

Site Selection in School District Research: A Measure of Effectiveness Using Hierarchical Longitudinal Growth Models of Performance¹

Alex J. Bowers²
Teachers College, Columbia University

ABSTRACT:

School districts in the United States are an active area of study in education research as findings have shown that some districts find success in certain contexts, while others struggle. However, while significant resources are needed to study the complex day to day practices of district and central office staff, including researcher time and effort, to date the research domain has had few actionable methods for site selection to determine the extent to which a district should be considered unusually effective, while taking into account the cautions from 40 years of school effectiveness research. The present study analyzes all school districts in the state of Ohio (n=610) from 2005-06 through 2012-13 using a two-level hierarchical linear growth model to identify districts that significantly outperform or underperform the background and demographic variables in the Ohio Performance Index Score, that are outside the influence of the district administration. The aim of the study is to provide a framework for researchers, practitioners and policy-makers looking to select comparison and contrast sites for deep district effectiveness research. This study builds upon recent work by capturing the school level variance within districts to control for overall average district performance over an eight year period analyzing the entire population of districts. Fifteen outperforming districts are named as potential sites for site selection for in-depth qualitative studies of districts in comparison to similar districts that are at the norm for the state.

Keywords: School District, District Effectiveness Research (DER), Hierarchical Modeling, Longitudinal Studies, School Effectiveness Research (SER), Student Attendance, Rural Urban Differences, Suburban, Municipalities, Enrollment

INTRODUCTION:

The purpose of this study is to present and discuss issues of site selection for district research from all possible districts in a state focusing on how best to identify where in the distribution of “effectiveness” any one district may be in relation to all other districts in the population as a means to provide contrasts and

generalizable findings across districts that may be outperforming, underperforming or at the norm. To date, much of the literature on the effectiveness of school district central offices has focused on in-depth qualitative studies of the actions of central office personnel and constituent school leaders (Honig, 2009; Leithwood, 2010; Rorrer, Skrla, & Scheurich, 2008). However, few of these studies link findings to measures of district effects on student achievement (Trujillo, 2013). One reason is the issue of site selection and the justification of labeling a district as “effective” (Bowers, 2010). Recently, Bowers (2010) proposed and tested a new method for site selection in District Effectiveness Research (DER), examining all districts in Ohio over seven years using hierarchical growth modeling of test scores across multiple subjects and grade levels, identifying districts that significantly outperformed districts at the norm while controlling for background characteristics. Following the recommendations from the literature on school effects (Cuban, 1984; Heck & Moriyama, 2010; Trujillo, 2013) this work provided a means to identify significantly unusual districts from comprehensive longitudinal datasets. The purpose of the present study is to replicate, update and extend this model to take into account additional factors including the amount of variance between schools within a district with a more recent and longer dataset.

Past district site selection models have failed to consider the variance between schools within districts (Bowers, 2010). This is an important consideration given that district effects can be heavily influenced by individually high performing schools within a district. The aim of the present study is to test an extended longitudinal growth model of student achievement controlling for background variables and between-school variance within districts. Outperforming districts would thus have significantly higher growth in scores than comparison districts with similar demographics, and would do so more evenly across their schools. Districts can then be identified for possible inclusion in in-depth qualitative case studies of central office practices from the population of districts in a state, through a deeper understanding of where in the distribution of effectiveness a district lies prior to investing the significant resources and time required for deep qualitative district research.

BACKGROUND AND LITERATURE REVIEW

Over the last two decades, school leadership has been shown to have a positive and significant effect on student achievement (Robinson, Lloyd, & Rowe, 2008), especially when accounting for principal time on the job and organizational growth (Bowers & White, 2014; Coelli & Green, 2012; Hallinger & Heck, 2011; Heck & Hallinger, 2009; Heck & Moriyama, 2010; White & Bowers, 2011). However, less is known about the effects of school districts on student achievement, and the contributions of different aspects of school districts to achievement growth or decline across the

¹ This document is a pre-print of this manuscript, which is published in the journal *School Leadership & Management*. Recommended Citation:

Bowers, A.J. (2015). Site Selection in School District Research: A Measure of Effectiveness Using Hierarchical Longitudinal Growth Models of Performance. *School Leadership & Management*, 35(1) p. 39-61. <http://dx.doi.org/10.1080/13632434.2014.962500>

² Teachers College, Columbia University; Bowers@tc.edu; 525 W. 120th Street, New York, New York 10027.

ORCID: 0000-0002-5140-6428, ResearcherID: C-1557-2013

Note: This document was last updated on July 9, 2015.

district's schools (Rorrer, et al., 2008). This is an important consideration as schools in the U.S. are nested within districts, providing substantial local control to communities around the affairs of their schools, as well as an added layer of bureaucracy that is aimed to help districts find efficiencies and centralize resources and leadership for the community (Berkman & Plutzer, 2005; Cuban, 1984; Purkey & Smith, 1985). Starting in the mid-1980's, as the industry of school effectiveness research began to take hold in the education research community in an effort to find schools that were able to demonstrate success with all students no matter the context (Edmonds, 1979; Luyten, Visscher, & Witziers, 2005), questions arose as to the contribution of school districts to effective school practices (Bowers, 2010; Trujillo, 2013). As an example, Purkey and Smith (1985) noted that: "If the school is the arena for change, how can district policy stimulate the sort of bottom-up change necessary to the school improvement process?" (p.364). More recently, research on the district as the unit of analysis in educational reform has begun to ask deeper questions about the leadership practices in districts that appear to find success across schools in a community. As noted by Thompson, Sykes & Skrla (2008):

Recent case literature and related research suggest that classroom and school level coherence can be promoted by coherence in district level leadership—leadership spurred or supported by political pressures and implemented by harmonizing a variety of administrative controls with the creation of a professional community to bring about a pervasive unity of purpose (Thompson, Sykes, & Skrla, 2008) (pg 3).

Thus, recent research on the school district as the unit of analysis has come to five major conclusions about effective district practices that lead to overall system-wide school improvement (Leithwood, 2010; Levin, Datnow, & Carrier, 2012). First is a district-level focus on aligning the various moving parts across a district to an overall coherent instructional system-wide focus (Bowers, 2008; Elmore & Burney, 1997; Knapp, 2008; Skrla, McKenzie, Scheurich, & Dickerson, 2011), termed as an instructional regime (Cohen, Raudenbush, & Ball, 2003), a coherent systemic strategy (Smith & O'Day, 1991), a coherent instructional focus (Thompson, et al., 2008), and networked improvement communities (Bryk, Gomez, & Grunow, 2011). This primarily involves the second main conclusion of district research - selecting an instructional and curricular focus (Bowers, 2008; Elmore & Burney, 1997; Leithwood, 2010) and then third providing sustained and ongoing professional development around that instructional focus for not only teachers and staff (Firestone, Mangin, Martinez, & Polovsky, 2005; Floden et al., 1988; Hightower, Knapp, Marsh, & McLaughlin, 2002; Opfer, Henry, & Mashburn, 2008; Spillane & Thompson, 1997) but also individual school leaders (Honig, 2012; Mitgang, 2013; Turnbull, Riley, Arcaira, Anderson, & MacFarlane, 2013). Fourth, districts across this literature were able to pull together resources from multiple funding streams in service to the core instructional focus in an effort to provide instructional and curricular resources and time for teachers and leaders to build capacity, as well as a general alignment of human resource systems such as hiring and teacher placement in service to the central instructional focus of the district (Bowers, 2008; Darling-Hammond et al., 2005; Elmore & Burney, 1999). And fifth, districts across this literature appear to make effective use of data and evidence in continuous improvement processes in service to the overall instructional goals articulated throughout the system (Farley-Ripple & Cho, 2014; Honig &

Coburn, 2008; Honig & Venkateswaran, 2012; McLaughlin & Talbert, 2002; Wohlstetter, Datnow, & Park, 2008).

However, a major critique of much of the literature on effective district practices has questioned the overly specific focus on a single monolithic conception of "The District" as a single actor, or as an extension of the superintendent and school board (Honig, 2003; Seashore Louis, 2008). The school district, as an organizational concept is an organization made up of many different people acting in a variety of ways across many different social relationships (Daly & Finnigan, 2011; Daly, Moolenaar, Bolivar, & Burke, 2010; Honig, 2009), more so for large suburban and urban districts (O'Day & Quick, 2009). Taking a more individual actor view, much of this recent research has begun to dig deeply into understanding the day-to-day practices and relationships of central office staff, and what it looks like to run and organize systems of schools as they attempt reform efforts (Daly & Finnigan, 2012; Honig, 2008, 2009; Honig & Coburn, 2008), mapping the types of collaborative conversations that change over time as central office practitioners work to implement district reform efforts. Indeed, as a result of these efforts there have been multiple recent reports about best practices for district improvement (Levin, et al., 2012; Mitgang, 2013; Wallace Foundation, 2013), working to generalize the findings from the many in-depth qualitative district case studies conducted to date.

