# Playing Charades: Failures, Fads, and Follies of School Improvement Plans

**David C. Coker**
*Fort Hays State University*

Most public K-12 schools in the United States enact a school improvement plan each year based upon the state testing results from the previous year. Despite the many frameworks, research, and recommendations, few scholars tackled what the plans contain and the results. Using a sequential explanatory mixed methods model, two stages examined school improvement plans for language arts tests over three years: a general linear model for 1316 schools in two states and a longitudinal thematic analysis of plans within one district. Most plans produced minor improvements in test scores, and most schools developed generic, nondescript plans to improve teaching practices. The school improvement plans failed to create substantive, visible change in most schools. Instead of the usual recommendations, a simple view of improvement finds schools need to focus on three criteria at the local level: students, curriculum, and highly effective teachers. Each criterion has several subvariables, but school improvement plans should move beyond a focus on student failure and seek to improve all students.

*Keywords:* school improvement, strategic leadership, educational administration, thematic analysis, mixed methods

Each Spring, most public K-12 schools in the United States partake in a familiar dance: A principal, with an advanced degree and often decades of experience, convenes a small group of staff members termed stakeholders to develop a school improvement plan (SIP) for the following year. State test results dictate the initiatives, with thousands of books, consultants, and a large bureaucratic framework guiding the would-be leaders. The cycle repeats year after year. Billions of dollars, many surefire plans, and even more in work hours tackle a problem which persists despite the best-laid plans (Backstrom, 2019).

Fads and fancies pushed much of the school improvement process since the 1980s (Alvy, 2017; Crandall et al., 1986). Over $7 billion dollars alone was spent on school improvement grants which produced little gains and even less direct connection with the boondoggle (American Enterprise Institute, 2017). Lofty goals and feel-good sentiments guided much of the advice and studies, but the gap was few empirical studies demonstrated the efficacy or sustained rigor of school improvement plans (Fernandez, 2011; Harris, 2001; Strunk et al., 2016). There remains a lack of agreement in what comprises an effective school improvement plan, with teacher agency—the boots on the ground—often missing from the discussion (Datnow, 2020; Thompson, 2018).

The literature review examined school improvement plans and three popular approaches. A sequential explanatory mixed methods investigation starts with quantitative analysis to examine the longitudinal results of school improvement plans over three years in two states. Then a longitudinal thematic analysis dissects the nature and focus of school improvement plans for one district over the same time period. A discussion and recommendations converge the results of the studies.

## Literature Review

There is a cottage industry in school improvement, with recommendations on every aspect of how to fix what ails schools, from frameworks to curriculum to leadership and practically everything in between (Duignan, 1986; Leithwood et al., 2020; Murphy, 2013; Sebring et al., 2006). For most schools, after the many processes and initiatives, repressive desublimation defined the process, with little changes and results bearing more resemblance to what existed before the regular order of school improvement plans (Coker, 2021a; Rowan, 2002). Three major issues permeated school improvement plans: turnaround initiatives, systems thinking, and research trends.

A more extreme form of school improvement plans focused on turnaround efforts (Carlson & Lavertu, 2018), or radical, holistic changes within a school. Even though there have been some successes, there has been a failure to translate in the face of immutable student characteristics, such as student mobility and absenteeism (Henry et al., 2020; Peck & Reitzug, 2014). Strong leadership, collaboration, and theories galore received glowing recommendations (Evans et al., 2012; Hargreaves, 1995; Hitt & Tucker, 2016; Ross & Gray, 2006), but alas, the initiatives failed to deliver the promises predicated on oversimplification and buzzwords. Leaders in educational administration programs make pronouncements of being data driven, transformational, and many other great sounding, theoretical-laden terms.

Though school improvement plans failed to produce the explicit promises and goals, apologists nevertheless persisted in celebrating the value (Strunk et al., 2016). Little was known

about efficacy (Hitt & Meyers, 2018; VanGronigen & Meyers, 2021), but one common factor was schools were in an endless cycle of improving and professionally developing the same problems and factors year after year. Like many educational practices and strategic leadership initiatives, such as mission statements (Coker, 2022a), practitioners take as a given from their educational professors one must do what one has always done with little evidence.

A central tenet in most school improvement plans was the concept of systems theory (Askell-Williams & Koh, 2020). Leaders claimed school strategic planning utilized systems thinking. Systems thinking purportedly influenced much of the conceptual basis for school improvement. Schools needed to define, understand, and change the different units and dimensions included in school improvement efforts, or so the conventional wisdom stated (Schneider et al., 2017). Though school improvement plans often lacked the sophistication found in the original conception of the theory of systems thinking (Kast & Rosenzweig, 1972; Morel & Ramanujam, 1999; Shaked & Schechter, 2020), leaders recognized the inherent complexity and nonlinear relationships needed mapped (Schneider & Somers, 2006).

The limitations of systems theory originated in the genesis: As stated by von Bertalanffy, systems were open, constantly in flux, and complex which required an organizational commitment to tackle problems (Adelman & Taylor, 2007; Heylighen & Joslyn, 1992; Hopkins & Higham, 2007). Like school improvement frameworks and buzzwords, systems theory lacked a unified definition and cogent research (Arnold & Wade, 2015; Harris et al., 2021) and often became obsequious to fads, sloganeering, and grandstanding. Calls for gaining and using insight from modelling (Arnold & Wade, 2017; Hung, 2008) gave little clear direction, with heuristics probably the true practice of systems thinking (Mintrop & Zumpe, 2019).

There were few large-scale studies in effectiveness of school improvement plans (Bohanon et al., 2021; Browne-Ferrigno et al., 2008; Good et al., 2005; Huber & Conway, 2015), but lack of success did not stop researchers and consultants continuing to believe schools improvement plans can still work if the recommendations were heeded. Much more common, in dissertations, peer review articles, and the popular literature were vignettes and single case studies (e.g., Hollingworth et al., 2018; McIntosh et al., 2021; Redding & Searby, 2020; Tran et al., 2018). There were research-based practices and programs, but the failure to translate led to schools in an endless spiral of school improvement plans.

The research in effectiveness and research-based initiatives provided an information overload, with the challenge of how to implement the different programs and ideas (Cohen-Vogel et al., 2016). Elgart (2017) recounted several key characteristics, such as focus and a sustainable culture, but what or how the characteristics looked remained mysterious. Calls for parental engagement, academic press, professional learning communities, and other quick wins (e.g., Bloom & Owens, 2013; Brown et al., 2017; Cannata et al., 2017; Meyers & Hitt, 2018; Padilla et al., 2020; Solone et al., 2020) sounded good and right when pronounced by experts. Alas, the widespread success failed to materialize (Purkey & Smith, 1983; Redding & Nguyen, 2020).

Universities teach what academics believe school leaders need to know (Tingle et al., 2019), but how the skills translated into practice, both from a content analysis and results over time, was lacking (Preston et al., 2017; Strunk et al., 2016). As stated by Quong and Walker (2010), school improvement required leaders to move beyond a vision and a mission and work toward a better, sustainable future. Quick adoption and abandonment of practices created an endless

cycle of fads and change fatigue with little value (Rohanna, 2017). School improvement should be defined by measures of student achievement.

## Significance and Research Question

Every educational administration program teaches school improvement, whether as a formal program or through instructional evaluation and planning. Longitudinal studies examining the effectiveness of school improvement plans are infrequent (Feldhoff et al., 2016). The results could be useful to educational professors, consultants, and practitioners. There is a gap in both how schools improve and what themes drive planning at the local level.

There were two research questions, which converged. For the quantitative component: Do schools improve from year to year? If so, what was the practical significance? For the qualitative component: What do schools do to improve student achievement in reading? Both questions shed light on the nature of improvement efforts from a macrolevel and explain what was done at the microlevel.

