional Information and Publication Timeline**

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 the Directory of Open Access Journals, and Cabell’s Directory of Publishing Opportunities. Articles are also archived in the ERIC collection.

Publication Schedule:

<table>
<thead>
<tr>
<th>Issue</th>
<th>Deadline to Submit Articles</th>
<th>Notification to Authors of Editorial Review Board Decisions</th>
<th>To AASA for Formatting and Editing</th>
<th>Issue Available on AASA website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>October 1</td>
<td>January 1</td>
<td>February 15</td>
<td>April 1</td>
</tr>
<tr>
<td>Summer</td>
<td>February 1</td>
<td>April 1</td>
<td>May 15</td>
<td>July 1</td>
</tr>
<tr>
<td>Fall</td>
<td>May 1</td>
<td>July 1</td>
<td>August 15</td>
<td>October 1</td>
</tr>
<tr>
<td>Winter</td>
<td>August 1</td>
<td>October 1</td>
<td>November 15</td>
<td>January 15</td>
</tr>
</tbody>
</table>

Submit articles to the editor electronically:
Christopher H. Tienken, EdD, Editor
AASA Journal of Scholarship and Practice
christopher.tienken@shu.edu

To contact the editor by postal mail:

Dr. Christopher Tienken
Assistant Professor
College of Education and Human Services
Department of Education Leadership, Management, and Policy
Seton Hall University
Jubilee Hall Room 405
400 South Orange Avenue
South Orange, NJ 07079
AASA Resources

- AASA has partnered with one of the nation’s premier leadership development organizations, The SUPES Academy, to create the **The AASA National Superintendent Certification Program**. The 2-year program, launched in 2013 in California, focuses on sharpening the skills that successful superintendents acknowledge are needed to thrive on the job.

  School system leaders who complete the program are eligible to receive 24 credits—a monetary value of approximately $25,000—toward a doctor of education degree administered by Nova Southeastern University.

  For more information go to AASA’s home page at www.aasa.org and click on the link to The AASA National Superintendent Certification Program or visit http://www.aasa.org/superintendent-certification.aspx or contact The AASA National Certification Program Offices, toll free, at 855-803-7547.

- 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**

  ✓ **2015 National Conference**, February 26-28, 2015, San Diego, Calif., San Diego Convention Center celebrating the 150th Anniversary of the founding of AASA
  Note: Call for Proposals is Now Open until May 26th