While this recent attention to district central office practices has helped to inform the literature as to what is known about how districts may go about instructional improvement efforts, almost all of the research to date on district effects have suffered from similar methodological issues as those of the early school effectiveness research (Bowers, 2010; Trujillo, 2013). As reviewed in Trujillo (2013), district-level research to date has mirrored the problems of school effectiveness research by a) focusing on small samples of districts and an overly narrow definition of effectiveness, such as mathematics and reading standardized test scores in a single grade; b) an almost exclusive focus on districts deemed to be unusually high performing; c) a focus on cross-sectional and "snapshot" research, rather than longitudinal change; and d) little to no information on school or classroom processes. In addition Trujillo (2013) noted the atheoretical nature of much of the research as a central issue across the domain:

One of the largest strengths of the district effectiveness literature is its relevance to issues of practice. Indeed, a primary impetus behind the research has been to isolate the concrete steps that central office leaders can take to cultivate the types of organizations that promote greater student success. Yet in concentrating chiefly on practical application, some of these studies may have repeated the same conceptual oversights as their school-level ancestors by steering too far from certain theoretical bodies of knowledge that can help explain some of their findings... The atheoretical nature of these studies is also seen in the widespread "lists" of effectiveness correlates that these studies often put forth. (p.441).

Trujillo (2013) notes that this focus on atheoretical lists of correlates does not allow researchers to interpret the findings of their studies more deeply as behaviors and processes deeply seated within the local context of the communities which the districts serve. As with the scarcity of theory in school effectiveness research (Scheerens, 2012), this would allow in district research the generation of complex theories of district effectiveness

grounded in the lived experiences of district and school personnel, and a means to provide useful theories for research and practice when attempting to identify how a district can move from less effective to more effective practices.

Nevertheless, as a practical matter, how does one go about finding districts to study for district effectiveness research? Researchers are looking to build theory, and practitioners and policy-makers are looking to find more “effective” districts and compare their practices to other districts deemed less effective in an effort to find useful “best practices” that may work in multiple district contexts to help improve teaching and learning across a system of schools. However, in-depth qualitative case studies of school districts are a significant undertaking, as the resources required are extensive. Deep qualitative research that addresses the district-as-monolithic-actor critique of past DER work (Honig, 2008) gets to the heart of the practices of central office staff and their interaction with the schools through long-term relationships between researchers and a district which then provides many opportunities to interview and observe all of the actors at the central office, all of the individual school leaders, a larger number of teachers, and a representative sample of students, parents, the school board and community members.

However, despite the large investment in time, effort and resources required for DER work, for almost all of the past research on districts, site selection has been haphazard, with the majority of studies having either no justification for site selection other than serendipity of proximity to the researcher or through consulting relationships, or based on asking state departments of education and policymakers which districts are “effective” (Bowers, 2010; Trujillo, 2013). This is a similar issue that has continued to plague the early school effectiveness literature (Trujillo, 2013). Without a systematic means of identifying districts on some broad set of measures prior to site selection for qualitative studies, researchers hamper the ability of future studies to compare and contrast findings across district contexts. With little to no justification in site selection for where a potential district may be situated within the distribution of effectiveness of districts across a state, however defined, future studies will find it problematic to relate findings across varied contexts with no baseline to relate to (Bowers, 2010; Trujillo, 2013). Indeed, in discussing this issue of examining effective district practices Leithwood (2010) notes that researchers need to include comparisons and contrasts across study sites, selecting multiple districts from across large policy-relevant datasets, noting that “future research should include efforts to collect evidence from large samples of districts that represent the full range of district performance (high to low), however defined” p. 284.

In a study of the entire population of school districts in the state of Ohio from academic years 2001-02 through 2007-08, Bowers (2010) detailed a hierarchical linear growth modeling approach to identifying the distribution of district effects across a state, and proposed specific districts as potential sites for in-depth analysis. In the study, in an attempt to address the past critiques of school effectiveness literature, for districts in Ohio the statistical model appropriately nested time in districts and modeled district growth in the Ohio Performance Index (PI) Score, which is a weighted aggregated measure of district standardized test performance across grades 3 through 12 in mathematics, English, science and social studies, controlling for known background and demographic variables that are outside the influence of district administration. Districts that statistically significantly “outperformed” the model-predicted growth in PI score were deemed to be potential sites for

in-depth DER studies, since given the model, the outperforming districts had significant achievement growth across multiple grade levels, subjects and seven years in comparison to the average across the entire population of districts in the state. Bowers (2010) noted that future DER studies should work to include sites for in-depth qualitative analysis that selected from across the distribution, such that much could be learned in a comparison of the outperforming districts to the districts at the norm, which represent the majority of districts in the state. These majority of districts at the norm that have stable performance trends that are mostly predictable based on their demographics could be well-served by understanding how their peers across the state who outperform the model have varied practices around the same policies and mandates that they work with on a daily basis (Bowers, 2010). Additionally, similar such studies could be conducted in comparing the underperforming districts with the norm. Thus, this type of identification strategy potentially helps to inform findings across studies, build theory, and aid in external validity in district effectiveness research (Trujillo, 2013).

However, the Bowers (2010) study is problematic in three main ways. First, the seven year trends in Ohio included the academic years 2001-02 through 2007-08, which corresponds to the beginning of the U.S. federal policy-era of NCLB, a markedly different period in education policy than currently (Furgol & Helms, 2012; McDonnell, 2013). Second, while the model in Bowers (2010) controlled for multiple covariates on the intercepts (year 2001-02 PIS), the slopes through time (i.e. the rate of growth over seven years) were modelled without covariates. Third, the districts model did not account for school-level variance. This is an important consideration since individual exceptionally high or low performing schools within a district may be overweighting the district-level performance measures in the model. From a DER theory standpoint in considering the assertion that districts can exert substantial positive influence on schools through a “coherent instructional focus” (Thompson, et al., 2008), understanding the extent of variance between schools within a district in any definition of effectiveness is an important caveat that helps to inform the “coherence” (high or low) across a district. Examining this type of issue is important, since district reform efforts may have a substantive impact on elementary school performance, but bringing coherence to secondary school performance from the district level is a tougher nut to crack (Bowers, 2008; Purkey & Smith, 1985), and could be discerned through a model that examined the amount of variance between schools within districts.

Framework of the present study:

Thus, the aim of the present study is to address these multiple issues in district research site selection by testing an updated model that examines the entire population of school districts in a state (Ohio) over an extended period of time that is more recent (2005-06 – 2012-13), includes a broader set of district performance background and control covariates across time and addresses the issue of including a measure of school-level variance in performance. Rather than address the issues with theory from the past DER studies (Trujillo, 2013), the purpose of the present study is to provide an example of a methodological tool in action as a means to help future district researchers identify potential sites for in-depth qualitative analysis while addressing the past critiques of school and district effectiveness research. Hence, the research question for this study is:

To what extent can school districts be identified from all of the districts in a state that significantly outperform or underperform long-term performance trends across multiple indicators?

Table 1: Descriptives of Ohio District Variables 2005-06 – 2012-13.

<i>Variable</i>	<i>Mean</i>	<i>(SD)</i>	<i>Min</i>	<i>Max</i>
Performance Index (PI) Score	97.25	(6.30)	69.80	112.40
% Disadvantaged students	0.35	(0.19)	0	1.00
% Asian students	0.01	(0.02)	0	0.17
% African American students	0.06	(0.14)	0	0.99
% Hispanic students	0.02	(0.04)	0	0.45
% High mobility students	0.39	(0.11)	0.15	1.00
% Disabled students	0.14	(0.04)	0	0.31
Small enrollment (<i>vs. medium</i>)	0.33	(0.47)	0	1.00
Large enrollment (<i>vs. medium</i>)	0.31	(0.46)	0	1.00
Extra Large enrollment (<i>vs. medium</i>)	0.02	(0.15)	0	1.00
Urban (<i>vs. suburban</i>)	0.04	(0.19)	0	1.00
Small Town (<i>vs. suburban</i>)	0.12	(0.33)	0	1.00
Rural (<i>vs. suburban</i>)	0.47	(0.50)	0	1.00
% Student attendance	0.95	(0.01)	0.89	0.99
School-level PI score variance	3.97	(2.89)	0	36.92
Avg. teacher years experience	14.84	(3.29)	4.00	34.00
Student-Teacher ratio	15.94	(1.95)	0.87	49.73
Avg. teacher salary (\$)	51859.48	(7228.81)	30810	81851
<i>n</i>	610			
Number of years of data	8			

METHOD:*Sample:*

This study is a secondary data analysis of publically available data from the Ohio Department of Education (ODE, 2014). The subsequent models discussed below include all currently available years of data at the time of this writing, from academic years 2005-06 through 2012-13 (eight timepoints), for all local school districts in the state with complete data ($n=610$).

Variables:

The dependent variable in the subsequent models is the Ohio Performance Index (PI) Score for each district in each of the eight years. The PI score is the overall indicator of district performance within any one year, as well as a district's rating for Adequate Yearly Progress (AYP) under NCLB (ODE, 2014). For more information on PI score use in this type of research, please refer to ODE (2014) or Bowers (2010). Briefly, the PI score ranges from 0 to 120 and is calculated as a weighted average of student performance in grades 3, 8 and 10 across all tested subjects in the district, with higher scores receiving higher weights, accounting for non-test takers (ODE, 2014). Tests include mathematics, reading, writing, science and social studies. Thus, this measure is a policy-relevant indicator that matters to school district central office staff across the state, as their districts and schools are rated on this assessment for sanctions or commendations under the current policy environment (Bowers, 2010).