## Conceptual Framework

Two frameworks influenced the research design and discussion: adaptive leadership and balanced leadership. Adaptive leadership stated successful organizations have continual diagnoses, seeks multiple perspectives, honors the past, encourages experimentation, and tests ideas and hypotheses (Heifetz et al., 2009). Balanced leadership examined leadership as the continuum between technical and adaptive; adaptive leadership was necessary for messy, complex situations with ill-defined boundaries (Goodwin et al., 2015). Combined, both theories suggested school improvement oscillates between the complex and the complicated, and each situation possessed unique features which required translation of research and programs with a backwards looking perspective on root causes which informed future practices.

## Methodology

The research methodology was a sequential explanatory mixed methods investigation (Creswell, 2021). First, two longitudinal samples from Washington and Florida were drawn over four years. Secondly, a longitudinal qualitative analysis using thematic analysis of school improvement plans from a large school district in Florida was then conducted to explore and qualify the quantitative results over the same time. The longitudinal qualitative study revealed patterns which could explain the quantitative research; under each step, there is further explanation.

All records were archival and publicly available; since the research did not involve humans, there was no institutional board review approval needed. There were attempts to draw from more states, but there were problems. First, only states which used the same tests over the entire period were considered; there could be problems comparing different tests. Secondly, many states either did not have records available or in an inaccessible format. Thirdly, states had to have robust school improvement plan requirements. Several more states were examined, but they failed to meet the criteria. Still, both states are from different areas of the US and provide a large, rich data set which could be generalizable to the population.

**Study 1: Quantitative Findings**

Using the Washington Department of Education and Florida Department of Education websites, a longitudinal study was conducted for elementary schools for three years from 2015-2018, with 2014-15 as the baseline. The records of 439 Washington elementary schools and 877 Florida elementary schools were collected for language arts scores for K-5 students for comparison to answer the question: Do schools improve from year to year? If so, what was the practical significance? All records were initially downloaded into Microsoft Excel; matching by county and district location, elementary schools with the same name were disaggregated and schools were assigned a code. Using IBM SPSS Statistics for Windows, version 28 (IBM Corp., Armonk, N.Y., USA), descriptive and inferential statistics examined if schools improved from year to year.

Categories were devised to examine the data for a combined sample of 1316 schools. Initially, for the baseline 2014-2015 school year, language arts test scores averaged 52.6 (SD = 16.4) and ended for the 2017-2018 school year averaging 55.8 (SD = 16.4). Yearly scores were also reported as quintiles. Year over year gains were broken down by >-10%, >-10 to -5%, <-5% to 5%, >5%-10%, and >10%. The rationale was -5% to 5% was trivial and due to random error (e.g., 80 students tested would need 4 students to improve or 1 per grade level, while 10% might be by chance but has less likelihood, etc.). The plan included descriptive statistics, correlation analysis, and a general linear model (GLM).

The general linear model examined school improvement results from 2015-2018. The covariate was the 2014-2015 baseline, centered. Examining Table 1 under the Greenhouse-Geisser correction, as sphericity was violated, suggested a statistically significant result. Sullivan and Feinn (2012) cautioned the need for context and to look beyond the *p* value. The partial eta square (0.008) suggested the growth was very minor; the large sample probably produced significant results, but the GLM did not answer the question of practical significance (Ranganathan et al., 2015). Further analysis provided context of what school improvement looked like.

**Table 1**

*Repeated Measures ANCOVA for 2015-2016, 2016-2017, & 2017-2018: Test of Within-Subject Effects*

| Source | | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|---|
| StudentAch | Sphericity Assumed | 3853.356 | 2 | 1926.678 | 37.924 | <.001 | .028 |
| | Greenhouse-Geisser | 3853.356 | 1.645 | 2342.652 | 37.924 | <.001 | .028 |
| | Lower-bound | 3853.356 | 1.000 | 3853.356 | 37.924 | <.001 | .028 |
| StudentAch * Base1415Cent | Sphericity Assumed | 1076.226 | 2 | 538.113 | 10.592 | <.001 | .008 |
| | Greenhouse-Geisser | 1076.226 | 1.645 | 654.293 | 10.592 | <.001 | .008 |
| | Lower-bound | 1076.226 | 1.000 | 1076.226 | 10.592 | .001 | .008 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Sphericity Assumed | 133512.071 | 2628 | 50.804 | | | |
| | Greenhouse-Geisser | 133512.071 | 2161.358 | 61.772 | | | |
| Error (StudentAch) | Lower-bound | 133512.071 | 1314.000 | 101.607 | | | |

*Note.* StudentAch = Student Achievement on state language arts assessment from 2015-2016, 2017-2018, & 2018-2019. Base1415Cent = State language arts scores for 2014-2015 centered.

Three analyses suggested the growth was small, possibly by chance, and insignificant in practice. First, the average growth over three years was 1.03% (SD = 3.21) per year, meaning most schools saw a change of 3.11% (SD = 9.64) in total. In any year, only 8-11% saw improvement greater than 10%. Secondly, there was a yoyo effect, suggesting regression to the mean—growth was sporadic and regressive—with over 90% of schools staying between ±10% from the baseline to the end (and less than 9% were >80% on student achievement in any year, meaning most schools had 2 for every 3 students fail for who succeeded or worse). Thirdly, as shown in Table 2, correlation showed as schools experienced growth in one year, there was negative growth the following year, and the final year approximated the first year. Year-over-year growth was uncommon, as growth for the first two years negatively predicted growth in the final year. Approximately 2.8% of schools had large gains or losses (>25%) after three years. Even trimming 70 outliers using Cook's distance, as suggested by Faraway (2016), produced an effect size which suggested minor improvement: Results, with the Greenhouse-Geisser correction, were F(1.530, 1905.877) = 5.977, p = .006, $\eta^2$ = 0.018.

**Table 2**

*Correlations of growth for 3-year period split by starting quintiles.*

| Control Variables | | | Yr1Gains | Yr2Gains | Yr3Gains |
|---|---|---|---|---|---|
| 1415Quint | Yr1Gains | Correlation | 1.000 | -.531 | -.040 |
| | | Significance (2-tailed) | . | <.001 | .148 |
| | | Df | 0 | 1313 | 1313 |
| | Yr2Gains | Correlation | -.531 | 1.000 | -.203 |
| | | Significance (2-tailed) | <.001 | . | <.001 |
| | | Df | 1313 | 0 | 1313 |
| | Yr3Gains | Correlation | -.040 | -.203 | 1.000 |
| | | Significance (2-tailed) | .148 | <.001 | . |
| | | Df | 1313 | 1313 | 0 |

*Note.* a. Computed using alpha = .05 b. 1415Quint = 2014-2015 Results by Quintile. Yr1Gains, Yr2 Gains, Yr3Gains = Yearly gains from previous year broken down by <-10%, >-10%-5%, >-5% to 5%, >5%-10%, & >10%.

The practical significance was illusory for most schools and suggested regression to the mean. The year-to-year difference revealed small overall gains of just over 1% each year—equivalent for many schools of one student improving. To put the situation in perspective, 46% of schools saw losses in the second year (another 5.8% saw no gains). Only 39% saw an increase of 5% or greater after three years (e.g., schools with losses in 2016-2017 saw 66% have gains the following year, while schools with gains 5% or greater saw 61% have a decline). The yoyo effect defined the sample.

Even a 10% improvement for most schools meant out of 100 students, approximately 3 students per tested grade level improved from the previous year before the gains were erased in the following year; the 3% improvement, for many schools, meant three students out of 100 improved after three years and might literally hinge on the improvement of only three questions (one per student). Noise, not systematic improvement, defined much of what passed for gains. The overall stark lack of success defined most schools in a dismal picture as well as inconsistency in gains. Even gains in one year were temporal. School improvement plans existed in name only and failed to produce continuous, widespread gains. A central question should be: Why do most schools persist as if there were no school improvement plans?

