This paper is a working draft from the research of Justin Collins and Jerry Valentine at the University of Missouri. The writers are considering refinement of this paper for submission to a professional journal. Do Not copy or reproduce this paper without written permission from Professor Valentine (requests must be made to Jerry Valentine, Professor Emeritus, University of Missouri at ValentineJ@missouri.edu).

Through the use of a series of disparately constructed Hierarchical Linear Models, the present study attempted to statistically determine the extent of the relationship between student engagement levels within schools that elected to incorporate the Instructional Practices Inventory (IPI) and the frequency with which student engagement data were collected and analyzed. The relationship between student engagement levels and school practices and processes was then considered. Influential site-level variables, in addition to the schools practices and student classroom engagement levels associated with the IPI, were also tested in relation to schools' standardized test achievement levels. All told, the affect of the IPI was tested by treating the student engagement data captured at the building level as a phenomena that is nested within districts and regions. Ultimately, good faith efforts to collect and analyze student engagement data within schools yielded heightened levels of student engagement in such schools. Additionally, the findings from the study were, with some exceptions, suggestive of a noteworthy relationship between higher and lower-order student engagement levels and standardized test achievement.

Introduction

Many questions involving instructional improvement efforts in today's public schools remain unresolved. Distinctly clear, however, is the reality that full and sustained changes to the nature and quality of classroom instruction is a tall challenge that many educational leaders have yet to fulfill. This paper attempts to strike a methodological balance between offering practical findings for educators while giving ample statistical consideration to building-level practices and processes whose complexities must be tested according to the busy and interrelated instructional efforts found within schools.

Hierarchical Linear Modeling (HLM) involves a statistical methodology that facilitates a whole-picture exploration of instructional reform. That is, the wide assortment of educational inputs at the building, district, and regional levels are likely to influence a faculty's ability to positively alter student engagement and achievement levels over time. The HLM models capture such scattered and layered environmental complexities in a mathematical form that is workable for not simply researchers, but practitioners as well.

The HLM models themselves provide a general framework to be more fully specified by the specific school, district, and regional-level inputs that are expected to influence engagement and achievement. Ultimately, this paper becomes more intently focused on exploring whether, and if so, how much, these engagement behaviors impact resulting standardized achievement levels. Before such a relationship can be completely understood, the factors that contribute to student engagement levels within buildings must be investigated. Finally, test scores, an educational outcome on the minds of all instructional leaders, can be tested in relationship to standardized engagement levels. Blunter questions that explain whether a relationship exists between classroom behaviors and standardized achievement levels can be further pursued by testing the discrete forms of desirable and undesirable student engagement levels in connection with the test scores that follow.

It is important that instructional leaders not lose sight of instructional quality in the accountability haze. Adopting instructional improvement programs can enable teachers to remain continuously diligent of the quality of classroom learning and instruction. The Instructional Practices Inventory (IPI), an observational instrument that enables the nature and extent of students' classroom engagement to be codified and measured, represents an especially attractive tool for school leaders to monitor instructional environments and facilitate changes in student cognitive engagement. The IPI is a school improvement process that allows classroom observers to record the nature of student engagement in each classroom within a school. As these observations, comprised of multiple, brief "snapshots" of student engagement levels, are then aggregated over a day-long collection cycle, this allows school leaders and faculty to create and study a telling statistical profile of the overall student engagement levels within the school.

The IPI was designed to be an informative rather than an evaluative tool by which to gauge the nature and extent of student learning across public school classroom. Concern existed that school leaders, accustomed to such practices, might view the IPI as yet another punitive measure. Indeed, it becomes especially important to remember that the IPI is not a process that can be readily and appropriately incorporated within schools in a one-size-fit all fashion. As a result, the IPI implementation survey was designed, in part, to capture the extent to which the IPI initiatives were implemented with fidelity or instead misapplied by school leaders or otherwise misdirected, signifying low levels of implementation fidelity. As "…many principals and teachers oppose performance monitoring," the researcher was diligent in acknowledging the extent to which "performance monitoring can serve the needs of policy makers charged with ensuring that all students receive a quality education while at the same time supporting the day-to-day work of principals and classroom teachers" (Willms, 1999, p. 474). Such a consideration also addresses the essence of the IPI and the rationale of the present study: simply adopting a school improvement mechanism to monitor and better the educational setting is of little benefit if these instructional treatments are not widely, and at least somewhat warmly, embraced. Hence, change initiatives that are incorporated within K-12 educational learning environments, but are not effective in accomplishing their desired effects, may be highly dependent on the organizational practices within schools. Ultimately, the IPI process offers the appropriate instructional supports needed by schools to enhance instructional change.

Prior Research Findings

1a: Employing Empirical Means that Address the On-the-ground Reality

Public schools are busy places. The statistical study of public schools is a complex and oftentimes thorny undertaking. Beyond the statistical noise associated with a vast and oftentimes heavily interrelated set of research variables is the spatial consideration of data. Indeed, students, and their respective schools, do not operate in a vacuum, and the wider contextual environment must be accounted for in statistical tests.

School-wide initiatives similar to IPI improvements have been demonstrated to take three to five years to successfully implement (Raudenbush & Bryk, 2002; Zhang, Shkolnik, & Fashola, 2005). Zhang, Shkolnik, and Fashola, (2005) found that student achievement improved slowly during the first two years of a curricular initiative, but then accelerated rapidly in the third year onward. As the level of implementation fidelity increases exponentially over time, the achievement growth might not concomitantly increase, but rather experience a lag time before manifesting similar exponential magnitudes of achievement growth/gain (Zhang, Shkolnik, & Fashola, 2005).

Hierarchical Linear Modeling represents an especially attractive methodology upon which to address such theoretical concerns. The structurally and spatially nested nature in which student learning and school processes are configured is duly accounted for by HLM Models, as the extent to which the IPI gains traction at the building level (as evidenced by IPI survey response scores) can be tested as independent variables that are embedded, or "nested within," the district level in which school leaders must execute the tenets of the IPI. While the methodology itself may be of little interest to school leaders or policymakers, the interactions of the many complex and oftentimes confounding building level variables may prove to be of far greater salience to such an audience. A three-level Hierarchical Linear Model can account for classroom engagement that is nested within distinctive districts and Regional Professional Development Centers

(RPDCs). Exceptionally large standardized factor loadings need not be evidenced to establish significant empirical findings. Marginal increases in student achievement are always a welcome prospect in public education, especially if such findings are evidenced under the condensed time horizon of approximately three years.