I used the previous literature on school and district effects to guide the variable selection for covariates and controls for this study (Bowers, 2008, 2010; Murphy & Hallinger, 1988; Skrla, et al., 2011; Trujillo, 2013). Table 1 provides descriptive statistics for the variables, including mean, standard deviation, minimum and maximum. Background control and demographic variables at the district level included *Percent Disadvantaged Students*, *Percent Asian Students*, *Percent African American Students*, *Percent*

Hispanic Students and *Percent Disabled Students*. *Percent High Mobility Students* was defined as the percentage of students who had attended schools in the district for two or less academic years. Using past research as a guide that used school enrollment categories (Bowers & Urick, 2011; Rumberger & Palardy, 2005), district enrollment categories were based on an examination of the U.S. Department of Education National Center for Education Statistics (NCES) Common Core of Data (CCD) for all school districts in the U.S. in 2005-06 (the first year of data included here from Ohio), $n=16,092$ with valid enrollment data (NCES, n.d.). District enrollment categories were defined as follows: *Small Enrollment* ranged from 1 to 1299 students (60.1% of all U.S. districts), *Large Enrollment* ranged from 2400 through 9999 students (20.2% of all U.S. districts), and *Extra Large enrollment* ranged from 10,000 students and up (5.3 % of all U.S. districts), with Medium Enrollment as the reference group.

Following the recommendations of past research at the district level (Bowers & Lee, 2013) district locale urbanicity categories were derived from the NCES CCD (NCES, n.d.) in which the U.S. Census metrocentric codes were converted into the variables *Urban*, *Small Town*, and *Rural*, with *Suburban* as the reference group. The variables *Percent Student Attendance*, *Average Years Teacher Experience*, *Student-Teacher Ratio*, and *Average Teacher Salary* replicate and extend the variables used in past studies into the present study (Bowers, 2010; Skrla, et al., 2011). To address the past critique discussed above of district effects models failing to incorporate a measure of school-level variance across the district, *School-Level PI Score Variance* was calculated in the following way. First, for each year of the dataset, the entire universe of all school level Performance Index Scores across all schools in Ohio were obtained ($n=3,677$). Schools were then aggregated within each of the 610 districts, and the standard deviation of Performance Index Scores was calculated for each district across its schools. The reason to aggregate this proxy of within district variance is due to the constraints of the hierarchical

linear growth model described below, in which the outcome of interest (here, district PI score for each year) must be at level 1 in the model (Hox, 2010; Raudenbush & Bryk, 2002).

Analysis:

The analytic model for this study was a hierarchical linear growth model. Hierarchical linear growth modeling, in which timepoints are nested within the unit of analysis (here the school district), have recently emerged as a useful means to address many of the past critiques of examining cross-sectional models, or value-added models, each of which fail to appropriately control for growth through time (Bowers, 2010; Bowers & White, 2014; Hallinger, Heck, & Murphy, 2014; Hallinger & Heck, 2011; Heck, Lam, & Thomas, 2014; White & Bowers, 2011). Following the recommendations from the hierarchical linear growth modeling literature (Heck, Thomas, & Tabata, 2013; Hox, 2010; Raudenbush & Bryk, 2002; Singer & Willett, 2003), the dataset was structured in a “unit-period” format, such that each school district was represented with eight rows of data, one for each year, and using the nomenclature nominated by Raudenbush & Bryk (2002) and Heck, Thomas & Tabata (2013), the full model used for the analysis here can be written as:

Level 1:

$$\text{PI Score}_{ti} = \pi_{0i} + \pi_{1i}\text{YEAR}_{ti} + e_{ti}$$

Level 2:

$$\pi_{0i} = \beta_{00} + \beta_{01}X_{ti} \dots + u_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}X_{ti} \dots + u_{1i}$$

in which:

PI Score_{ti} = district PI Score for district i in time t

YEAR_{ti} = year for each district's data

X_{ti} = time-varying covariates for each district in each year

π_{0i} = the slope of the intercepts varying randomly across districts (district i 's estimated PI score in 2005-06)

π_{1i} = the slope of YEAR varying randomly across districts (the annual change in district i 's PI scores growing or declining between 2005-06 and 2012-13)

e_{ti} = error in predicting average PI score for district i in time t

u_{0i} = level 2 random component describing differences in average district starting PI scores

u_{1i} = level 2 random component describing differences in average district PI score growth rates

Thus, the hierarchical linear growth model, models the average Ohio district growth in PI score through the eight year timespan, controlling for PI score in the first year of the model (intercepts) and estimating parameters for all covariates on both the intercepts (year 2005-06 PI score) and slopes (growth in PI score over the eight years). All models were analyzed using SPSS version 22. The syntax for the final model is included in Appendix A. Additionally, as will be noted in the results, following the recommendations for specifying the covariance structure for repeated measures (Heck, et al., 2013) an autoregressive error covariance matrix (AR1) was specified at level 1 while a diagonal error covariance matrix (DIAG) was specified at level 2.

Finally, model predicted PI scores were calculated for each of the 610 districts in the dataset. Predicted gain was calculated by subtracting year 2005-06 predicted PI score from 2012-13 predicted PI score, and the same was done for actual PI score gain.

These two gain score calculations were then plotted against each other to generate the final figure.

RESULTS:

The purpose of this study is to examine the extent to which individual school districts can be identified from all districts in a state comparing growth in longitudinal assessments across their constituent schools in multiple grade levels and subjects, controlling for background and control variables that are outside the control of the district administration. Districts that significantly outperform model predicted performance are then categorized as outperforming, at the norm, or underperforming to provide additional baseline context data for informed site selection for in-depth qualitative district studies. In this section, I first present a description of the overall growth trajectories in PI score for districts in Ohio, to provide background and justification for the study. Second, I note the fit of the unconditional hierarchical linear growth model and the significant amount of between district variance available to model. Third, I detail the fit of the full model to the data, and then present the comparison of predicted PI score gains versus actual PI score gains. I then turn to the discussion of the results, along with a list of the 15 model-identified “outperforming” districts with descriptive information to aid in site selection for future district studies.

Examining Ohio School District PI Score Trajectories

Figure 1 plots the twelve year longitudinal trajectories in PI score for each of the 610 school districts in Ohio, from academic year 2001-02 to 2012-13. PI scores can range from zero to 120, however, no districts scored below 50 in their performance index, so here they are plotted from 40 to 120. While the present study focuses on the academic years 2005-06 through 2012-13, which represent the years with complete data on all variables accessible from the state of Ohio, trends from academic year 2001-02 forward are plotted as the author had access to this data from the past Bowers (2010) study, indicated in the lower portion of Figure 1. The data are presented here in Figure 1 to demonstrate the difference in the two studies. Note that a significant shift in PI score trends occurred around 2005-06, and that while trends on average do continue to rise (as will be noted in the model below) the dataset analyzed in the present study represents a different epoch of time in Ohio's measures of district performance. This demonstrates that while the argument here is that the method presented to identify districts is robust and applicable to any state, states as a whole change over time, and so any method to identify districts for site selection should be updated to reflect the overall policy and performance context of a state, as is the purpose of the present study.

Hierarchical Linear Growth Model of Ohio PI Score Change

As noted in the methods, I used a hierarchical linear growth model to estimate the predicted eight year growth in district PI score. Following the recommendations for this type of growth modeling (Heck, et al., 2013; Hox, 2010), I first estimated the unconditional (null) model, and examined differences in model fit between different specifications of the error covariance matrix specified for repeated measures. As is recommended (Heck, et al., 2013), an iterative set of combinations of specifications of the error covariance matrix were tested for fit at level 1 and level 2 in the model, including scaled identity, diagonal, autoregressive and unstructured (data not shown). The specification with the best fit was an autoregressive error covariance matrix (AR1) at level 1 and a diagonal error covariance matrix at level 2 with a -2 log likelihood of 20051.374, AIC 20059.374 and BIC of 20085.331.