## Study 2: Thematic Analysis

School improvement plans for 16 Florida schools were examined using longitudinal thematic analysis. Using one school district, school improvement plans for elementary schools (K-5) over three consecutive school years, 2015-2016, 2016-2017, and 2017-2018, were selected if the schools had reading goals and action plans for improvements. Schools with missing data or lack of reading goals were excluded. The *a priori* plan was to select at minimum 12 schools, as it was thought 12 schools would provide a representative sample to understand school improvement plans which could be generalized to the population.

The research question was broad: What do schools do to improve student achievement in reading? Following Coker's (2021b) and Fereday and Muir-Cochrane's (2006) recommendations, thematic analysis was conducted using Microsoft Excel and Word. After downloading the school improvement plans, the schools were given a pseudonym (letters A-P), and each year and process were numbered to root the data to be able to return to the original source. Demographic data were collected. To analyze reading goals, the coding schema included in vivo, descriptive, processes, categories, subelements, elements, dimensions, themes, and meta-themes. A meta-narrative transcribed the findings into a story. There were also memos and aha comments as necessary. Temporality and directionality were coded by frequency and who/where implementation of the processes. Dimensions and themes were structured using the ethnographic approach every answer has an embedded question. The embedded question was developed by reverse root cause analysis, or every action was in response to a hypothesized problem and diagnosis. Coding and thematic formation were formed inductively and deductively.

Because there were approximately 10,000 codes generated, reliability and validity were improved by extensive cataloging using number codes. Every code received a dummy code, which allowed for a frequentist approach to ensuring consistency horizontally and vertically of each code. The frequency of codes was run with JASP Team (2021). JASP (Version 0.16)[Computer software] and Microsoft Excel. Using this data, reconciliation of data was conducted to examine if similar codes were analyzed consistently since there were multiple levels of agreement needed for the holistic framework. Any problems led to codes being reworked.

### *Sample*

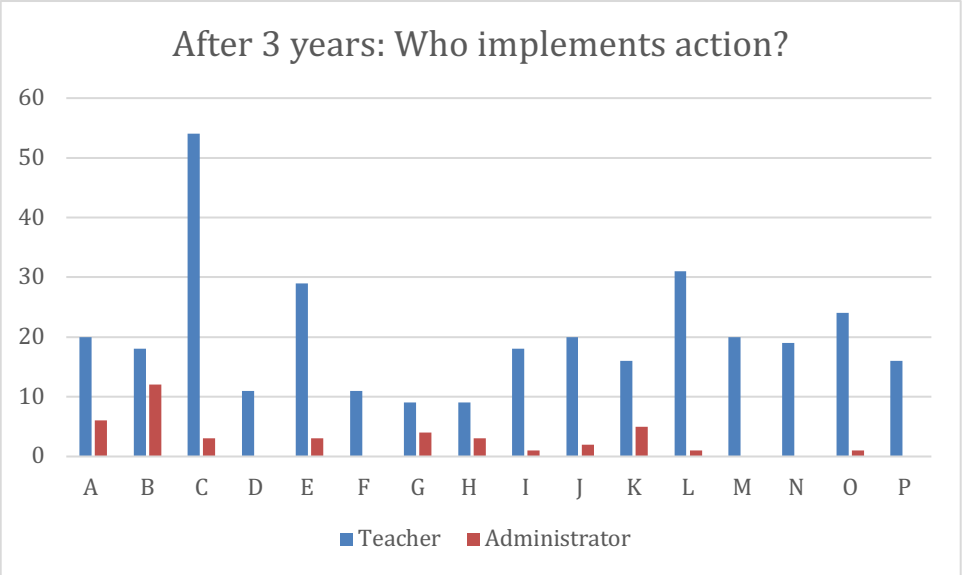
There were 16 elementary (K-5) schools from one large Florida school district comprising approximately 8,150 students (average school size 510, *SD* = 93.9). The schools were diverse,

with 26-100% of students minority and an average of 63.5% low socioeconomic status. A team developed most plans, with a range of 1-12 members (*M* = 8.1; *SD* = 3.9); the principal was on each team, and generally an assistant principal, teachers, and school psychologists/social workers rounded out the membership. The number of students retained was generally low (range of 0-13, *M* = 5.1; *SD* = 3.8). All schools had reading goals and action plans for the 2015-2016, 2016-2017, and 2017-2018 school years.

***Results***

The 16 schools, over a 3-year period, planned 366 actions in total. Yearly actions per school ranged from 2 to 34 per year. At the end of the third year, schools averaged 22.6 actions (range 11-57; *SD* = 11.5), of which 20.31, on average, fell to teachers and less than 3 to administrators. As shown in Figure 1, schools placed most of the burden on teachers, with a daily expectation of a change in instructional practices being the major driver. There was never any mention of abandonment, which leads one to conclude each action was in addition to the previous year's school improvement plan.

**Figure 1**
*Implementation of Actions for Each School by Role*

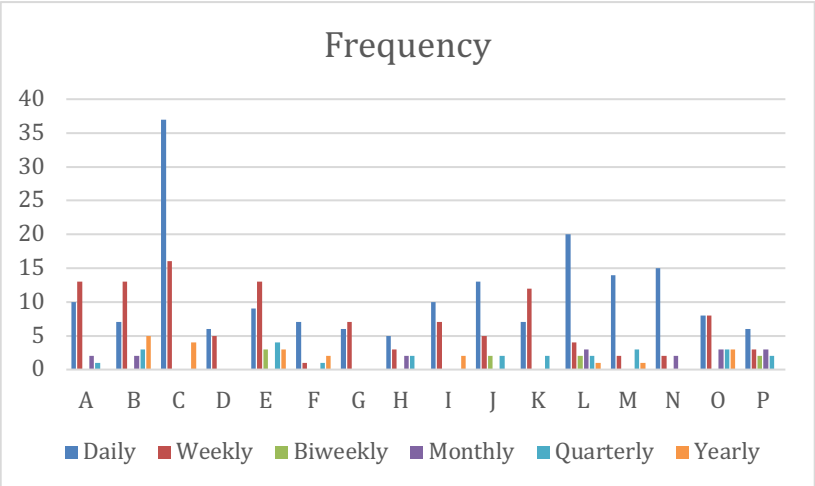



Frequency of actions was broken down from daily to yearly, as shown in Figure 2. As an aggregate, schools put forth action plans which required teachers to implement different actions on a daily or weekly basis. Much less common were long-term goals, such as quarterly and yearly requirements. Broken down yearly, schools on average implemented 5 to 8 new teacher actions per year, every year. Administrators had much fewer expectations. There was no mention of

when or how actions were monitored except by observation of the administrators and, or coaches.

**Figure 2**

*Frequency of Actions in SIPs by School After 3 Years*




Three themes defined the plans: T1) Schools cannot deliver the necessary curriculum due to the following: lack a coherent, standards-based curriculum, an optimal schedule, and sufficient staff (frequency = 20; 5.5%); T2) Behind the Scenes: Teachers need professional development and regular collaboration time to develop effective instructional techniques, intervention and differentiation, and processes for data-driven instruction. (frequency = 140; 38.3%); T3) Classroom Level: Teachers lack effective teaching strategies and interventions with proper supervision. (frequency = 202; 55.2%). As shown in the Appendices A-D, a complete catalog of categories (C1-99) with sub-elements (SE1-99), elements (E1-99), dimensions (D1-99), and themes (T1-3) connect every component of the analysis together.

Of the themes, T3 dominated action plans by a wide margin. The two dimensions directly related to instruction, D2 (staff lack effective, research-based strategies aligned to state standards to engage students; 145 count; 39.6% of all dimensions) and D3 (there is not enough staff or time, but staff members lack the skills to correctly intervene anyway; 19; 5.2%) dominated T3. The top suggestions for D2 were the following: guided reading, small groups, vocabulary instruction, and independent reading. For D3 within T3, feedback was the major goal to improve teacher performance; D4, lack of meaningful interventions, was closely behind. There were many other recommendations, from planned physical movement, culturally relevant instruction, and mini-lessons. T3 overwhelmingly expected teachers to directly change classroom practices on a regular basis, as poor instructional practices needed improved to raise student achievement.