1b: The Present Study: Whole Picture Statistics

Very complex problems are rarely explained by simplistic mathematical devices. Though it does little good to employ empirical methodologies that abandon parsimony in favor of unneeded degrees of statistical complexity, flimsy modeling can only be expected to yield flimsy results. An attractive feature of HLM involves the ability to test the complexities found within school buildings while exploring the outcome in terms that fully inform both the educational research community and instructional leaders in search of pragmatic answers to pesky building-level problems.

For the purposes of this study, a number of HLM models are constructed to investigate the relationship between educational inputs and student engagement levels. Moreover, the impact that classroom behaviors might then exact on standardized achievement levels are fully tested. Within this paper, the researcher seeks to definitively establish whether such a relationship exists within the broader sample of public schools of all types.

2: Inputs: Demographic Influence on Engagement and Achievement

The present study is grounded in the supposition that student learning is an input that schools can directly attempt to manipulate. (Phelps & Addonizio, 2006). As such, it is postulated that the socio-economic status of students, along with other demographic variables, might be a hollower indicator than is conventionally presumed. Indeed, the level of student engagement and the nature of such engagement is perhaps the factor that is most controllable by school districts, schools, and ultimately teachers.

Clearly, a certain funding threshold must be surpassed for schools to provide essential instructional and more general education functions. Considering only the magnitude of funding associated with school improvement planning is an inadequate metric, however (Walberg & Fowler, 1987). Indeed, budget differences were not found to account for students' educational performance. Instead, it might be more fruitful to focus on the incentives for allocating school budgetary outlays (Hanushek, 1995). It is also important to consider the socioeconomic status (SES) of the student population when "…measuring the possible independent effect of per student educational expenditures and size of enrollment because numerous studies show that children from families of high SES generally do better on achievement tests than children of lower SES" (Walberg & Fowler, 1987, p. 5). Remarkably, a school population's family structure evidenced a much stronger relationship than the socioeconomic status or the racial composition of the school's student population. (Caldas & Bankston, 1999). As a students' family structure is perhaps the most uncontrollable factor of a school and school district, this implies that a school district's ability to directly control student test performance levels might be more muted than initially presumed. It should also be noted that the proportion of students who receive free and reduced lunch (FRL) has also been found to be significant determinant of student test performance in the expected direction (Boscardin, Aguirre-Munoz, Stoker, Kim, Kim, & Lee, 2005; Witte & Walsh, 1990).

2: The Present Study: Literature-Informed Model Testing

The structure of the HLM methodology provides a mechanism that can accommodate the full complement of practices, processes, and input variables that unfold in school buildings that undertake instructional reform. While many studies for many years have tested the effects of traditional input variables on student achievement, the quantitative research efforts that test these educational inputs in relationship to corresponding student engagement levels has been entirely lacking. Two points are to be made on this count. First, important contributions to student engagement and standardized achievement can be made by testing both the traditional, as well as other influential inputs, in relationship to student engagement levels. Such a formulation offers the fullest contribution to a research base where such findings are unexplored. Efforts to statistically link discretely identified classroom behaviors with resulting achievement will, therefore, offer a very rich addition to the student engagement literature.

3: Higher Order Thinking

The current literature is largely bereft of studies that empirically probe whether students within those schools where faculties provide higher-order instruction to actively engage students perform better on standardized tests than students in schools with mostly teacher-directed pedagogy. Intuition suggests that instructional practices that encourage critical and higher-order thinking skills, while providing students with sufficient background content knowledge, should enhance students' test-taking abilities. Research confirms that critical and higher-order thinking skills can be taught to students (Lawson, 1993; Marzano, 1993; Tishman & Perkins, 1993). However, new curricular and instructional initiatives that include efforts to promote higher order-thinking within classrooms do not unfold spontaneously, but rather require organizational learning (Valentine, 2005).

Students can engage in varying degrees of critical thinking. An advanced form of critical thinking that is termed “higher-order thinking” incorporates desirable aspects of complex student learning, such as abstraction, extrapolation, and conceptual synthesis (Geertsen, 2003; Lewis, 1978; Underbakke, Borg, & Peterson, 1993). Higher-order thinking is an important intellectual enterprise for all learners. To promote higher-order thinking skills, it is imperative that teachers not simply didactically convey factual information to students. Whether on standardized tests or in their professionals lives, students will be forced to think critically and creatively (Geertsen, 2003). Higher-order thinking can be equated with a more exacting form of critical thinking (Cotton, et al., 1989; Lewis, 1978, Underbakke, Borg, & Peterson, 1993). Lewis (1978) constructs a useful definition of higher-order thinking, in which he suggests that “higher-order thinking occurs when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers to perplexing situations” (Lewis 1978, p. 136). Such critical and higher-order thinking processes are not innate or readily embraced student practices. Instead, effective teachers who provide instruction to high achieving students have been found to be more likely to engage their students in critical thinking and problem solving (Brophy, 1990). Moreover, tracing this observable manifestation of learning to quantifiable achievement outcomes presents a straightforward means of studying instructional improvement in today’s public schools.

3: The Present Study: Incorporating Student Engagement

Most teachers apply the “I know it when I see it” standard to diagnosing their instructional quality. These split-second visual inspections, instructional leaders believe, enables them to accurately gauge the nature of student engagement within their classrooms. Snap judgments from the eyes of educators are insufficient in enabling these instructional leaders to accurately measure classroom engagement. It first becomes important to more meaningfully segment student engagement behaviors into categories that clearly signify desirable classroom behaviors from those that are destructive to the mission of providing instructional excellence. Quantifying the engagement levels to determine the proportion of desirable and undesirable classroom behaviors leaves faculties more meaningfully informed. Educators are, therefore, given a better sense of the shape such instructional practices assume within classrooms. As a result, they can more definitively gauge how much work remains to be done. As importantly, discrete engagement levels can be tested alongside identifiable fluctuations in resulting test scores to allow for a fuller picture of the relationship between student engagement and achievement over time.

4: Effective Schools

Amassing evidence as to what constitutes effective schooling more generally, and good teaching in particular, is far from a trivial undertaking. It is important to determine those traits that define and characterize quality teaching and to ensure that such practices occur in all classrooms, as student learning is strongly impacted by the quality of such teacher instruction (Druian & Butler, 1987). Much like differentiating a school’s contribution to student learning from other contributing sources of student achievement can be exceedingly useful in determining effective school practices, so too can narrowing the research focus to teacher effects on student learning (Druian & Butler, 1987). Students become more actively engaged in the learning process, for instance, when teachers stress relevance in their curriculum (Brophy, 1990). Caldas and Bankston (1999) defined an effective school by the demonstration of academic excellence, as manifested by student test performance. Certain irreplaceable components of effective schools might be documentable, as it is often the case that “schools that hold high expectations for their students and maintain an orderly learning environment realize higher student achievement scores on standardized tests” Henderson, Buehler, Stein, Dalton, Robinson and Anfara, p. 64, 2005).