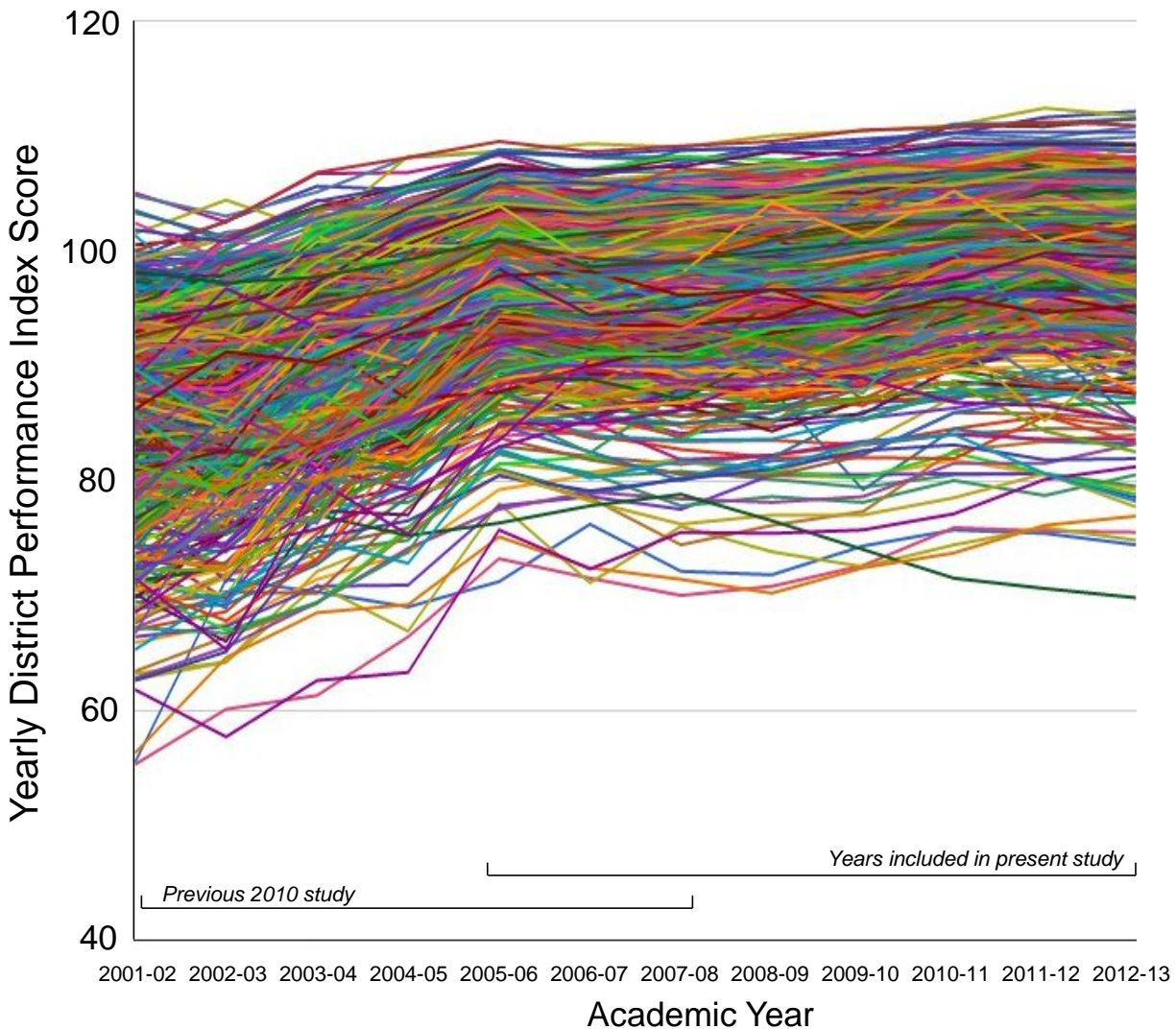


Figure 1: Twelve years of Performance Index (PI) scores for every district in Ohio. Twelve years of consecutive performance index score data for the entire population of school districts in the state of Ohio are plotted. Each line represents the trend of one of the 610 individual district twelve year history of performance index score change. District trends are plotted from 40-120 PI score, since no districts scored below 50 PI score at any time. Years included in the previous Bowers (2010) study, the present study and the overlap between the two models are shown at the bottom of the panel.

A benefit of the AR1 error covariance matrix is that it models the covariances between occasions as differing and assumes that correlations may weaken between occasions separated by multiple timepoints as well as provides an estimate of ρ (ρ), an autocorrelation coefficient, which represents the correlation in the outcome variable between two adjacent occasions (Heck, et al., 2013). The Wald Z for the level 2 variance in the unconditional model was significant (Wald Z = 3.620, $p < 0.001$) and the Intraclass Correlation Coefficient (ICC) was 0.589, indicating that 58.9% of the variance in PI score in the state of Ohio from 2005-06 through 2012-13 was between districts.

Table 2 reports the full hierarchical linear growth model results. All continuous variables in the model were standardized (z -scored), except for *school-level PI score variance*. The ethnicity variables were also square root transformed. Thus, the intercept in the model represents the average PI score in the first year of the model, 2005-06, for a district at the average for all continuous

variables and was a medium enrollment suburban district with no variance between schools in PI score for that year (zeroes on the dichotomous enrollment and locale variables). The Year coefficient represents the slope through time, and thus was the average yearly growth in PI score, accounting for the control variables. ρ for the model was 0.480, indicating that when controlling for the variables in the model, the average district's PI score in any one year was correlated at 0.480 with the subsequent year PI score. As noted at the bottom of Table 2, the model fit the data well and accounted for 84.325% of the variance within districts and 44.137% of the variance between, an improvement over past models (Bowers, 2010), and had a $-2 \log$ likelihood of 19454.740, an AIC of 19462.740 and a BIC of 19488.669.

Table 2: Parameter estimates for a hierarchical linear growth model of Ohio PI score change 2005-06 – 2012-13

<i>Variable</i>	<i>Coeff.</i>	<i>(SE)</i>	<i>Stand. Coeff.</i>	<i>p-value</i>
Intercept	96.726***	(0.365)		<0.001
Year	0.492***	(0.044)	0.179	<0.001
<i>Intercepts</i>				
% Disadvantaged students	-0.864***	(0.097)	-0.137	<0.001
% Asian students ^a	0.668***	(0.130)	0.106	<0.001
% African American students ^a	-0.849***	(0.154)	-0.135	<0.001
% Hispanic students ^a	0.268*	(0.106)	0.043	0.011
% High mobility students	-0.122	(0.068)		0.073
% Disabled students	-0.617***	(0.095)	-0.098	<0.001
Small enrollment	-0.595	(0.406)		0.143
Large enrollment	0.562	(0.427)		0.189
Extra Large enrollment	-1.511	(1.204)		0.210
Urban	-7.380***	(0.965)	-0.219	<0.001
Small Town	-2.133***	(0.542)	-0.111	<0.001
Rural	-1.031*	(0.406)	-0.082	0.011
% Student attendance	0.825***	(0.080)	0.131	<0.001
School-level PI score variance	-0.004	(0.016)		0.824
Avg. teacher years experience	0.405***	(0.075)	0.064	<0.001
Student-Teacher ratio	-0.186**	(0.070)	-0.030	0.008
Avg. teacher salary	-0.041	(0.112)		0.713
<i>Slopes</i>				
% Disadvantaged	0.152***	(0.020)	0.103	<0.001
% Asian students	-0.057*	(0.024)	-0.040	0.015
% African American students	-0.082***	(0.023)	-0.054	<0.001
% Hispanic students	-0.037*	(0.017)	-0.026	0.032
% High mobility students	0.004	(0.017)		0.835
% Disabled students	0.032	(0.019)		0.099
Small enrollment	-0.031	(0.043)		0.469
Large enrollment	-0.140**	(0.046)	-0.046	0.002
Extra Large enrollment	-0.081	(0.128)		0.527
Urban	0.092	(0.102)		0.370
Small Town	0.055	(0.058)		0.347
Rural	0.128**	(0.044)	0.048	0.003
% Student attendance	0.024	(0.018)		0.177
School-level PI score variance	-0.023***	(0.005)	-0.047	<0.001
Avg. teacher years experience	-0.089***	(0.015)	-0.060	<0.001
Student-Teacher ratio	0.014	(0.014)		0.325
Avg. teacher salary	0.141***	(0.021)	0.097	<0.001
Amount of variance between districts (ICC) (%)				
	58.869			
<i>Variance Explained</i>				
Level 1: Within districts (%)	84.325			
Level 2: Between districts (%)	44.137			

Notes:

^a: % Asian, African American and Hispanic student enrollment variables are square root transformed.

All variables are standardized (z-scored) except for School-level PI score variance and the enrollment and locale dichotomous variables.

Turning next to the significant parameters in the model, there were multiple significant variables on the estimate of the intercept – the average PI score in the first year of the model, 2005-06. *Percent disadvantaged students*, *percent Asian students* and *percent African American students* and *percent disabled students* were all significant and in the expected directions given the past research noted above. However, *percent Hispanic students* was positive and significant. This is an opposite finding from the previous Bowers (2010) study in which *percent Hispanic students* was negative and significant. This may perhaps indicate a change in influence on district performance in 2005-06 versus the estimates in the previous study on year 2001-02. Enrollment appeared to have no relationship in the full model to year 2005-06 PI score, controlling for other variables in the model. In contrast, district locale was strongly related to the intercept, with districts in *urban*, *small town* and *rural* contexts all starting lower in PI score in the model than suburban districts (reference group). As in the previous research, *percent student attendance*, *average teacher years experience*, and *student-teacher ratio* were all significant on the intercept and in the predicted directions in which controlling for the other variables in the model, districts with higher student attendance or more experienced teachers started on average in the model with higher PI scores than other districts while districts that on average had larger class sizes as indicated by the student-teacher ratio, had lower PI scores on average in 2005-06. I note here that the standardized coefficient for percent student attendance on the intercept of 0.131 is comparable to the strong negative standardized coefficient of percent disadvantaged students, suggesting that a one standard deviation increase in percent student attendance in a district (a 1% increase in average student attendance) is associated with an average increase in PI score of 0.131 in 2005-06, controlling for other variables in the model - an independent effect. *School level PI score variance* for the districts was unrelated to average district PI score in the first year of the model.