The second theme sought to develop teachers through collaboration, professional development, and common assessments. The thought seemed to be teachers lacked the skills and knowledge with which to adequately perform instruction at a high level, even after receiving

a bachelor's degree at minimum and many having a master's degree. For T3, three dimensions stood out: professional development or PD (D5), professional learning communities or PLCs (D6), and data-driven instruction (D7). Professional development centered on using instructional coaches and standards, while PLCs had general advice on how to operate or focused on data-driven instruction.

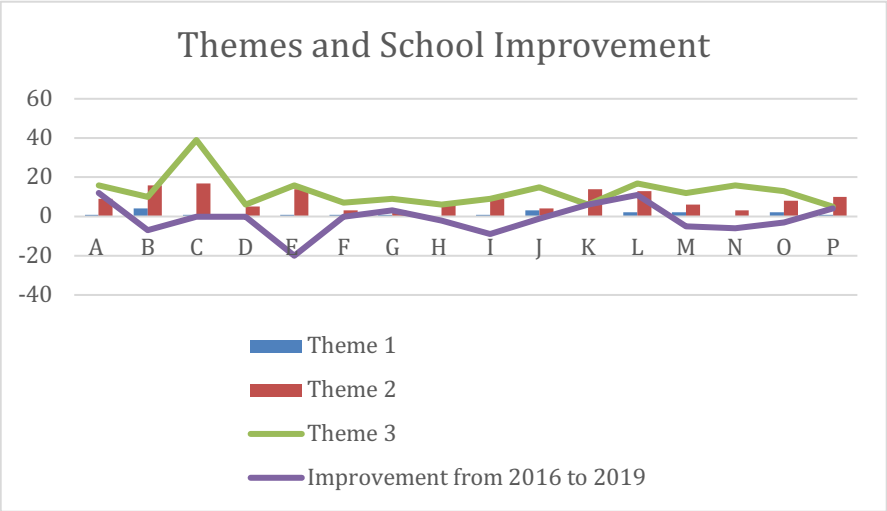
Schoolwide, T3, was infrequent and suggested supervision/lack of leadership and misaligned curriculum (D1) caused poor reading gains. The crossover of dimensions justify why themes were logically separated the way they were. All three themes discussed standards-based curriculum and instruction, but the focus was different for each theme. T1 looked at aligning the curriculum before the school year; T2 sought either PD or PLCs for teachers to work throughout the year aligning standards; T3 required teachers to align instructional content with the curriculum All three had one dimension, but the implementation and responsibilities varied widely.

Another example of the splitting of dimensions by themes was data-driven instruction (D7, D3, and D6). At the school level, some schools used the previous year's standardized testing to plan groups and interventions before the school year started. Behind the scenes, PLCs and less frequently PD instructed teachers to come together to analyze data and implement plans accordingly. Still, other plans pushed data-driven instruction down to the classroom level and either wanted decisions in real time or teachers to develop formative assessments to respond flexibly.

By and large, schools showed either negative or little improvement, as shown in Figure 3. During the 3-year period, 11 schools had either negative or zero results, and three schools had results which could be counted as trivial. Results around 5% or less might have been noise; since the results do not measure the same cohorts, there could be the natural variance in small numbers, i.e., 5% is only 2-3 students per grade level for a building with two teachers per grade level, which could be the change with new students. Two schools had results >10% (A and L), but a comparison to elements, categories, dimensions, and themes did not give any direct effect. There must be a difference which the SIPs did not account. For the three years, the change was up and down for every school except three which showed consistent gains year in and year out, as shown in Appendix E. Figures 1 and 2 do not provide much insight beyond change by chance.

**Figure 3**

*School Improvement Results After 3 Years of SIPs Compared to Total Number of Each Theme.*




***Meta-themes***

There were three meta-themes which applied to the results and cut across every theme: activity over value, the re-creation and repackaging of the status quo, and the lack of data-driven decision making. School improvement plans had a common connotation and denotation, but the results after three years showed most schools did not improve, and of the schools which did, there was no clear connection to any action or goal. School improvement plans were a rite of passage, required each spring to build the illusion of improvement.

Activity over value detailed no risk in any action. Everything was a certainty, as there was no risk in any action. The activity was the objective, with no measurement or performance indicators of anything except implementing a policy or procedure in hopes of improving student learning. From guided reading to a standards-based curriculum to PLCs and PD, doing stuff was what defined all improvement efforts. Even before anything started, as long as the policy or program was implemented, success was guaranteed. SMART goals were not so smart after all.

There was little evidence any action was any more than a continuation of the status quo repackaged; student learning be damned, as the quality was destined to remain the same. Assuredly, every school used small groups, PLCs, PDs, coaches, and the many other actions before the current plans. There were no adaptations, experimentation, or innovation in all activities. Doing what one had always done with the protestations of a different result defined the norm rather than the exception; instead of a change at the most basic level, SIPs were largely surface level without any substantive improvements.

Many actions called for the use of data to guide and drive the curriculum, instruction, and interventions. Yet, there was not a connection of any action to a specific data point. The results revealed there were no diagnoses or actions rooted in the specifics. Every action was generic and lacked nuances which should define the plans. The achievement of individual students, classrooms, and teacher efficacy should provide a great deal of variation suggesting actions localized and differentiated by the units of analysis (e.g., if one teacher had significantly higher scores than another teacher, there should be an investigation to see what could be done both within and between classrooms, etc.). Nested variance could tell a more complete story, especially with small numbers.

Consilience—or the convergence of multiple perspectives and disciplines—was missing. There was neither systematic nor systemic considerations, as there were no questions or connections to proximal and distal causes. Schools wrote documents to appease a bureaucratic system by going through the motions to make lofty goals with a plethora of actions decidedly lacking in calls for improved student learning. SIPs lacked a conjunctural analysis of the components and drivers of the schools under study. The result was schools aligned perfectly with the unstated goal: no change, no improvement, carry on with the status quo.

A *single-state theory* connected all themes. All students—in all classes, all teachers, all grade levels, in all schools—lacked any divergence. What was needed for one was needed for all. To use a hospital analogy, every diagnosis was the same. There was a familiar narrative. Get 8-10 staff members together. The principal already possessed the necessary knowledge and skills, often imparted from the central office. Teachers needed to teach better. Administrators needed to keep on their employees. Sprinkle some buzzwords, and PLCs, collaboration, and data-driven would cure that which ails. Everything was a quick fix, and repeat the same dance year in and year out.

### *Reliability and Validity*

Reliability and validity in research can be enhanced and established by producing a transparent record of the qualitative process (Carcary, 2009; Creswell & Miller, 2000; Rodgers & Cowles, 1993). There was a clear coding schema, with findings rooted and grounded in the data. By producing an extensive cataloging system, the data were verified and cross-checked both horizontally and vertically for logic and consistency (as shown in the appendices) throughout the process.

Becker and Geer (1957) pointed out long ago inferences were not what happened, and not only can there be mistakes and mistaken assumptions, the further one gets from the data, the less reliability and validity. Up to themes, the conclusions and categories were closely connected to the data. The meta-themes developed form an intimate knowledge of both academic ideas of strategic leadership and school improvement and a practitioner's experience with school improvement, both direct and vicariously. Still, by producing a thorough record, the reader can see how the conclusions were formed and find alternatives which could provide an explanation.