Common features of effective schools can be gleaned from the literature (Wilson, 2007). For instance, it is in effective schools that school leaders prioritize the curricular and instructional objectives for the school (Wilson, 2007; Cohen, 2007). The research appears to suggest that initially focusing on achieving modest, yet demonstrable gains within the first twelve months of the change effort is advisable (Cohen, 2007). In the current accountability era, schools can afford to dedicate neither the time nor the effort to loose experimentation with curricular practices in their attempts to best and most effectively educate students.

Effective school initiatives can be successfully introduced within schools largely because such schools are evolving organizations, rather than static institutions that are unable to change (Bolman & Deal, 2003). Effective schools are not a pre-existing institutional phenomenon. Instead, organizations must gather institutional knowledge, and organizational leaders must incorporate such knowledge into desirable procedural routines.

4: The Present Study: Charting Outcomes

No study on instructional improvement is complete without a full consideration of standardized achievement levels that result from building-level inputs and practices. While instructional leaders are interested in making their classrooms centers of educational excellence, they must also remain mindful of the importance that pedagogical quality has on resulting standardized achievement levels. It is not likely the case that student engagement levels will exhibit an undetectable influence on test scores. Also implausible is any supposition that varied forms of student engagement will yield similar outcomes for student populations. While students who are turned onto learning will likely benefit on test day, tuned out learners will likely pay a price for their inattentiveness. The magnitudes of the relationship helpfully informs instructional leaders across the nation. As the linkage between engagement type and test score fluctuations becomes reduced to hard numbers, teachers will be informed and motivated to most fully promote what works within the classroom. All the while, they can steer clear of what doesn't in formulating their planed instructional behaviors.

Methods

Research Questions:

To empirically explore whether student engagement and achievement levels are statistically related requires a statistical modeling process that is multifold. Considering the student engagement-achievement relationship without investigating the contributory components that impact engagement levels is misguided and misleading. As such, the researcher first sought to consider whether a nexus exists between student engagement levels in schools that elect to incorporate the Instructional Practices Inventory process and the frequency with which IPI data were collected and studied. After answering this research question, further statistical tests were performed to ascertain the relationship between the student engagement levels that result after the IPI treatment has had time to take effect and the student achievement levels on Missouri's standardized tests (the MAP). The two research questions are, therefore, as follows:

1) Does a statistically significant relationship exist between the frequency with which student engagement data are collected and analyzed according to the prescribed IPI instructional treatment protocols and the measurable student engagement levels captured by classrooms observations?
2) Does a statistically significant relationship exist between student engagement levels within classrooms, as measured by IPI observational data, and the standardized test performance of the students within such schools?

This study, as a holistic empirical approach to accounting for the student engagement and achievement relationship, amounts to a two-pronged empirical consideration of the IPI treatment schools with both (1) the relationship between student engagement levels and IPI data collection and analysis, and (2) the extent to which such engagement levels evidence a statistical relationship with standardized test performance levels. By first establishing the extent to which instructional treatments influence student engagement levels, the researcher can make a more active determination of the linkage that might exist between IPI treatment adoption and student achievement levels. It next becomes useful to consider the implications of not simply the nature of classroom engagement levels, but the consequent standardized test performance of students that inhabit such instructional environments. In the rigid accountability era, a consideration of the value of the IPI as a low-cost school improvement initiative is of particular interest to school leaders at a time of resource paucity.

School leaders' acquired fluency with the IPI tenets is believed to galvanize higher-order student engagement levels. Furthermore, the relationship between both the higher and lower-order student engagement levels and standardized test achievement levels is suspected to dictate student achievement levels, a postulation that can be most definitively addressed by employing Hierarchical Linear Modeling (HLM).

The IPI Instrumentation

The Instructional Practices Inventory (IPI) is a process employed by classroom observers to ascertain the nature of student engagement across classrooms within a school. The IPI is comprised of "a set of observational categories complex enough to provide substantive data grounded in the knowledge of best practice (valid) yet easily understood and interpreted" (Valentine, 2007). The IPI instrumentation allows a trained classroom observer to collect scores of observational codes that capture student engagement behaviors for each school. The observation categories included in the IPI observation protocol are: (1) student disengagement, (2) student engagement in non-higher order activity without teacher participation, (3) student engagement in non-higher order activity with teacher support, (4) teacher-directed instruction, (5) student engagement in higher-order classroom discussion, and (6) all other higher-order student learning.

One of the more complex methodological challenges presented by the present study is not formulaic in nature, but rather involves the adequate and accurate definition of student engagement levels and what constitutes higher and lower ordering thinking.

Such attempts to delineate meaningfully nuanced distinctions between various types of student engagement can quickly become obscured and fruitless if student engagement behaviors are hyper-parsed by the researcher. The IPI strikes a methodologically appropriate balance between meaningfully categorizing student engagement categories without deconstructing the classroom environment to such an extent that coding the minutia of student behavior becomes an untenable task for the classroom observer. More importantly, as the categories become more numerous (and indistinguishable), the reliability of such classroom observations can become greatly diminished. With this in mind, the Instructional Practices Inventory categorizes student engagement levels on a continuum from 1 to 6, which is designed to account for the spectrum of student engagement that one can expect to find in any given classroom at a particular moment.

Table One offers an explanation of each of the six coding categories. It is important to note that while the higher-order categories ("5" and "6") represent desirable forms of student learning whereas the lower-order categories ("1" and "2") represent less effective and generally undesirable, indefensible forms of student activity within classrooms, it is not always possible, nor desirable, for students to be engaged solely in higher-order activities. As such, categories "3" and "4" account for those moments during classroom instructional time when the teacher is primarily involved in informing and directing the students' activities in the classroom, as student engagement becomes mostly passive and inactive. This might come in the form of teachers informing students of certain tasks or logistical considerations or teacher-directed learning, both of which are inevitable components of effective teacher pedagogy and student learning.

Insert Table 1 approx. here

The IPI Survey

This exploratory quantitative study will be largely driven by the data yielded from classroom walk-through observations that employ the Instructional Practices Inventory instrumentation (IPI) and the corresponding data collection protocols. The IPI-trained data collectors conduct classroom observations in which they ascertain both the nature and the proportion of higher-order versus lower-order student learning that occurs within classrooms. IPI school profiles are generated from the trainers' observation data, which ultimately provide a holistic statistical depiction of student engagement levels within schools. Such IPI profile data will allow the researcher to determine if a relationship exists between the level of higher-order and lower-order thinking and schools' exposure to the IPI treatment and the vigor and integrity with which IPI treatment tenets evidence enhancements in the nature of students' classroom engagement levels. The IPI data provided by interviewees includes a statistical profile of the nature and level of student engagement within a school (Valentine 2007; 2008). The IPI profiles provide a statistical representation of engagement, including whether students are inactive, are engaged in knowledge acquisition with or without teacher attention, are the recipients of didactic teaching, or are engaged in higher-order thinking and reflection **(**Valentine, 2005; 2007; 2008).