Also of interest are the parameter estimates on the slopes through the eight year time period. This is the first study to date that reports on these variables in this type of modeling framework for district level research. *Percent disadvantaged* and *percent Asian students* were significant in the model, but in the opposite directions as expected. This most likely is due to a known ceiling effect with these variables and this type of data (Bowers & White, 2014; Ho, 2008), presented here for the first time at the district level. These variables a) present a skewed distribution across the state of Ohio and b) are strongly associated with the intercepts in the model such that districts with very high percent disadvantaged populations have room in the distribution to grow through time, with the opposite for percent Asian students. Thus, controlling for these variables on the intercepts and then the slopes, *percent African American students* and *percent Hispanic students* were negative and significant on the slopes. Interestingly, *large enrollment* districts grew significantly slower than medium enrollment districts, while none of the other enrollment categories were related to change in PI score through time, controlling for the other variables in the model. In contrast, *rural* districts grew significantly faster in PI score than suburban districts, controlling for the other variables in the model, at a rate comparable to the negative effect on large enrollment districts when considering the standardized coefficients. This perhaps indicates a significant district context effect in Ohio on PI score change through time.

As a test for the first time in this domain, *school-level PI score variance* was negative and significant on the slopes through time, controlling for the other variables in the model. This finding

suggests that the higher the amount of variance between schools within a district, here measured through the standard deviation in school-level PI scores within a district, the slower the growth in PI score through these eight years in Ohio - an independent effect. However, this finding could be critiqued as an effect of enrollment size in a district, rather than PI score variance across schools, and thus is accounting for some variance in the number of schools used in the standard deviation calculation. However, this is most likely minimized by the inclusion of the enrollment and locale control variables, and thus I posit that *school-level PI score variance* is a reasonable proxy that attempts to capture this important variance for the first time in this type of model. More importantly, the inclusion of this variable in the model captures the variance within districts between schools, and allows districts with low variance between schools to rise faster in the model, and more likely be identified as “outperforming” than districts that may have just one or two very high performing schools versus schools that perform at or below the average for the state.

And finally in Table 2, *average teacher experience* was negative and significant on the slopes through time while *average teacher salary* was positive and significant. As with some of the other variables on the slopes, the average teacher experience finding is most likely a ceiling effect as more experienced teachers tend to work in higher performing schools, and thus districts, and so these districts have less room to grow through time, especially for the most experienced teachers and when accounting for average teacher salary in the model as I do here.

Figure 2 presents the final comparison of the districts during this time period in Ohio. Following the recommendations of the past research in this domain (Bowers, 2010; Trujillo, 2013), as noted in the methods, the full specification of the hierarchical linear growth model detailed in Table 2 was used to predict the PI scores for each of the 610 Ohio school districts for each academic year, 2005-06 through 2012-13, which in turn were used to calculate the gain in PI score over the eight years. Gains in actual PI scores were also calculated. Figure 2 presents the comparison of the model predicted gains versus the actual gains with each district represented by a symbol that relates to the 2012-13 Ohio Department of Education “letter grade” for the number of standards that the district met. In relating to the purpose of the present study to aid in providing a baseline for comparisons of all school districts in a state for site selection for in-depth qualitative studies of district and central office practices, Figure 2 identifies three different types of districts. First, above the $p=0.05$ 95% confidence interval line, 15 districts are identified as “outperforming” the model predicted eight year gain in PI score, termed here as outperforming the district’s background and demographic variables in relation to the state averages, and for the first time in this literature controlling for school-level PI score variance and the effect of the control variables through time. Second, the majority of the districts fall within the 95% confidence interval of their model predicted PI score gain, while third, multiple districts fall below the 95% confidence interval and thus are deemed here as “underperforming”.

Note in Figure 2 the high prevalence of districts receiving an “A” by the state in the number of standards met being below the 95% confidence interval, with conversely many districts receiving an “F” by the state that are above the 95% confidence interval (see Figure 2). This is a core feature of this type of district modeling and identification strategy and highlights the problematic issues with district effectiveness research in which researchers select sites based on state or local perception of “effectiveness” noted in the

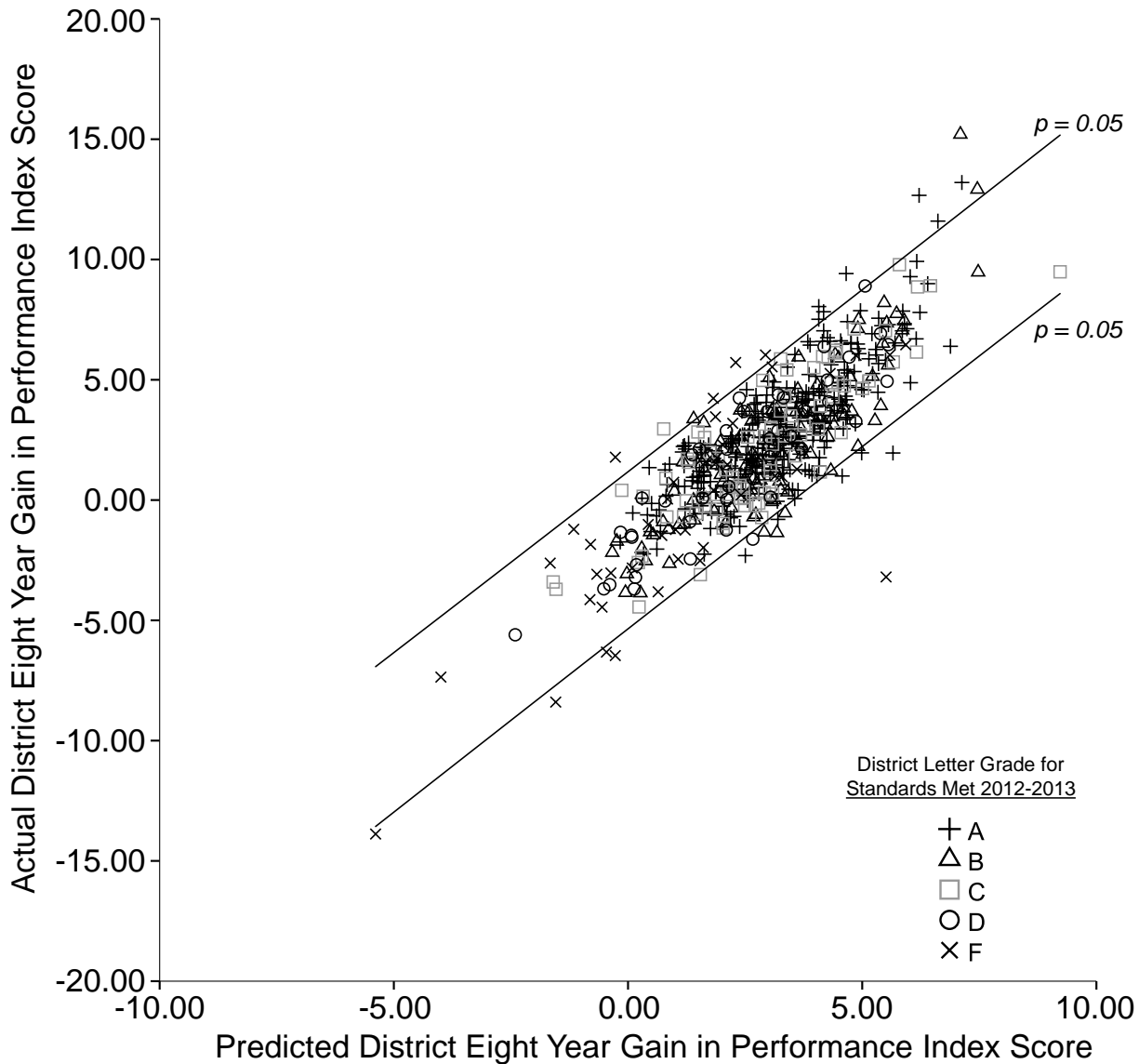


Figure 2: Comparison of predicted versus actual eight year Performance Index (PI) score gain in Ohio, 2005-06 – 2012-13. Hierarchical linear growth modeling predicted PI score gains are plotted on the x-axis for each of the 610 schools in Ohio over the time period, versus the actual gains from 2005-06 to 2012-13. The 95% confidence intervals are shown. Individual districts are plotted as symbols representing the Ohio Department of Education letter grade for adequate yearly progress on the number of standards met, with A being the highest (90-100% of standards met) and F the lowest (0-49%). Districts that are outside the 95% confidence interval are designated as significantly outperforming or underperforming their demographic and context variables.

previous research (Trujillo, 2013). While state designations are important, the central argument here of this study is that this type of modeling framework controls for district context, such that districts in relatively advantageous situations, such as with very high percentages of non-economically disadvantaged students, are not nominated as “outperforming” since the model controls for the effect of these variables. Additionally, this type of modeling framework is focused on change over time rather than on any one year of performance, again helping to account for local perception in which a single year of “success” may not provide a full picture of how well a district may be performing, addressing one of the central critiques from the school effectiveness research domain (Bowers, 2010; Rowan, Bossert, & Dwyer, 1983; Trujillo, 2013). Moreover, I acknowledge that researchers may not be as interested in examining districts that grow over time in relation to the state

averages, and may instead be interested in district PI score stability in contexts in which demographics may either be stable or changing. The model identifies these districts as the central “at the norm” districts, and as has been argued previously (Bowers, 2010; Trujillo, 2013), provides researchers interested in district effects with a very interesting and useful comparison group to the “outperforming” or “underperforming” districts. The argument is such that this type of baseline data for district comparisons provides a much more valid framework of district practices, since districts at the norm are the majority and thus could benefit the most from a deeper understanding of how their practices differ from those who outperform or underperform in similar contexts, rather than a singular focus on comparing effective districts to other effective districts or, more problematic, districts deemed

highly effective versus highly in-effective, both of which present radically different contexts.