One way to enhance reliability and validity is to compare and contrast with previous findings. Meyers and VanGronigen (2020) found a similarity in root-cause analysis: Those darn teachers need to do a better job. Other researchers found calls for clarity, data-driven instruction, and improved instructional practices (Gonzales et al., 2020; Leithwood et al., 2019), but a contradiction existed in the present study. Beyond very general SMART goals, there was never any mention of actually being data-driven. Nowhere did anyone have specific processes and products to direct and align activities; there was homage to the different buzzwords, but the activity was the objective.

Raining a plethora of solutions for problems not defined and offering a blanketed, one-size-fits-all approach dominated most school improvement plans (Meyers & VanGronigen, 2019; Scherer & Nilsen, 2019; Slavin, 2017). The plans, all from one large district, had significant overlap, where often the same goals appeared across multiple plans. Truly innovative or

individualized ideas were not to be found. There was an agonizing oversimplification, generally rooted in improved teachers and teacher behavior. Strikingly, what and how students would do something differently were missing.

## Discussion

Around this part of most research articles, recommendations permeate the discussion. New goals and objectives, replete with a well-developed, *novel* framework, litter tens of thousands of dissertations, articles, books, and consultants' documents. All the plans and frameworks offer promises of success and improvement; the stark reality is research is often overstated and practically everything studied claimed effectiveness (Coker, 2022b; Evans, 2022). Canned plans and trivial gains—often erased year to year—defined a lack of progress by a group of planners with a great deal of experience and advanced degrees looking either from the outside or at a distance (Alvy, 2017; Coe, 2009). Most school leaders believe they are Atlas when they are really Sisyphus. Expediency often defined what was done (Harris, 2000; Meyers & VanGronigen, 2019). Three main drivers offer improvement: jettisoning business as usual, developing an antiframework, and a simple view of improvement (SVI).

What if schools eliminated professional development? What if there were no school improvement plans? What if consultants, instructional coaches, and collaboration time were eliminated? What if programs, such as RtI, MTSS, and PBIS, were done away with? Why do leaders recycle fads and buzzwords endlessly? The reality was, for most teachers, few would see or do anything differently. Worse yet, for most students, there would be nothing visible as a drop in student achievement already existed (Camera, 2021). A magical realism defined plans, with a bluntness revealing the drivel: false sense of stakeholders, group work as a panacea, and defining student behavior downward. The paradox of calls for data-driven instruction existed, yet one saw no data in the plans or recognition that some classrooms were successful already and do not need an umbrella of reform. Doing and changing—not achieving—supplanted the goals of school improvement plans.

Plans were and are designed to focus on activities over achievement, taking a forced perspective of the clutter of disconnected ideas from a low angle. Focusing on several connected ideas, the background was blurred, and a high angle, forced perspective—the bird's eye view—was uncommon. Without a bird's eye view, plans were neither systematic nor long range. There was no measurement of any failure, and any leader could ask themselves: How did schools improve, and how does one know? The myths of school improvement drive the status quo, some of which are listed:

- Myth of professional development. Schools know what effective professional development entails, but few do the necessary work, monitor results/implementation, and often end with a net negative (Desimone et al., 2002; Garet et al., 2001; Guskey & Yoon, 2009; Heissel & Ladd, 2018; Kraft et al., 2018).
- Myth of data driven. The SIPs in this study, and one could generalize to most, called for data-driven instruction but lacked the data-driven aspect within their own plans. Using data proved unworkable and a waste for most teachers at the classroom level (Gleason et al., 2019; Hamilton et al., 2009; Neuman, 2016; Schildkamp, 2019).

- Myth of instructional coaches and leadership. Research suggested what was highly effective and ineffective in teaching and learning, and group work was not a factor in the effective part (Rosenshine, 2012; Stockard et al., 2018). Teacher leadership, instructional coaching, and instructional leadership have not produced widespread positive effects in student achievement (Hallinger et al., 2020; Ingersoll et al., 2018; Moody, 2019; Wong, 1997).

There are other myths. One push was to avoid nonexclusionary discipline and act like teachers could manage extreme behaviors while simultaneously teaching and maintaining the wellbeing of everyone in the classroom, but promoting incivility, disruptive behaviors, and violence erodes the already-mentioned slim chances of school improvement plans (Chambers Mack et al., 2019; MacNeil et al., 2009; Polanin et al., 2021; Shindler et al., 2016). All tie into the learning environment, but the calls for research-based practices struggle with competing demands.

The antiframework operates on three premises: teachers--not administrators, professors, or consultants—are responsible; highly personalized, highly contextualized improvements; and systematic not systemic drives change. Magical realism controls most school improvement plans: Get together, implement what the central office and principal want, influenced by college staff members and consultants who have no accountability, and after crossing one's finger, repeat the following year while taking credit for any improvement (no matter how small and trivial) and erasing continued failure. Cookie-cutter approaches and *pro forma* actions must be jettisoned.

Teachers, as a group, lack meaningful chances to participate in school leadership, to receive promotions, and give way to a singular leader at the top (Chambers Mack et al., 2019; Timperley, 2005). Not surprisingly, lack of control was a factor in teachers leaving the profession (Madigan & Kim, 2021). Sham stakeholders act like teachers have a say, but the bureaucracy and trends drive much of what passes for teacher involvement. Administrators mistake placing responsibility as a singular factor and not a shared purpose (Connolly et al., 2019). The teachers are the most knowledgeable, experienced ones to know and understand the issues and needs facing students, the classroom, and the school. Furthermore, teachers will be the ones to implement any proposed changes.

A highly personalized, highly contextualized school improvement process was missing from the literature and the current studies. Dixon and Palmer (2020) pointed out improvement depends on failure and tackling the problem; nowhere did they or anyone else mention a blanketed, one-size-fits-all approach. Instead of promoting plans which label every student and classroom with the same ailments, schools must also improve students who already succeed as well as highly effective classrooms. Not only should school leaders recognize and reward successes—instead of a complete focus on failure—school improvement plans *must* adapt the radical idea of two strands: How to improve students and teachers who were failing (Benoliel & Berkovich, 2020) while also improving those who were succeeding. The lived experiences and marks of success within the schools should be part of the roadmap to whole school improvement.

Stoll (2009) called for a shift away from generic recommendations to capacity building. The antiframework includes a vision, mission, and goals (Meyer et al., 2020) as well as research of the extant literature (Wallace et al., 2001) and a rejection of the status quo (Betts, 1992), but the central axes transform the external frameworks and integrate them within the existing system. Systems thinking is futile and incomplete once analyzed, but systematic is the ability to pick out the criteria which matter. While many frameworks will be applicable, schools will have

local contexts and issues which both hinder what need done and externalize the problems at the detriment of student success.

The simple view of school improvement can be reduced to the following equation: *student learning x rigorous curriculum x effective teacher = school improvement.* There are many subvariables and factors, but schools need to stop acting like only teachers are the problem. Drift is a real concern; a central goal of all elementary schools is reading competency, yet SIPs get inundated with all kinds of *disponsible-par hasard-objets trouves* (Dougherty & Weiner, 2019). Schools must define their domains of inquiry and work to change what they can, driven and monitored by the boots on the grounds doing the real work. A root cause analysis could also find examples of positive deviance, or classrooms and teachers with uncommon success which could serve as a model for the entire school. Abandonment of fads and programs would probably be at a 2:1 initiative; for each new program, two should be abandoned. In the antiframework, teachers would be the staff where decision making was pushed down to and the ones who could "pull the rope," or stop and point out failures and problems in real time.

Making school improvement plans revolutionary would also be rooted in the past—both proximal and distal. Schools would have to change to accomplish the antiframework. Why and how students fail would need documented (Schmoker & Wilson, 1993). Test makers could follow cohorts, student level variables, and triangulate other tests to flesh out patterns. There should not be a year-long wait to see if any of the random forays into improvement made a marked difference, with the need for rapid, small changes (Breakspear & Jones, 2020; Rubenstein-Montano et al., 2001). Caputo and Rastelli (2014) presented a way forward: School improvement involves leaders—teachers and principals—as researchers. Self-construction, contextualization of knowledge, and testing for results (Peurach et al., 2016), along with an institution of clinical collaboration and professional development, could directly disrupt the business-as-usual approach (Gonzales et al., 2020).