The IPI Survey

The IPI implementation survey questionnaire serves as the primary instrument by which to capture data about the nature of the implementation of the IPI process. The IPI survey was constructed by the researcher to capture several environmental factors demonstrated to directly affect student performance. More specifically, the IPI survey reveals the frequency and processes by which school leaders both collected and analyzed IPI-coded observational data.

The survey questionnaire items of interest in the present study were principally designed to capture the number of times and the duration that the IPI practices were executed within a given school (Questions 4 and 5, respectively) (Valentine, 2007; 2008). The frequency and duration of IPI practices undertaken within schools serves as a proxy for the extent to which the schools value academic achievement (Hoy, Tarter, & Hoy, 2006). For the purposes of the present study, the extent to which the IPI is implemented with integrity is assumed to be a robustly telling proxy that manifests other features of the school's culture and climate (Valentine, 2007; 2008). More specifically, a good faith IPI implementation effort can be equated with a school's desire to ensure that a challenging climate of academic excellence exists at the building level (Hoy, Tarter, & Hoy, 2006).

Statistical Procedures

Statistical Model Configurations

Hierarchical Linear Modeling (HLM) represents an especially attractive methodology upon which to address the theoretical concerns underlying Research Questions One and Two of this study. The structurally and spatially nested nature in which student learning and school processes are configured can be duly accounted for by HLM Modeling, as the extent to which the IPI gains traction at the building level (as evidenced by IPI survey response scores) can be tested as independent variables that are embedded ("nested") within a district level in which school leaders must execute the tenets of the IPI.

After investigating the more isolated and narrow consideration of school practices and processes associated with the IPI treatment on engagement levels, it becomes necessary to consider the site-level variables and their contemporaneous interactions with one another in a more holistic manner. The researcher began testing for the IPI treatment's influence on the practices and processes within schools on student engagement levels by employing a two level structure. Such a model most realistically captures the wider contextual environment in which the student engagement levels are evidenced in, and influenced by, the respective school districts in which students are situated.

For the purposes of the present study, the engagement within classrooms among schools that have incorporated the IPI process will be used as a starting point to accumulate the data needed to address the extent to which student engagement levels are altered as a result of IPI implementation, while also investigating the student engagement and standardized test performance relationship. To adequately account for the nesting of student engagement within classrooms in the broader environmental context, the introduction of a third level to the model that incorporates the region level (level three) can additionally be considered by the researcher as he attempts to account for the structure inherent in student learning. Furthermore, knowledge of more elaborately constructed HLM models enabled the researcher to immediately evaluate the proportion of variance explained among each of these levels to determine whether a parsimonious pruning of entire levels of the HLM Models is warranted.

<u>Level-One School-wide engagement</u>: Level-One of the model represents the student engagement levels within the classrooms. Raw percentage breakdowns are computed for each of the 139 schools that provided three or more IPI classroom data profiles, in the form of singular disengagement codes for core and total classrooms (C1, T1 and C2, T2), higher order and an aggregated metric of distinctive higher order categories C5 and C6 (T56 and C56). As multiple classrooms observations are coded for each classroom with the building over the course of a school day, a statistically representative depiction of student engagement levels within the school can be introduced into the multilevel statistical study at Level One. The assignment of student engagement levels as dependent variables in the model to test against the corresponding IPI practices and processes, as captured by coded IPI survey responses, has been emphasized at this level.

<u>Level-Two School Districts</u>: School districts comprise the second level of the multilevel statistical study that incorporates student engagement data from within and across classrooms. The schools that provided IPI classroom data were located within 63 distinctive school districts across the state. While not categorically the case, anecdotal evidence and more cursory observations suggest that school districts exhibit a pronounced and inescapable influence on the health and effectiveness of the schools that operate within them. The demographic data provided by Missouri’s Department of Elementary and Secondary Education (DESE) is quite exhaustive. For the purposes of

this study, traditional socioeconomic, and controllable and uncontrollable educational resources and input factors were collected and recorded for the corresponding school districts containing the schools that provided data for the current research undertaking.

Level-Three Regional Professional Development Centers: Not unlike many states across the nation, Missouri is comprised of several disparate regions. Impoverished urban centers in Kansas City and St. Louis are surrounded by more affluent suburban districts, where stronger standardized test performance levels are reflective of these socioeconomic and demographic endowments. In addition to the two metropolitan areas, suburban regions of the state give way to a more scattered arrangement of small towns/cities.

To determine whether a nexus exists between student engagement levels and the practices and processes associated with the IPI, the data collected from IPI walkthrough observations were subjected to Hierarchical Linear Modeling (HLM) analysis employing HLM 6.4 software. Such student engagement levels represent a variable that the school leaders are able to directly influence (Lee & Weimer, 2006). The Hierarchical Linear model will account for within school (Level-1), within district (Level-2), and within region (Level-3) data nesting effects.

To analyze student achievement at the school and district level, a two-level HLM model was constructed, using the following variables:

1. Level-one (within school) predictor variables:
    a. The annual disciplinary infraction rate

b. The IPI student engagement categories (C1, T1, C2, C56);
c. Teacher experience, as defined by years of instructional experience;
d. The traditional metric of socioeconomic status, Free and Reduced Lunch (FLR).
e. Ethnicity (calculated as the percentage of non-Caucasian students)

2. The Level-two (within district) variables:
a. District-level Free and Reduced Lunch Rates
b. Per pupil expenditures of the districts
c. Average Daily Attendance Rate within the districts
d. Percentage of families below the poverty line
e. Percentages of students within the district that attend post-secondary institutions
f. District-level teacher experience, as defined by years of instructional experience

Finally, a third level of the achievement-engagement model was established to test whether significant differences exist when incorporating the regional level to the school and district-level model:

3. The Level-three, within-region variables were:
a. Regional-level Free and Reduced Lunch Rates (FRL)
b. Regional-level percentage of families under the poverty line;
c. Regional Ethnicity (calculated as the percentage of non-Caucasian students)
d. District average ACT scores.