DISCUSSION

The findings from this study point to three main conclusions for researchers, practitioners and policymakers wishing to identify a set of baseline data across a selected state to help select sites for in-depth district effectiveness research in an effort to understand, communicate and replicate useful practices across district contexts. First, the results demonstrate that through using only publically available data, the distribution of district “effectiveness” across an entire state, defined here as a district outperforming the background and demographic variables in relation to the state averages, can be estimated with identification of individual districts as either significantly outperforming, at the norm, or underperforming. Second, the findings highlight the utility of using a hierarchical linear growth model. Controlling for the first year in the dataset, this modelling framework provides an attractive and useful method to examine school district growth through time, controlling for the multiple covariates that are known to be outside of the influence of district central offices, and thus focusing the PI gain score comparison on the residual error in the model which is attributed here to some extent as possibly relating to what a district may be doing differently from other districts. At the least, I argue that the identified “outperforming” districts are of enough interest to justify the significant amount of time, effort and resources required for an in-depth qualitative study of these districts in comparison to districts at the norm. Third, while the overall parameter estimates in the model help add to the literature on what is known about the predictors of district performance over time, the inclusion here for the first time in this type of modeling framework of covariates on the slopes through time and a measure of the school-level within district variance is a unique and significant contribution to the literature on modeling district performance.

Of note, I posit that the finding of the significant negative effect of *school-level PI score variance* on the growth in overall district PI score over the eight years may be the first quantitative measure across an entire state of a positive effect of the postulated district system-wide instructional coherence theory noted in the literature above (Bowers, 2008; Thompson, et al., 2008). Districts with lower variance across their schools grew significantly faster on average than districts with higher variance. This effect may indicate that districts that are able to create a stable system-wide focus manifest that effect through low variance between schools. However, this is but one of many possible explanations, as high variance may merely indicate a larger school district with more schools, although the inclusion of the enrollment and locale variables would appear to control for this. I encourage further research around this intriguing finding.

Table 3 presents descriptive data on the 15 districts identified in this study as outperforming. Districts are listed in order from the highest PI score gain difference to lowest, with the state averages for each indicator listed in the bottom row. As the central aim of this study, Table 3 is meant to help provide a baseline and a means to select districts for in-depth qualitative studies out of the entire universe of districts in a state, helping to make the justification for devoting the significant resources to such studies more evidence based and data informed. The suggestion here, based on the previous literature (Bowers, 2010; Trujillo, 2013) is to select a set of districts from a list such as Table 3 as the “effective” or “outperforming” districts in a state, and then select a set of districts from the norm that are similar on the multiple different types of

context variables. As noted in Table 3, these 15 districts provide a range of contexts, demographics, and performance levels. At a minimum, even if one were to argue that the model may only be identifying districts that are “doing something different” however defined, rather than denoting effectiveness, I would still posit that working to understand which districts are “unusual” from all of the districts in a state prior to site selection for qualitative studies is a reasonable assertion to help provide context and increased external validity and generalizability to the results of such studies.

Moreover, I wish to stress that the districts listed in Table 3 are not similar to each other, other than in their model identified gains in PI score over this time period. As detailed in Table 3, the Southeastern school district is a very different context from Youngstown. The argument here is that working to create a deeper understanding of the practices, routines, leadership styles, resource allocation, and overall organizational practices of these types of districts is important for not only the other districts in Table 3, but all of the districts in the state of Ohio. It may well be that each of the 15 districts are engaging in completely different processes that contribute to their being identified in the statistical model. But there is no way to know unless they are studied through interviews, observations, and document analysis - in short deep qualitative research. Thus, I argue for a new resurgence and focus in school district research, with the aim to help provide a means to communicate how some districts find success in circumstances similar to other districts that are at the averages for their state, or who struggle. Using the data here, or replicating this type of identification model in another state, such a study could select multiple districts from a table such as Table 3, focusing either on variety or similarity across contexts, and then just as importantly, select districts in similar contexts that are at the norm. Alternatively, a comparable study could be done in which the comparison is on the underperforming districts and the norm. Helping these districts at the least reach the averages of their peers is an important, and often overlooked, area of research in this domain. These recommendations for districts follow directly from similar recommendations for school effectiveness research (Bowers, 2010; Klitgaard & Hall, 1975; Purkey & Smith, 1983; Stringfield, 1994; Teddlie, Reynolds, & Sammons, 2000; Trujillo, 2013).

Limitations

While I argue that the findings presented in this study are robust, the study is limited in the following ways. First, the dependent variable and the unit of analysis are at the school district level. The interest in this study is on examining differences between districts to help isolate a “district effect”. However, multiple levels of data are available publically; especially school-level data, as noted here in the work to capture the school-level within district variance in PI score for each year. Conceptually, this could indicate the need for additional levels in the model, such as a three level model, modeling time in schools in districts. However, the focus here is on a district level outcome, working from the district research that notes that district effects are not just a product of adding all of the school effects together, but that there is a unique contribution of the central office that adds an additional layer of complexity to the schooling system (Bowers, 2008; Honig, 2008; Thompson, et al., 2008). Thus the outcome of interest here is uniquely at the district level, not the school level. However, a central requirement of multilevel modeling is that the outcome is at the lowest level in the model (Hox, 2010; Raudenbush & Bryk, 2002). Thus, for this study with a focus on the district level, capturing the school-level variance in PI score was important, but required aggregating the

Table 3: Descriptive Data for the 15 Model-Identified Districts that were above the 95% Confidence Interval.

<i>District</i>	<i>PI Score 2012-13</i>	<i>Actual 8-year PI Score gain</i>	<i>Model Predicted 8-year PI Score Gain</i>	<i>Ohio Letter Grade of Standards Met 2012-13</i>	<i>Enrollment</i>	<i>Locale</i>	<i>% Disadvant. Students</i>	<i>Number of Schools</i>	<i>School PI Score Var.</i>
Southeastern	98.7	15.2	7.10	B	1247	Rural	51%	3	5.58
East Knox	99.6	12.7	6.21	A	1104	Rural	42%	3	3.06
Jackson Center	103.2	13.2	7.13	A	541	Rural	35%	2	0.20
Southern	97.0	12.9	7.46	B	663	Rural	65%	2	2.04
Jackson	100.4	11.6	6.62	A	2489	Small Town	53%	5	3.01
Madison	98.3	9.41	4.66	A	3009	Suburban	62%	5	1.32
Newton Falls	98.2	8.04	4.07	A	1286	Suburban	54%	4	4.72
Groveport Madison	91.8	8.90	5.06	D	5587	Suburban	62%	10	3.69
Swanton	100.1	7.83	4.18	A	1311	Suburban	42%	4	2.68
Nordonia Hills	104.5	7.52	4.07	A	3720	Suburban	17%	6	1.65
Springfield	83.2	5.72	2.30	F	7245	Urban	80%	15	9.10
Ridgedale	92.0	6.03	2.93	F	784	Rural	49%	3	2.22
East Liverpool	88.9	4.23	1.82	F	2161	Small Town	74%	5	4.74
West Carrollton	94.7	2.96	0.76	C	3636	Suburban	47%	8	2.54
Youngstown	76.9	1.78	-0.27	F	5239	Urban	93%	16	14.23
<i>State Averages</i>	<i>98.8</i>	<i>2.37</i>	<i>2.95</i>	<i>52.6% A</i>	<i>2612</i>	<i>47% Rural</i>	<i>41%</i>	<i>5</i>	<i>3.47</i>

indicator to the district level. I look to future research in this area to work to specify additional modeling frameworks that can help account for this issue. Second, the hierarchical linear growth model depended here on a linear specification of growth through time. It may be that additional nonlinear terms could help improve the fit of the model and hence the specification of the gain in PI scores over the years. I look forward to pursuing this type of additional specification in future research.