## Limitations

There were several limitations. First, the quantitative data examined two states and three years. Though the sample was large, schools in different states with different tests might have competing findings. Other factors, such as school climate, demographics, and school size, could improve the generalizability of the findings. Secondly, the qualitative findings were from one district. Comparing plans between districts and for a longer time period would add to the understanding of the process. A more complex statistical analysis could give a model of important factors.

Future research needs to consider different ecological variables which impact student achievement (Feldhoff et al., 2022). The school's demographic factors, curriculum, and staff characteristics could offer a holistic view of the context of school improvement plans. Goals and objectives in SIPs should be firmly tied to concrete metrics. Finally, highly successful turnaround schools and extreme failures of plans would give specifics in the hows and the whats.

## Conclusion

Educational professionals, from practitioners to professors, have a moral imperative to own their failures and recognize current business-as-usual approaches lack positive results (Fullan, 2018; Timperley & Robinson, 2001). Research on school improvement was often fragmented and contradictory (Dragoset et al., 2019; Feldhoff & Radisch, 2021; Sun et al., 2021), yet at the school level, there was consistency: the same results despite binder fillers—school improvement plans—which sat on a shelf and produced little value. Long-time 20th century school union leaders Bob Hurst and Bill Dobbles described the problem: A teacher was promoted to principal, and the administrators would take the new principal in the backroom and give them the secrets to success; instantly, they knew it all. Scholars and consultants mean well, but the late Robert Slavin (2017) pointed out the fallacies of such a disconnect. Complexity and ambiguity of systems defy fads and single-order change (von Bertalanffy, 1972).

Argyris (2000) claimed jargon, buzzwords, and dreamy visions, even when research backed, failed to translate. School improvement results at the individual level were black swans, predicated on overly simplistic solutions (Shaked & Schechter, 2020). Organizational dysfunction matters (Kiliçoğlu et al., 2019), and generic, external answers failed to address the idiosyncrasies of the local context of each school building. Schools must both go it alone and connect with others going through the same experience (Bryk, 2015); otherwise, schools would be apt to discontinue the current charade and save the time, money, and resources.

## References


Adelman, H. S., & Taylor, L. (2007). Systemic change for school improvement. *Journal of Educational and Psychological Consultation, 17*(1), 55-77. Retrieved from https://www.tandfonline.com/journals/hepc20

Alvy, H. (2017). *Fighting for change in your school: How to avoid fads and focus on substance.* ASCD.

American Enterprise Institute. (2017). The $7 billion school improvement grant program: Greatest failure in the history of the US Department of Education? Retrieved from https://www.aei.org/education/greatest-failure-in-history-us-department-of-education/

Argyris, C. (2000). *Flawed advice and the management trap.* Oxford University Press.

Arnold, R. D., & Wade, J. P. (2015). A definition of systems thinking: A systems approach. *Procedia Computer Science, 44*, 669-678. https://doi.org/10.1016/j.procs.2015.03.050

Arnold, R. D., & Wade, J. P. (2017). A complete set of systems thinking skills. *Insight, 20*(3), 9-17. https://doi.org/10.1002/inst.12159

Askell-Williams, H., & Koh, G. A. (2020). Enhancing the sustainability of school improvement initiatives. *School Effectiveness and School Improvement, 31*(4), 660-678. https://www.tandfonline.com/doi/abs/10.1080/09243453.2020.1767657

Backstrom, B. (2019). School turnaround efforts: What's been tried, why those efforts failed, and what to do now. *Nelson A. Rockefeller Institute of Government*. Retrieved from https://rockinst.org/

Becker, H., & Geer, B. (1957). Participant observation and interviewing: A comparison. *Human Organization, 16*(3), 28-32. https://doi.org/10.17730/humo.16.3.k687822132323013

Benoliel, P., & Berkovich, I. (2020). Learning from intelligent failure: An organizational resource for school improvement. *Journal of Educational Administration, 59*(4), 405-421. https://doi.org/10.1108/JEA-07-2020-0155

Betts, F. (1992). How systems thinking applies to education. *Educational Leadership.* Retrieved from https://www.ascd.org/

Bloom, C. M., & Owens, E. W. (2013). Principals' perception of influence on factors affecting student achievement in low-and high-achieving urban high schools. *Education and Urban Society, 45*(2), 208-233. https://doi.org/10.1177%2F0013124511406916

Bohanon, H. S., Wu, M. J., Kushki, A., LeVesseur, C., Harms, A., Vera, E., ... & Shriberg, D. (2021). The role of school improvement planning in the implementation of MTSS in secondary schools. *Preventing School Failure: Alternative Education for Children and Youth, 65*(3), 230-242. https://doi.org/10.1080/1045988X.2021.1908215

Breakspear, S., & Jones, B. R. (2020). *Teaching sprints: How overloaded educators can keep getting better.* Corwin Press.

Brown, C., Schildkamp, K., & Hubers, M. D. (2017). Combining the best of two worlds: A conceptual proposal for evidence-informed school improvement. *Educational Research, 59*(2), 154-172. https://doi.org/10.1080/00131881.2017.1304327

Browne-Ferrigno, T., Allen, L. W., & Hurt, P. (2008). Longitudinal reform and renewal efforts to improve public schools: Kentucky's standards and indicators for school improvement. *Leadership and Policy in Schools, 7*(4), 401-427. https://doi.org/10.1080/15700760701748477

Bryk, A. S. (2015). 2014 AERA distinguished lecture: Accelerating how we learn to improve. *Educational Researcher, 44*(9), 467-477. https://doi.org/10.3102%2F0013189X15621543

Camera, L. (2021). America's kids earn disappointing grades on nation's report card. *US News.* Retrieved from https://www.usnews.com/news/education-news/articles/2021-10-14/americas-kids-earn-disappointing-grades-on-nations-report-card

Cannata, M. A., Smith, T. M., & Taylor Haynes, K. (2017). Integrating academic press and support by increasing student ownership and responsibility. *AERA Open, 3*(3). https://doi.org/10.1177/2332858417713181

Caputo, A., & Rastelli, V. (2014). School improvement plans and student achievement: Preliminary evidence from the Quality and Merit Project in Italy. *Improving Schools, 17*(1), 72-98. https://doi.org/10.1177%2F1365480213515800

Carcary, M. (2009). The research audit trial—enhancing trustworthiness in qualitative inquiry. *Electronic Journal of Business Research Methods, 7*(1), 11-24. Retrieved from https://www.ejbrm.com

Carlson, D., & Lavertu, S. (2018). School improvement grants in Ohio: Effects on student achievement and school administration. *Educational Evaluation and Policy Analysis, 40*(3), 287-315. https://doi.org/10.3102%2F0162373718760218

Chambers Mack, J., Johnson, A., Jones-Rincon, A., Tsatenawa, V., & Howard, K. (2019). Why do teachers leave? A comprehensive occupational health study evaluating intent-to-quit in public school teachers. *Journal of Applied Biobehavioral Research, 24*(1), e12160. https://doi.org/10.1111/jabr.12160

Coe, R. (2009). School improvement: Reality and illusion. *British Journal of Educational Studies, 57*(4), 363-379. https://doi.org/10.1111/j.1467-8527.2009.00444.x

Cohen-Vogel, L., Cannata, M., Rutledge, S. A., & Socol, A. R. (2016). A model of continuous improvement in high schools: A process for research, innovation design, implementation, and scale. *Teachers College Record, 118*(13), 1-26. https://doi.org/10.1177%2F016146811611801301

Coker, D. C. (2021a). The forgotten child: Juvenile delinquents as a metaphor for change. In *Handbook of research on barriers for teaching 21st-century competencies and the impact of digitalization* (pp. 182-196). IGI Global.