The student engagement outcome of greatest interest for this study is the percentage of higher-order student engagement in core content area classrooms coded as a "5" or a "6" (AV_C56). The construct was assigned as the dependent variable in both the two and three level models. Conversely, the non-higher order student engagement levels were also tested as dependent variables, to ascertain not only fluctuations that result from independent variable manipulations, but to compare any fluctuations to their higher-order counterparts. To test this theoretical proposition, the percentage of classrooms coded as either a "1" or "2", (either student disengagement (AV_C1 or teacher inattentiveness, AV_C2, within core classrooms) was assigned to be the dependent variable in HLM Model. Ultimately, the student engagement and achievement relationship can be more thoroughly and holistically explored by testing data under a HLM statistical framework. Simply put, the HLM models enable the researcher to determine the extent to which the IPI more directly influenced student engagement levels, which might, in turn, also exhibit influence on standardize test score levels of schools.

A Brief Note on the IPI Practices and Processes Fidelity Levels

As the practices and processes associated with the IPI have been delineated and elaborated upon, it is important for the reader to appreciate that the IPI treatment fidelity is defined as the measurable level of data collection and analysis which is demonstrative of a broader commitment to the IPI treatment processes. The IPI survey responses crafted to identify specific building level information regarding the IPI implementation were summed, resulting in a cumulative, holistic picture of the fidelity of the IPI process. Thus, discretely coded variables served as the IPI fidelity input variable.

Explanation of Population Sample and Descriptive Data

The population of this study met two basic criteria. First, only Missouri public schools are included in the data set. Second, school leaders in the study group must have attended an Instructional Practices Inventory (IPI) workshop and subsequently employed the IPI methodology within their schools. (Valentine, 2005; 2007; 2008).

In 2005, numerous schools across Missouri and the nation began to conduct IPI classroom walkthroughs. These walkthroughs enable the level of student engagement in each classroom within a school to be captured and documented by a trained observer. At the time of this study, approximately 300 Missouri public school utilize the Instructional

Practices Inventory with some degree of fidelity. The eventual population size of 163 schools offers evidence of a robust response rate to the electronic data collection period.

Data Collection Procedure

To collect IPI data, a certified data collector moves continuously from (classroom to classroom throughout the school day, observing and coding between 120 and 170 observed classroom behaviors. Two points are to be stressed at this point as it relates to the trained IPI classroom observers who conduct classroom walkthroughs: First, teacher and school leaders other than principals are designated as data collectors to diminish the possibility of bias in data collection or concern about the instrument as a mechanism for supervision or evaluation. Second, all IPI data collectors are to have an IPI reliability measure of .90 on a post-workshop assessment.

Designing HLM Models to Address Research Question One

Research Question One is principally concerned with testing whether a statistical relationship exists between the quantifiable magnitude with which the IPI is implemented over time, captured by the raw number of data collections and analyses by school leaders, with the extent to which student engagement levels are impacted. Research Question One represents a sound and necessary starting point, as Hierarchical Linear Modeling captures the nested nature whereby school practices and processes unfold not only at the building level but within the districts which have decided to implement the IPI treatment. At the district level, demographic and input variables are of interest, as these variables might influence the very potency and durability of the IPI treatment.

The researchers concluded the consideration of the engagement-process relationship by testing for the IPI treatment's influence on the practices and processes within schools on student engagement levels using a three level structure. Such a model realistically captures the wider contextual environment in which the student engagement levels are evidenced. Furthermore, inspecting the intraclass correlations (ICC) of the three-level HLM models enabled the researcher to immediately evaluate the proportion of variance explained among each of these levels to determine whether a parsimonious pruning of the regional level of the HLM Models is warranted in future studies.

Ultimately, six models were constructed to test for evidence of a relationship between IPI practices and processes and student engagement levels. The model testing was ambitious in scale due to the exploratory nature of the study. The most extensively-tested models were the two-level models that capture school-level practices and processes at level one and the corresponding district-level variables representing a wider context in which such practices unfold. The primary dependent variables of interest were higher order ("AV_C56") and non-higher order disengagement ("AV_C1" and "AV_C2" constructs).

Designing HLM Models to Address Research Question Two

Research Question Two is principally concerned with testing the relationship between the higher and lower order student engagement levels within classrooms, as captured by the IPI observational data, with the resulting standardized test performance of these schools. It was postulated that the good faith implementation of the IPI will materially elevate higher-order student engagement levels and depress lower order-student classroom activities. Such findings are noteworthy, as it is not only the student engagement-school processes relationship that is of interest but also the student engagement-achievement relationship which can be more narrowly considered by models tailored to address Research Question Two. Were an empirical nexus to be established between such heightened levels of intellectual classroom engagement and standardized test performance, school leaders would have little reason not to actively promote such curriculum and pedagogical practices across all classrooms (Marzano, 1993; Nickerson, 1988; Underbakke, Borg, & Peterson, 1993).

Analysis

A summary of the descriptive statistics associated with the variables incorporated at Level-One, Level-Two, and Level-Three of the HLM models that capture every school which provided data for the current study appears in Table Four. As is evidenced from the HLM descriptive output, 243 schools contributed data that were nested within the 105-district level variables, which were in turn nested within 9 RPDCS. It is worthwhile to compare the student engagement levels within the population of the present study with both the successful and ineffective schools described in the 2004-2005 Tables. The large and representative sample of 243 schools that provided IPI data for the present sample manifested student engagement levels that are presented in Table Two.

Insert Table 2 approx. here

Analyzing the Statistical Findings

The findings were substantial, with several centrally relevant findings to the student engagement-instructional treatment nexus. Namely, it was determined that the percentage of Free and Reduced Lunch predictor (FRL) exhibited a magnitude that was fully three times more pronounced at the district as opposed to the building level.

As data collection and analysis are continuous events, slopes for both the disengagement and higher order thinking are both considerable, especially when one considers that these slopes might become even further enhanced as school leaders acquire greater fluency with the IPI process. Survey question 4 (data collection frequency) and survey question 5 (data analysis frequency), were tested in isolation of one another for a host of reasons, but most notably because of collection analysis disparities of both follow-up on collection periods (for example four collections with only one accompanying data analysis) and the quality of such interactions. It appears to be the case for both lower and higher order thinking that the analysis of the data can exact more appreciable enhancements to student engagement levels than mere iterative data observation/collections.

The variance across schools and districts between student engagement collection frequency and data analysis was sufficiently marked to warrant parsing each survey question into independent HLM models. Data yielded by Questions 4 and 5 of the IPI Survey enable the researcher to ascertain the number of times collections have been conducted within the school. Such information can be explanatorily powerful in revealing the extent to which the frequency of data collections impact student engagement patterns but might additionally serve as an integrity proxy, as an appropriate frequency of data collections evidences administrative diligence within public schools.