Third, while the model included average teacher salary, given the well-known problematic persistent state of the state of Ohio's school funding issues (Ingle, Bowers, & Davis, 2012; Porter, 2010), future research should consider working to incorporate more indicators of savvy district financial systems. Much of the district qualitative case study research has shown that a central aspect of effective district practices is how central office personnel find ways to focus all of the different funding streams available to a district towards the central instructional and curricular goals of the district (Bowers, 2008; Elmore & Burney, 1999; Leithwood, 2010). One indication of this focus on district finance would be that a district is able to outperform other districts in a longitudinal model while maintaining spending per pupil across the system at the state averages, or below, perhaps indicating that the district is both effective and efficient. I encourage future education finance research to investigate this type of measure of district financial efficiency.

Conclusions

In conclusion, this study provides an analysis framework to help identify significantly unusual school districts from the population of districts in a state over extended periods of time. Given that much of the research on school districts focuses either on the largest of metropolitan regions in the U.S. (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; O'Day, Bitter, & Gomez, 2011; O'Day & Quick, 2009; Supovitz, 2006), or on specific districts that researchers may have entrée into through multiple different ongoing relationships (as critiqued in Bowers, 2010), the present study provides a means to examine the multiple and diverse contexts of the approximately 16,000 individual school districts across the U.S. As a uniquely American concept, the construct of local educational control through the local school district is a fruitful area of study, from the education, socio-political, organizational, and sociological frameworks (Berkman & Plutzer, 2005). Recent research has begun to examine a wider variety of district contexts, such as a focus on suburban districts (Cuban, Lichtenstein, Evenchik, Tombari, & Pozzoboni, 2010; Holme, Diem, & Welton, 2014), and so the goal of this study is to hopefully provide a richer description of the types of contexts and differences we see across school district organizations as a means to help district administrators find success in their schools and communities.

NOTES:

An earlier version of this manuscript was presented at the 2014 annual meeting of the American Educational Research Association (AERA), Philadelphia PA.

RECOMMENDED CITATION FORMAT:

Bowers, A.J. (2015). Site Selection in School District Research: A Measure of Effectiveness Using Hierarchical Longitudinal Growth Models of Performance. *School Leadership & Management*, 35(1) p.39-61.

REFERENCES:

- Berkman, M. B., & Plutzer, E. (2005). *Ten thousand democracies: Politics and public opinion in America's school districts*. Washington, DC: Georgetown University Press.
- Bowers, A. J. (2008). Promoting Excellence: Good to great, NYC's district 2, and the case of a high performing school district. *Leadership and Policy in Schools*, 7(2), 154-177.
- Bowers, A. J. (2010). Toward Addressing the Issues of Site Selection in District Effectiveness Research: A Two-Level Hierarchical Linear Growth Model. *Educational Administration Quarterly*, 46(3), 395-425. doi: 10.1177/0013161X10375271
- Bowers, A. J., & Lee, J. (2013). Carried or Defeated? Examining the Factors Associated With Passing School District Bond Elections in Texas, 1997-2009. *Educational Administration Quarterly*, 49(5), 732-767. doi: 10.1177/0013161x13486278
- Bowers, A. J., & Urlick, A. (2011). Does high school facility quality affect student achievement? A 2-level hierarchical linear model. *Journal of Education Finance*, 37(1), 72-94.
- Bowers, A. J., & White, B. R. (2014). Do Principal Preparation and Teacher Qualifications Influence Different Types of School Growth Trajectories in Illinois? A Growth Mixture Model Analysis. *Journal of Educational Administration*, 52(5), 705-736. doi: 10.1108/JEA-12-2012-0134
- Bryk, A. S., Gomez, L., & Grunow, A. (2011). Getting Ideas into Action: Building Networked Improvement Communities in Education. In M. T. Hallinan (Ed.), *Frontiers in Sociology of Education* (Vol. 1, pp. 127-162): Springer Netherlands.
- Bryk, A. S., Sebring, P. B., Allensworth, E. M., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: The University of Chicago Press.
- Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, 31(1), 92-109. doi: 10.1016/j.econedurev.2011.09.001
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction and research. *Educational Evaluation and Policy Analysis*, 25(2), 119-142.
- Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy and practice at the district level. *Harvard Educational Review*, 54(2), 129-151.
- Cuban, L., Lichtenstein, G., Evenchik, A., Tombari, M., & Pozzoboni, K. (2010). *Against the Odds: Insights from One District's Small School Reform*. Cambridge, MA: Harvard Education Press.
- Daly, A. J., & Finnigan, K. S. (2011). The Ebb and Flow of Social Network Ties Between District Leaders Under High-Stakes Accountability. *American Educational Research Journal*, 48(1), 39-79. doi: 10.3102/0002831210368990
- Daly, A. J., & Finnigan, K. S. (2012). Exploring the Space Between: Social Networks, Trust, and Urban School District Leaders. [Article]. *Journal of School Leadership*, 22(3), 493-530.
- Daly, A. J., Moolenaar, N. M., Bolivar, J. M., & Burke, P. H. (2010). Relationships in reform: the role of teachers' social networks. *Journal of Educational Administration*, 48(3), 359-391.
- Darling-Hammond, L., Hightower, A. M., Husbands, J. L., Lafors, J. R., Young, V. M., & Christopher, C. (2005).

- Instructional leadership for systemic change: The story of San Diego's reform (Leading systemic school improvement)*. Lanham: ScarecrowEducation.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 31(1), 15-24.
- Elmore, R. F., & Burney, D. (1997). School variation and systemic instructional improvement in community school district #2, New York City (pp. 1-41). Pittsburgh: Pittsburgh University.
- Elmore, R. F., & Burney, D. (1999). Investing in teacher learning: Staff development and instructional improvement. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession : Handbook of policy and practice* (pp. 263-291). San Francisco: Jossey-Bass.
- Farley-Ripple, E. N., & Cho, V. (2014). Depth of Use: How District Decision-makers Did and Did Not Engage with Evidence. In A. J. Bowers, A. R. Shoho & B. G. Barnett (Eds.), *Using Data in Schools to Inform Leadership and Decision Making* (pp. 229-252). Charlotte, NIC: Information Age Publishing.
- Firestone, W. A., Mangin, M. M., Martinez, M. C., & Polovsky, T. (2005). Leading coherent professional development: A comparison of three districts. *Educational Administration Quarterly*, 41(3), 413-448.
- Floden, R. E., Porter, A. C., Alford, L. E., Freeman, D. J., Irwin, S., & Schmidt, W. H. (1988). Instructional leadership at the district level: A closer look at autonomy and control. *Educational Administration Quarterly*, 24(2), 96-124.
- Furgol, K. E., & Helms, L. B. (2012). Lessons in Leveraging Implementation: Rulemaking, Growth Models, and Policy Dynamics Under NCLB. *Educational Policy*, 26(6), 777-812. doi: 10.1177/0895904811417588
- Hallinger, P., Heck, R., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 1-24. doi: 10.1007/s11092-013-9179-5
- Hallinger, P., & Heck, R. H. (2011). Exploring the journey of school improvement: classifying and analyzing patterns of change in school improvement processes and learning outcomes. *School Effectiveness and School Improvement*, 22(1), 1-27. doi: 10.1080/09243453.2010.536322
- Heck, R. H., & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal*, 46(3), 659-689.
- Heck, R. H., Lam, W. S., & Thomas, S. L. (2014). State Political Culture, Higher Education Spending Indicators, and Undergraduate Graduation Outcomes. *Educational Policy*, 28(1), 3-39. doi: 10.1177/0895904812453996
- Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: a regression discontinuity approach. *School Effectiveness and School Improvement*, 21(4), 377-408. doi: 10.1080/09243453.2010.500097
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and Longitudinal Modeling with IBM SPSS* (2nd ed.). New York, NY: Routledge.
- Hightower, A. M., Knapp, M. S., Marsh, J. A., & McLaughlin, M. W. (2002). The district role in instructional renewal: Setting the stage for dialogue. In A. M. Hightower, M. S. Knapp, J. A. Marsh & M. W. McLaughlin (Eds.), *School districts and instructional renewal* (pp. 1-6). New York: Teachers College Press.
- Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under no child left behind. *Educational Researcher*, 37(6), 351-360.
- Holme, J. J., Diem, S., & Welton, A. (2014). Suburban School Districts and Demographic Change: The Technical, Normative, and Political Dimensions of Response. *Educational Administration Quarterly*, 50(1), 34-66. doi: 10.1177/0013161x13484038
- Honig, M. I. (2003). Building Policy from Practice: District Central Office Administrators' Roles and Capacity for Implementing Collaborative Education Policy. *Educational Administration Quarterly*, 39(3), 292-338. doi: 10.1177/0013161x03253414
- Honig, M. I. (2008). District Central Offices as Learning Organizations: How Sociocultural and Organizational Learning Theories Elaborate District Central Office Administrators' Participation in Teaching and Learning Improvement Efforts. *American Journal of Education*, 114(4), 627-664. doi: 10.1086/589317
- Honig, M. I. (2009). No small thing: School district central office bureaucracies and the implementation of new small autonomous school initiatives. *American Educational Research Journal*, 46(2), 387-422.
- Honig, M. I. (2012). District Central Office Leadership as Teaching: How Central Office Administrators Support Principals' Development as Instructional Leaders. *Educational Administration Quarterly*, 48(4), 733-774. doi: 10.1177/0013161x12443258
- Honig, M. I., & Coburn, C. E. (2008). Evidence-based decision making in school district central offices. *Educational Policy*, 22(4), 578-608.
- Honig, M. I., & Venkateswaran, N. (2012). School-Central Office Relationships in Evidence Use: Understanding Evidence Use as a Systems Problem. *American Journal of Education*, 118(2), 199-222. doi: 10.1086/663282
- Hox, J. (2010). *Multilevel analysis: Techniques and applications* (Second ed.). New York: Routledge.
- Ingle, W. K., Bowers, A. J., & Davis, T. (2012). *Which School Districts Qualified for Federal School Facility Funding Under ARRA? Evidence from Ohio*. Paper presented at the Annual meeting of the University Council for Educational Administration (UCEA), Denver, CO.
- Klitgaard, R. E., & Hall, G. R. (1975). Are there unusually effective schools? *The Journal of Human Resources*, 10(1), 90-106.
- Knapp, Michael S. (2008). How Can Organizational and Sociocultural Learning Theories Shed Light on District Instructional Reform? *American Journal of Education*, 114(4), 521-539. doi: 10.1086/589313
- Leithwood, K. (2010). Characteristics of school districts that are exceptionally effective in closing the achievement gap. *Leadership and Policy in Schools*, 9, 245-291. doi: 10.1080/15700761003731500
- Levin, B., Datnow, A., & Carrier, N. (2012). Changing school district practices. Boston, MA: Students at the Center: Teaching and Learning in the Era of the Common Core: A Jobs for the Future Project.
- Luyten, H., Visscher, A., & Witziers, B. (2005). School effectiveness research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement*, 16(3), 249-279.
- McDonnell, L. M. (2013). Educational Accountability and Policy Feedback. *Educational Policy*, 27(2), 170-189. doi: 10.1177/0895904812465119