Coker, D. C. (2021b). Making thematic analysis systematic: The seven deadly sins. *Journal of Studies in Education, 11*(3), 126-146. https://doi.org/10.5296/jse.v11i3.18882

Coker, D. C. (2022a). A mission statement does not a mission make: A mixed methods investigation in public education. *International Education Studies, 15*(1), 210-225. http://dx.doi.org/10.5539/ies.v15n1p210

Coker, D.C. (2022b). Multiverses, myths, and the Big Bang: Problems in educational research. In V. Wang (ed.), *Handbook of research on educational leadership and research methodology.* IGI Global.

Connolly, M., James, C., & Fertig, M. (2019). The difference between educational management and educational leadership and the importance of educational responsibility. *Educational Management Administration & Leadership, 47*(4), 504-519. https://doi.org/10.1177/1741143217745880

Crandall, D. P., Eiseman, J. W., & Louis, K. S. (1986). Strategic planning issues that bear on the success of school improvement efforts. *Educational Administration Quarterly, 22*(3), 21-53. https://doi.org/10.1177/0013161X86022003002

Creswell, J. W. (2021). *A concise introduction to mixed methods research.* Sage Publications.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124-130. https://doi.org/10.1207/s15430421tip3903_2

Datnow, A. (2020). The role of teachers in educational reform: A 20-year perspective. *Journal of Educational Change, 21*(3), 431-441. https://doi.org/10.1007/s10833-020-09372-5

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(2), 81-112. https://doi.org/10.3102/01623737024002081

Dixon, C. J., & Palmer, S. N. (2020). Transforming educational systems toward continuous improvement: A reflection guide for K-12 executive leaders. *Carnegie Foundation for the Advancement of Teaching*. Retrieved from https://www.carnegiefoundation.org/

Dougherty, S. M., & Weiner, J. M. (2019). The Rhode to turnaround: The impact of waivers to No Child Left Behind on school performance. *Educational Policy, 33*(4), 555-586. https://doi.org/10.1177%2F0895904817719520

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., & Lee Luca, D. (2019). The impact of school improvement grants on student outcomes: Findings from a national evaluation using a regression discontinuity design. *Journal of Research on Educational Effectiveness, 12*(2), 215-250. https://doi.org/10.1080/19345747.2019.1571654

Duignan, P. (1986). Research on effective schooling: Some implications for school improvement. *Journal of Educational Administration, 24*(1), 59-73. https://doi.org/10.1108/eb009909

Elgart, M. A. (2017). Can schools meet the promise of continuous improvement? *Phi Delta Kappan, 99*(4), 54-59. https://doi.org/10.1177%2F0031721717745546

Evans, L. (2022). Is educational leadership (still) worth studying? An epistemic worthiness-informed analysis. *Educational Management Administration & Leadership, 50*(2), 325-348. https://doi.org/10.1177%2F17411432211066273

Evans, L., Thornton, B., & Usinger, J. (2012). Theoretical frameworks to guide school improvement. *NASSP Bulletin, 96*(2), 154-171. https://doi.org/10.1177%2F0192636512444714

Faraway, J. J. (2016). *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models.* CRC Press.

Feldhoff, T., Emmerich, M., Radisch, F., Wurster, S., & Bischof, L. M. (2022). Unique problems require unique solutions—models and problems of linking school effectiveness and school improvement. *Education Sciences, 12*(3), 158. https://doi.org/10.3390/educsci12030158

Feldhoff, T., & Radisch, F. (2021). Why must everything be so complicated? Demands and challenges on methods for analyzing school improvement processes. In A. Oude Groote Beverborg, T. Feldhoff, K. Maag Merki, F. Radisch (eds.), *Concept and design developments in school improvement research* (pp. 9-26). Springer Nature. https://doi.org/10.1007/978-3-030-69345-9

Feldhoff, T., Radisch, F., & Bischof, L. M. (2016). Designs and methods in school improvement research: A systematic review. *Journal of Educational Administration, 54*(2), 209-240. https://doi.org/10.1108/JEA-07-2014-0083

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80-92. https://doi.org/10.1177%2F160940690600500107

Fernandez, K. E. (2011). Evaluating school improvement plans and their effect on academic performance. *Educational Policy, 25*(2), 338-367. https://doi.org/10.1177/0895904809351693

Fullan, M. (2018). *Nuance: Why some leaders succeed and others fail.* Corwin Press.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945. https://doi.org/10.3102/00028312038004915

Gleason, P., Crissey, S., Chojnacki, G., Zukiewicz, M., Silva, T., Costelloe, S., & O'Reilly, F. (2019). Evaluation of support for using student data to inform teachers' instruction. NCEE 2019-4008. *National Center for Education Evaluation and Regional Assistance*. Retrieved from http://ies.ed.gov/ncee .

Gonzales, M. M., Bickmore, D. L., & Roberts, M. B. (2020). Implementing school improvement plans: Perceptions and implications of aspiring principals for educational leadership programs. *Journal of Research on Leadership Education, 17*(2), 160-180. https://doi.org/10.1177%2F1942775120964821

Good, T., Burross, H., & McCaslin, M. (2005). Comprehensive school reform: A longitudinal study of school improvement in one state. *Teachers College Record, 107*(10), 2205-2226. Retrieved from https://www.tcrecord.org/

Goodwin, B., Cameron, G., & Hein, H. (2015). *Balanced leadership for powerful learning.* ASCD.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495-500. https://doi.org/10.1177/003172170909000709

Hallinger, P., Gümüş, S., & Bellibaş, M. Ş. (2020). 'Are principals instructional leaders yet?' A science map of the knowledge base on instructional leadership, 1940–2018. *Scientometrics, 122*(3), 1629-1650. https://doi.org/10.1007/s11192-020-03360-5

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., Wayman, J. C., ... & Steele, J. L. (2009). Using student achievement data to support instructional decision making. Institute of Education Sciences. Retrieved from https://ies.ed.gov/

Hargreaves, D. H. (1995). School culture, school effectiveness and school improvement. *School Effectiveness and School Improvement, 6*(1), 23-46. https://doi.org/10.1080/0924345950060102

Harris, A. (2000). What works in school improvement? Lessons from the field and future directions. *Educational Research, 42*(1), 1-11. https://doi.org/10.1080/001318800363872

Harris, A. (2001). Building the capacity for school improvement. *School Leadership & Management, 21*(3), 261-270. https://doi.org/10.1080/13632430120074419

Harris, A., Jones, M., & Hashim, N. (2021). System leaders and system leadership: Exploring the contemporary evidence base. *School Leadership & Management, 41*(4-5), 387-408. https://doi.org/10.1080/13632434.2021.1889492

Heifetz, R. A., Grashow, A., & Linsky, M. (2009). *The practice of adaptive leadership: Tools and tactics for changing your organization and the world.* Harvard Business Press.