Accounting for Level One and Level Two Variance

Bryk and Raudenbush (1992) demonstrate the usefulness of the computed sigma-squared and tau values in explaining the percentage of variance that is accounted for by the HLM models and their respective predictors. Testing student engagement levels within the classroom by nesting classroom data within districts and schools represents a sound starting point for such in-depth empirical exploration. It is the case, however, that the educational processes undertaken at the school (the site level) also warrants empirical investigation. The classroom environment is inextricably linked to the wider building-level and the processes therein. For the purposes of this study, the integrity and fidelity with which the Instructional Practices Inventory treatment is adopted and cultivated within a school, using Questions 4 and 5 as powerful proxies for implementation fidelity, enable the researchers to ascertain the extent to which the interplay of effective school-level processes and effective classrooms is manifested in the current population sample. The school-level predictors explained relatively little of the overall variance at this level (ranging from 2.81-9.87%). The variance explained at the district level was nonexistent in some instances, while it was as high as 66.38% in others. As for the three-level model, while the predictors did not account for any of the models variance at level one, it accounted for 31.81% of level two variance, and fully 100% of the level three variance of the model.

Ultimately, the integrity and fidelity of the IPI implementation practices and processes demonstrated a positive statistical relationship with student engagement levels, as good faith efforts to adhere to the prescribed methods of IPI adoption within schools yielded heightened levels of student engagement. Additionally, the findings from the study were, with some exceptions, suggestive of a noteworthy yet moderate relationship between higher and lower-order student engagement levels and standardized test achievement.

Impact on Student Achievement

The empirical findings yielded by the HLM models constructed to test for a relationship between student engagement levels and the collection and analysis of data appears to be significant. Such results are also corroborative of the presupposition of a demonstrable relationship between the vigor of data collection and analysis and higher and lower order student engagement levels. As such, it next became desirable to determine whether student engagement levels were, in turn, statistically related to standardized test achievement, allowing the relationship between Instructional Practices Inventory (IPI) and the MAP test performance of schools to be more definitively broached. While FRL was slightly more augmented at the district level than the school level, the magnitude of the FRL at the building level was fully four times larger when achievement was assigned to be the dependent variable than was the case when student engagement was assigned as the outcome variable. This squarely comports with intuition, as the FRL rate of a school population has been definitively established as a contributing factor to student achievement, but should not dictate either instructional pedagogy or, more specifically, student engagement with nearly the same intensity, a finding that is brightly illustrated in the current models.

Full Model – Mathematics

The fully specified HLM Mathematics Model, which assigned 2007 Mathematics achievement as the dependent variable, was exhaustively specified with school and district-level predictors to determine which of these educational input factors may evidence a statistically significant influence on mathematics standardized achievement levels. The skeletal model of the fully specified HLM Mathematics models is depicted in Figure One below. An inspection of the computational output reveals that the percentage of disciplinary infractions (DISCIPLI), free and reduced lunch students (FRL) and percentage of classrooms with higher order student engagement in core classrooms (AV_C56) were found to be significant. At the district level (level two), the percentage of students that attend postsecondary institutions (TO_COLLE) and the percentage of students that receive free and reduced lunch were found to be statistically significant, as well. Consequently, these predictors, determined to be significantly linked to mathematics achievement, were incorporated in subsequent HLM models that assigned mathematics achievement as the dependent variable.

Insert Figure 1 approx. here

Several two-level HLM models were outfitted with predictors that were found to be statistically significant within the fully specified model that tested the relationship between the nature and levels of student engagement and test performance. Additionally, three-level hierarchical models were employed to test the relationship between the higher order thinking dependent variable, AV_C56, and the educational input independent variables. The three-level HLM model might appear, on its face, to be unwieldy, and more likely to confound the effects of the distinctive independent variables. Sympathetic to this argument, the researcher constructed the model to be both appropriately parsimonious in scope and discriminating in terms of independent variable inclusion.

Only the variables revealed to be the most highly-impactful independent input variables, as evidenced in the full model, were included. The researcher was, therefore, assured of methodological rationality rather than a capricious exploratory undertaking.

Accounting for Implementation Integrity Differences

It is methodologically desirable to differentiate between the varying degrees of fidelity with which the IPI is employed at the building level. Indeed, schools that more scrupulously adhere to the tenets of the IPI should evidence augmented higher-order thinking and depressed lower-order thinking levels. Furthermore, the impact of student engagement on achievement might be disparately impactful, depending on the corresponding instructional treatment implementation fidelity levels. For the purposes of the achievement-engagement models, high integrity, low integrity, and the entire population sample of schools in the study was incorporated within the HLM model configurations. The distinction between high integrity, low integrity, and all schools in the population sample is as follows:

<u>High Integrity</u>: High-integrity schools are defined as those schools whose faculty responses on a global question of the IPI survey designed to capture the benefit of the IPI as a school improvement instructional treatment was suggestive of its current and anticipated ameliorative impact to the school.

<u>Low Integrity</u>: Using the same global survey question as for high-integrity schools, low integrity schools are those whose staff suggested the IPI was either met by faculty intransigence, was incorporated in a manner that strayed from the prescribed tenets of the

instructional treatment, or was implemented in good faith but did not achieve its desired effect over the course of the school year (s).

Insert Table 3 approx. here

The magnitude associated with the student engagement predictors was found to be fully twice as great for schools that were identified as “high integrity” implementers as opposed to the wider set of all IPI schools across Missouri. Of perhaps greater interest is the glaring discrepancy between the magnitudes of higher order and lower thinking predictors. Lower order thinking predictors exhibit magnitudes that are elevated more than three times greater than the higher-order thinking predictors. While further investigation of student disengagement and higher order thinking is certainly warranted, such findings are not unexpected. Indeed, the very nature of the standardized testing instruments designed to measure student and school achievement and effectiveness are not tailored to reward creative or higher-order expression in many instances. The content material presented to students requires attentiveness and attention to content area detail, however. That is, were students entirely disconnected during the course of the school day, the likelihood that they ultimately perform satisfactorily on standardized tests could be severely diminished. Contrastingly, higher-order student conversations and classroom activities, while epitomizing robust levels higher-order thinking, might do little to benefit students on testing instruments fashioned to reward “skill and drill” practice, which would not be assigned a code of “5” or “6” during IPI classroom observations.

The explanatory power of the predictors at both levels one and two were demonstrably greater in the Achievement-Engagement models than in the Engagement-Treatment Processes models. More specifically, the level one variance explained by the Achievement-Engagement predictors ranged from 9.96% to 33.45%. The level two variance explained was more uniformly robust, ranging from 64.07%-86.87%. The three-level model, while not accounting for any of the level one variance, accounted for the entire variance at both level two and level three.