- McLaughlin, M. W., & Talbert, J. E. (2002). Reforming districts. In A. M. Hightower, M. S. Knapp, J. A. Marsh & M. W. McLaughlin (Eds.), *School districts and instructional renewal* (pp. 193-202). New York: Teachers College Press.
- Mitgang, L. (2013). *Districts Matter: Cultivating the Principals Urban Schools Need*. New York, NY: The Wallace Foundation.
- Murphy, J., & Hallinger, P. (1988). Characteristics of instructionally effective school districts. *Journal of Educational Research, 81*(3), 175-181.
- NCES. (n.d.). Common Core of Data Retrieved December 18, 2006, from <http://nces.ed.gov/ccd/>
- O'Day, J. A., Bitter, C. S., & Gomez, L. M. (2011). *Education reform in New York City: Ambitious change in the nation's most complex school system*. Cambridge, MA: Harvard Education Press.
- O'Day, J. A., & Quick, H. (2009). Assessing instructional reform in San Diego: A theory-based approach. *Journal of Education for Students Placed at Risk, 14*(1), 1-16.
- ODE. (2014). *Office of data services*. Columbus, OH: Ohio Department of Education Retrieved from <http://education.ohio.gov/>.
- Opfer, V. D., Henry, G. T., & Mashburn, A. J. (2008). The district effect: Systemic responses to high stakes accountability policies in six southern states. *American Journal of Education, 114*(2), 299-332.
- Porter, T. S. (2010). Trends in the Fiscal Capacity of Ohio's Public Schools. [Article]. *Journal of Education Finance, 35*(3), 217-244. doi: 10.1353/jef.0.0016
- Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 426-452.
- Purkey, S. C., & Smith, M. S. (1985). School reform: The district policy implications of the effective schools literature. *The Elementary School Journal, 85*(3), 352-389.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.
- Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 635-674.
- Rorrer, A. K., Skrla, L., & Scheurich, J. J. (2008). Districts as institutional actors in educational reform. *Educational Administration Quarterly, 44*(3), 307-357.
- Rowan, B., Bossert, S. T., & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. *Educational Researcher, 12*(4), 24-31.
- Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal, 42*(1), 3-42. doi: 10.3102/00028312042001003
- Scheerens, J. (2012). The use of theory in school effectiveness research revisited. *School Effectiveness and School Improvement, 24*(1), 1-38. doi: 10.1080/09243453.2012.691100
- Seashore Louis, K. (2008). Learning to Support Improvement—Next Steps for Research on District Practice. *American Journal of Education, 114*(4), 681-689. doi: 10.1086/589320
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Skrla, L., McKenzie, K. B., Scheurich, J. J., & Dickerson, K. L. (2011). Home-town values and high accountability: A Texas recipe for districtwide success in an urban school district. *Journal of Education for Students Placed at Risk, 16*(2), 137-165. doi: 10.1080/10824669.2011.559902
- Smith, M. S., & O'Day, J. A. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing: The 1990 yearbook of the politics of education association* (pp. 233-267). New York: The Falmer Press.
- Spillane, J. P., & Thompson, C. L. (1997). Reconstructing Conceptions of Local Capacity: The Local Education Agency's Capacity for Ambitious Instructional Reform. *Educational Evaluation and Policy Analysis, 19*(2), 185-203. doi: 10.3102/01623737019002185
- Stringfield, S. (1994). Outlier studies of school effectiveness. In D. Reynolds, B. P. M. Creemers, P. S. Nesselrodt, E. C. Schaffer, S. Stringfield & C. Teddlie (Eds.), *Advances in school effectiveness research and practice* (pp. 73-84). Tarrytown NY: Elsevier Science.
- Supovitz, J. A. (2006). *The case for district-based reform: Leading, building, and sustaining school improvement*. Cambridge, Massachusetts: Harvard Education Press.
- Teddlie, C., Reynolds, D., & Sammons, P. (2000). The methodology and scientific properties of school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 55-134). New York, NY: Falmer Press.
- Thompson, C. L., Sykes, G., & Skrla, L. (2008). Coherent, instructionally-focused district leadership: Toward a theoretical account, from http://education.msu.edu/epc/forms/Thompson_et_al_2008_Coherent_Instructionally_Focused_District_Leadership.pdf
- Trujillo, T. (2013). The Reincarnation of the Effective Schools Research: Rethinking the Literature on District Effectiveness. *Journal of Educational Administration, 51*(4), 426-452. doi: 10.1108/09578231311325640
- Turnbull, B. J., Riley, D. L., Arcaira, E. R., Anderson, L. M., & MacFarlane, J. R. (2013). *Six Districts Begin the Principal Pipeline Initiative*. New York, NY: The Wallace Foundation.
- Wallace Foundation. (2013). *Districts Matter: Cultivating the principals urban schools need*. New York, NY.
- White, B. R., & Bowers, A. J. (2011). Principal effects in Illinois: A research brief (IERC 2011-3). Edwardsville, IL: The Illinois Education Research Council (IERC), Southern Illinois University Edwardsville.
- Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: applying the principal-agent framework. *School Effectiveness and School Improvement, 19*(3), 239-259. doi: 10.1080/09243450802246376

Appendix A:
SPSS model syntax:

```

MIXED PIS WITH TIME ZPerc_Disadvant ZSQRT_Perc_Asian ZSQRT_Perc_African_Am
ZSQRT_Perc_Hispanic ZPerc_Hi_Mobile ZPerc_Disable Small_Enroll Large_Enroll
XL_Enroll Urban SmallTown Rural ZAttendance Schools_SD_PIS ZAvgYrsTeacherExp
ZStu_Teacher_Ratio ZAvgTeacherSalary

/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0,ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)

/FIXED=TIME ZPercDisadvant ZSQRT_Perc_Asian ZSQRT_Perc_African_Am
ZSQRT_Perc_Hispanic ZPerc_Hi_Mobile_dec ZPerc_Disable Small_Enroll Large_Enroll
XL_Enroll Urban SmallTown Rural ZAttendance Schools_SD_PIS ZAvgYrsTeacherExp
ZStu_Teacher_Ratio ZAvgTeacherSalary
ZPercDisadvant*TIME ZSQRT_Perc_Asian*TIME ZSQRT_Perc_African_Am*TIME
ZSQRT_Perc_Hispanic*TIME ZPerc_Hi_Mobile*TIME ZPerc_Disable*TIME Small_Enroll*TIME
Large_Enroll*TIME XL_Enroll*TIME Urban*TIME SmallTown*TIME Rural*TIME
ZAttendance*TIME Schools_SD_PIS*TIME ZAvgYrsTeacherExp*TIME ZStu_Teacher_Ratio*TIME
ZAvgTeacherSalary*TIME | SSTYPE(3)
/METHOD=REML
/PRINT=G SOLUTION TESTCOV
/RANDOM=INTERCEPT TIME | SUBJECT(DistrictIRN) COVTYPE(DIAG)
/REPEATED=Index1 | SUBJECT(DistrictIRN) COVTYPE(AR1).

```