Of further note is relatively unimpressive capability of the most fully specified model to explain the measured variance at each level. This leads the researcher to conclude that an empirical accounting of all school and building level factors is an implausible objective. A comparison of the variance explained by the fully specified model was greater than in the lesser-specified models. But not strikingly so, however, lending additional support for the parsimonious construction of HLM models. School and district- level educational inputs are notorious for introducing confounding and multicollinear influences on one another. As such, sacrificing slight enhancements in explained variance for a mitigated potential for these confounding effects is highly desirable.

Achievement-Engagement-Communication Arts

The achievement-engagement relationship was similarly tested with communication arts designated as the dependent variable this time. Again, and as depicted in Figure Two, a two-level model was constructed that employed a fully-specified panoply of predictors at both the school and district levels. The statistical significance of the FRL predictors at both the first and second levels were not as strong as was the case when mathematics was assigned as the dependent variable. The coefficients associated with the total percentage of disengaged students fixed effect also exhibited a more depressed magnitude than its mathematics counterpart. Such a finding is perhaps attributable the nature of the communication arts test which, at least to a limited extent, captures students' prior knowledge bases and innate skill sets. The mastery of the mathematics segment of the MAP test, on the other hand, most likely requires attention to more detailed, formulaic instruction. It was further the case, however, that lower-order thinking coefficients were fully three times greater than their higher order thinking fixed effect coefficients. Among other deductions to be made from such a finding is the universally pernicious effect that student disengagement evidences on standardized test achievement, irrespective of the subject matter tested. Furthermore, such a finding appears to suggest that in starkly different content areas, the effects of lower and higher order thinking on student achievement, remain nearly identical in magnitude.

Insert Figure 2 approx. here

While the lower-order thinking magnitude on achievement was less for Communication Arts than for mathematics, the coefficients on the higher-order thinking constructs bore a stronger resemblance to one another when the communication arts and mathematics dependent variables were tested. The variance explained with the Communication Arts models, shown in Table Four, was also slightly less than was the case for the Mathematics HLM models. More specifically, between 1.84 and 19.4% of the level one variance was explained by the assigned predictors, while 26.16%- 72.38% of the variance was explained at the district level (level two of the models). It was again the case that the fully specified model, comprised of an exhaustive list of school and district level demographic and input variables, did not provide an overwhelmingly greater explanatory power of the models' variance.

Insert Table 4 approx. here

What is quite evident from an inspection of both the mathematics and Communication Arts HLM output is that not only is there strong statistical evidence to support the existence of a demonstrable nexus between student engagement and student achievement, but such a relationship is sufficiently robust to suggest the importance of the IPI in facilitating school leaders in their mission to provide students with excellent and challenging learning environments as they also prepare these to succeed as they sit for standardized tests.

Discussion

It is most informative to first discuss the findings according to the relationship of building-level inputs on student engagement levels. After this key relationship is established, the empirical discussion can then shift to how these engagement levels impact standardized achievement scores. In short, HLM output reveals that student engagement behaviors can be greatly manipulated under workable time schedules of instructional reform. With the benefit of these findings, the feasibility of such reforms can be spelled out in more precisely quantified terms, both for collection cycles and the length of time these efforts ultimately require. To entirely eliminate student disengagement "1" by student engagement observations alone, 23.25 collections are needed. At a quarterly rate of observation collections, this would take school leaders 5.8 years to eliminate such undesirable engagement behaviors in the public school setting.

Were the student engagement data analyzed and discussed by instructional leaders, 12.86 analysis cycles, or 3.21 years are then required to rid school classrooms of student disengagement. Doubling the percentage of higher order thinking in core classrooms (C56) could prove to be an immensely powerful psychological boost to school faculties. For faculties to be successful in this endeavor, they would be required to undertake 11 years of data collection. The time span is reduced to 7.5 years were schools leaders to also analyze student engagement data. In all likelihood, the combined effect of data collection and analysis probably considerably inflates the 7.5 year projection. Put differently, the sheer act of taking stock of student engagement can yield more ideally optimal levels over sufficiently sustained time horizons.

Fluctuations of student engagement levels in the wake of instructional reforms, show the findings, leave an appreciable mark on standardized achievement, as well. As importantly, the findings make clear that disengagement can depress achievement levels with considerably greater impact than higher order thinking acts to elevate them. Also established is that diligent instructional leaders, who promote higher order thinking and eradicate lower order thinking, will see substantial test score growth that signifies enhanced levels of instructional quality. All the while, schools will gain considerable momentum in aligning their achievement performance with prescribed accountability levels.

Evidence of the universally deleterious impact that student disengagement (“1”) on mathematics achievement was confirmed by the study. Noteworthy is the finding that a 10 point increase in total disengagement would lead to a 4.8 point decline in mathematics achievement for the public schools included in the study. Likewise, a 20 point increase in total disengagement would stage a 9.6 point decline in mathematics achievement for these same schools.

High integrity implementers’ achievement levels are not only significantly impacted by total disengagement levels, but on a magnitude of over twice that of the entire population sample of schools. More specifically, a 10 point increase in total disengagement would lead to a 10 point decline in mathematics achievement for all schools in the study. A 20 point increase in total disengagement would lead to a 20 point decline in mathematics achievement for all schools in the study.

To enable meaningful comparisons between fidelity implementation levels, statistically significant findings would need to be evidenced for high, low, and all IPI implementers. Unfortunately, only the entire sample of schools evidenced a significant predictor value for higher order student engagement in core classrooms (AV_C56). Doubling the value (roughly a twenty-five point increase in this observed value) would lead to a 3.5 point increase in mathematics achievement.

Important findings were unearthed as it related to educational inputs, as well. The aggregated FRL predictor values were considerably greater for low- integrity implementers as opposed to all schools in the population sample. Indeed, in comparison to the entire population sample, a low integrity IPI implementer with a 50% FRL population would evidence 10.5 point lower achievement levels than the entire sample.

Communication Arts achievement levels were affected to a more muted extent than was the case for mathematics achievement. For all schools in the population sample, a 10 point increase in total disengagement would lead to a 3.7 point decline in mathematics achievement for all schools in the study. More strikingly, when total disengagement comprises 20 percent of all classroom behaviors, a 7.4 point decline in mathematics achievement results.

High integrity implementers were again impacted to a greater extent than the wider population sample. For these high integrity implementers, a 10 point increase in total disengagement would lead to a 6.8 point decline in mathematics achievement for all schools in the study. Moreover, a 20 point increase in total disengagement would lead to a 13.6 point declines in mathematics achievement for all schools in the study.

Again, statistically insignificant higher-order predictor coefficients were evidenced for high and low IPI implementers. For the entire sample of schools, however, significant predictor values for higher order student engagement in core classrooms (AV_C56) were discovered. In fact, enhancing the value by an ambitious 40 points would lead to a 4 point increase in Communication Arts achievement, using a conservative estimate, and an 8.8 point increase according more complex model designs.

Student engagement matters not only in impacting the quality of student learning, but also how well public schoolchildren perform on test day. To date, many studies have crafted quite specific suggestions of what student engagement should look like.   Stated differently, quantifying the sorts of engagement behaviors that constitute educational excellence have not been empirically charted by scholars. Far fewer efforts to quantify the fluctuations in classroom behaviors on the basis of controllable and uncontrollable school inputs have been considered. Not surprisingly, the literature on the engagement-achievement nexus is also entirely lacking.   As a result, this paper fills an important gap in the current educational research on student engagement and achievement. Though the statistical methods employed were somewhat complex, they were also the most amenable to yielding findings that starkly illuminate what matters in governing student engagement levels. How these resulting fluctuations in classroom behaviors can be expected to impact resulting student achievement levels can then be answered, as well.

The purpose of this study was two-pronged: First, the researcher sought to ascertain whether the levels of higher-order thinking and active student engagement within schools were statistically related to the fidelity with which the Instructional Practices Inventory treatment was incorporated within schools. Second, the researcher explored whether enhanced student engagement was, in turn, statistically related to concomitantly augmented standardized test score performance of these schools.

The results of this study were compelling on several counts and should shed additional light on the Missouri Department of Elementary and Secondary Education's demographic data reporting practices. The findings from this study should also allow for informed postulation as to which demographic, curricular, and pedagogical variables are most instrumental in directly impacting student learning and achievement. When considering any of these variables, school leaders must remember that higher-order thinking represents a conscious decision and not a direct pecuniary cost. As such, school district administrators are unable to disqualify any prescriptive suggestions solely on the basis of a funding deficiency rationale. As school, district, and state-level leaders focus on how to best utilize resources and best educate children, this will affect the lives and prosperity of students and the greater society that they will eventually inhabit.

References

Bolman, L., & Deal, T. (2003). Reframing organizations: artistry, choice and leadership. San Francisco, CA: Jossey-Bass.

Boscardin, C. K., Aguirre-Munoz, Z., Stoker, G., Kim, J., Kim, M., & Lee, J. (2005). Relationship between opportunity to learn and student performance on English and algebra assessments. *Educational Assessment*, *10*(4), 307-332.

Brophy, J. (1990). Teaching social studies for understanding and higher-order applications. *The Elementary School Journal*, *90*(4), 351-417.

Caldas, S. J., & Bankston III, C. L. (1999). Multilevel examination of student, school, and district-level effects on academic achievement. *The Journal of Educational Research*, *93*(2), 91-100.

Cohen, J. (2007). Evaluating and improving school climate: Creating a climate for learning. *Independent School, 67*(1), 19-26.

Cotton, K. (1989). Critical thinking across the curriculum: Aloha High School. *School Improvement Research, Series III*, 57-61.

Druian, G., & Butler, J. A. (1987). *Effective schooling practices and at-risk youth: What the research shows*. Portland, OR: Northwest Regional Educational Laboratory.

Geertsen, H. R. (2003). Rethinking thinking about higher-level thinking. *Teaching Sociology*, *31*(1), 1-19.

Hanushek, E. (1995). The quest for equalized mediocrity: School finance reform without consideration of school performance," in Lawrence O. Picus and James L. Wattenbarger (ed.). *Where does the money go? Resource allocation in elementary and secondary schools* Thousand Oaks, CA: Corwin Press, pp. 20-43.

Henderson, C. L., Buehler, A. E., Stein, W. L., Dalton, J. E., Robinson, T. R., & Anfara, J., V.A. (2005). Organizational health and student achievement in Tennessee middle level schools. *NASSP Bulletin*, *89*(644), 54-75.

Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal, 43*(3), 425-446.
Lawson, A. E. (1993). At what levels of education is the teaching of thinking effective. *Theory into Practice, 32*(3), 170-178.
Lee, T.P. (In Press). Using systems thinking to improve organizational learning in the public sector: Perspective of public officials. Thunghai University: Taiwan.
Lewis, E. S. (1978). Teacher influence on pupil cognitive performance. *Peabody Journal of Education, 55*(3), 252-264.
Marzano, R. (1993). How classroom teachers approach the teaching of thinking. *Theory into Practice*, *32*(3), 154-160.
Missouri Department of Elementary and Secondary Education (2008). Website. http://dese.mo.gov
Nickerson, R. S. (1988). On improving thinking through instruction. *Review of Research in Education, 15*, 3-57.
Phelps, J. L., & Addonizio, M. F. (2006). How much do schools matter? A production function approach to school accountability. *Educational Considerations, 33*(2), 51-62.
Pritchett, L., & Filmer, D. (1997). What education production functions really show: A positive theory of education expenditures (pp. 1-48): World Bank.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 11). Thousand Oaks, CA: Sage.
Tishman, S., Jay, E., & Perkins, D. N. (1993). Teaching thinking dispositions: From transformation to enculturation. *Theory into Practice, 32*(3), 146-153.
Underbakke, M., Borg, J. M., & Peterson, D. (1993). Researching and developing the knowledge base for teaching higher order thinking. *Theory into Practice, 32*(3), 138-146.
Valentine, J. W. (2005). *Instructional practices inventory: Profiling student engagement for school improvement.* Columbia, MO: Middle Level Leadership Center, University of Missouri (www.MLLC.org).
Valentine, J. W. (2007). *The instructional practices inventory: Using a student learning assessment to foster organizational learning*. Columbia, MO: University of Missouri (www.MLLC.org).
Valentine, J.W. (2008) Middle Level Leadership Center (MLLC) Website material published by Director Jerry W. Valentine. University of Missouri.
Walberg, H. J., & Fowler, W. J. (1987). Expenditure and size efficiencies of public school districts. *Educational Researcher, 16*(7)5-13.
Willms, J.D. (1999). Basic concepts in hierarchical linear modeling with applications for policy analysis, in Gregory J. Cizek (Ed.). *The Handbook of Educational Policy*. San Diego, CA: Academic Press.
Witte, J. F., & Walsh, D. J. (1990). A systematic test of the effective schools model. *Educational Evaluation and Policy Analysis, 12*(2), 188-212.
Wilson, L. (2007). Great American schools: The power and culture of passion. *Education Horizon, 86*(1), 35-44.
Zhang, Y., Shkolnik, J., & Fashola, A. (2005). Evaluating the implementation of

comprehensive school reform and its impact on growth in student achievement, *American Institutes for Research.* Washington, D.C.