Heissel, J. A., & Ladd, H. F. (2018). School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review, 62*, 302-320. https://doi.org/10.1016/j.econedurev.2017.08.001

Henry, G. T., Pham, L. D., Kho, A., & Zimmer, R. (2020). Peeking into the black box of school turnaround: A formal test of mediators and suppressors. *Educational Evaluation and Policy Analysis, 42*(2), 232-256. https://doi.org/10.3102%2F0162373720908600

Heylighen, F., & Joslyn, C. (1992). What is systems theory. *Principia Cybernetica Web*, 1. Retrieved from http://pespmc1.vub.ac.be/SYSTHEOR.html

Hitt, D. H., & Meyers, C. V. (2018). Beyond turnaround: A synthesis of relevant frameworks for leaders of sustained improvement in previously low-performing schools. *School Leadership & Management, 38*(1), 4-31. https://doi.org/10.1080/13632434.2017.1374943

Hitt, D. H., & Tucker, P. D. (2016). Systematic review of key leader practices found to influence student achievement: A unified framework. *Review of Educational Research, 86*(2), 531-569. https://doi.org/10.3102/0034654315614911

Hollingworth, L., Olsen, D., Asikin-Garmager, A., & Winn, K. M. (2018). Initiating conversations and opening doors: How principals establish a positive building culture to sustain school improvement efforts. *Educational Management Administration & Leadership, 46*(6), 1014-1034. https://doi.org/10.1177/1741143217720461

Hopkins, D., & Higham, R. (2007). System leadership: mapping the landscape. *School Leadership and Management, 27*(2), 147-166. https://doi.org/10.1080/13632430701237289

Huber, D., & Conway, J. M. (2015). The effect of school improvement planning on student achievement. *Planning and Changing, 46* (1/2), 57–70. Retrieved from https://education.illinoisstate.edu/planning/

Hung, W. (2008). Enhancing systems-thinking skills with modelling. *British Journal of Educational Technology, 39*(6), 1099-1120. https://doi.org/10.1111/j.1467-8535.2007.00791.x

Ingersoll, R. M., Sirinides, P., & Dougherty, P. (2018). Leadership matters: Teachers' roles in school decision making and school performance. *American Educator, 42*(1), 13-17. Retrieved from https://www.aft.org/ae

Kast, F. E., & Rosenzweig, J. E. (1972). General systems theory: Applications for organization and management. *Academy of Management Journal, 15*(4), 447-465. https://doi.org/10.5465/255141

Kiliçoğlu, G., Kiliçoğlu, D. Y., & Karadağ, E. (2019). Do schools fail to “walk their talk”? Development and validation of a scale measuring organizational hypocrisy. *Leadership and Policy in Schools, 18*(1), 52-82. https://doi.org/10.1080/15700763.2017.1371762

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547-588. https://doi.org/10.3102%2F0034654318759268

Leithwood, K., Sun, J., & McCullough, C. (2019). How school districts influence student achievement. *Journal of Educational Administration, 57*(5), 519-539. https://doi.org/10.1108/JEA-09-2018-0175

Leithwood, K., Sun, J., & Schumacker, R. (2020). How school leadership influences student learning: A test of "The four paths model". *Educational Administration Quarterly, 56*(4), 570-599. https://doi.org/10.1177%2F0013161X19878772

MacNeil, A. J., Prater, D. L., & Busch, S. (2009). The effects of school culture and climate on student achievement. *International Journal of Leadership in Education, 12*(1), 73-84. https://doi.org/10.1080/13603120701576241

Madigan, D. J., & Kim, L. E. (2021). Towards an understanding of teacher attrition: A meta-analysis of burnout, job satisfaction, and teachers' intentions to quit. *Teaching and Teacher Education*, 105, 103425. https://doi.org/10.1016/j.tate.2021.103425

McIntosh, K., Girvan, E. J., McDaniel, S. C., Santiago-Rosario, M. R., St. Joseph, S., Fairbanks Falcon, S., ... & Bastable, E. (2021). Effects of an equity-focused PBIS approach to school improvement on exclusionary discipline and school climate. *Preventing School Failure: Alternative Education for Children and Youth*, 1-8. https://doi.org/10.1080/1045988X.2021.1937027

Meyer, F., Bendikson, L., & Le Fevre, D. M. (2020). Leading school improvement through goal-setting: Evidence from New Zealand schools. *Educational Management Administration & Leadership*, 1741143220979711. https://doi.org/10.1177/1741143220979711

Meyers, C. V., & Hitt, D. H. (2018). Planning for school turnaround in the United States: An analysis of the quality of principal-developed quick wins. *School Effectiveness and School Improvement, 29*(3), 362-382. https://doi.org/10.1080/09243453.2018.1428202

Meyers, C. V., & VanGronigen, B. A. (2019). A lack of authentic school improvement plan development: Evidence of principal satisficing behavior. *Journal of Educational Administration, 57*(3), 261-278. https://doi.org/10.1108/JEA-09-2018-0154

Meyers, C. V., & VanGronigen, B. A. (2020). Planning for what? An analysis of root cause quality and content in school improvement plans. *Journal of Educational Administration, 59*(4), 437-453. https://doi.org/10.1108/JEA-07-2020-0156

Mintrop, R., & Zumpe, E. (2019). Solving real-life problems of practice and education leaders' school improvement mind-set. *American Journal of Education, 125*(3), 295-344. https://doi.org/10.1086/702733

Moody, M. S. (2019). If instructional coaching really works, why isn't it working. *Educational Leadership, 77*(3), 30-35. Retrieved from http://www.ascd.org

Morel, B., & Ramanujam, R. (1999). Through the looking glass of complexity: The dynamics of organizations as adaptive and evolving systems. *Organization Science, 10*(3), 278-293. https://doi.org/10.1287/orsc.10.3.278

Murphy, J. (2013). The architecture of school improvement. *Journal of Educational Administration, 51*(3), 252-263. https://doi.org/10.1108/09578231311311465

Neuman, S. B. (2016). Code red: The danger of data-driven instruction. *Educational Leadership, 74*(3), 24-29. Retrieved from https://www.ascd.org/

Padilla, G., Guerra, F., & Zamora, R. (2020). Effective school practices in Title I schools exceeding educational expectations (E 3). *International Journal of Educational Reform, 29*(2), 103-122. https://doi.org/10.1177%2F1056787919886582

Peck, C., & Reitzug, U. C. (2014). School turnaround fever: The paradoxes of a historical practice promoted as a new reform. *Urban Education, 49*(1), 8-38. https://doi.org/10.1177%2F0042085912472511

Peurach, D. J., Glazer, J. L., & Winchell Lenhoff, S. (2016). The developmental evaluation of school improvement networks. *Educational Policy, 30*(4), 606-648. https://doi.org/10.1177/0895904814557592

Polanin, J. R., Espelage, D. L., Grotpeter, J. K., Spinney, E., Ingram, K. M., Valido, A., El Sheikh, A., Torgal, C., & Robinson, L. (2021). A meta-analysis of longitudinal partial correlations between school violence and mental health, school performance, and criminal or delinquent acts. *Psychological Bulletin, 147*(2), 115–133. https://doi.org/10.1037/bul0000314

Preston, C., Goldring, E., Guthrie, J. E., Ramsey, R., & Huff, J. (2017). Conceptualizing essential components of effective high schools. *Leadership and Policy in Schools, 16*(4), 525-562. https://doi.org/10.1080/15700763.2016.1205198

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 427-452. https://doi.org/10.1086/461325

Quong, T., & Walker, A. (2010). Seven principles of strategic leadership. *International Studies in Educational Administration* (Commonwealth Council for Educational Administration & Management (CCEAM)), *38*(1), 22-34. Retrieved from http://cceam.net/publications/isea/

Ranganathan, P., Pramesh, C. S., & Buyse, M. (2015). Common pitfalls in statistical analysis: Clinical versus statistical significance. *Perspectives in Clinical Research, 6*(3), 169-170. https://dx.doi.org/10.4103%2F2229-3485.159943

Redding, C., & Nguyen, T. D. (2020). The relationship between school turnaround and student outcomes: A meta-analysis. *Educational Evaluation and Policy Analysis, 42*(4), 493-519. https://doi.org/10.3102%2F0162373720949513

Redding, C., & Searby, L. (2020). The map is not the territory: Considering the role of school improvement plans in turnaround schools. *Journal of Cases in Educational Leadership, 23*(3), 63-75. https://doi.org/10.1177/1555458920938854

Rodgers, B. L., & Cowles, K. V. (1993). The qualitative research audit trail: A complex collection of documentation. *Research in Nursing & Health, 16*(3), 219-226. https://doi.org/10.1002/nur.4770160309

Rohanna, K. (2017). Breaking the "adopt, attack, abandon" cycle: A case for improvement science in K–12 education. *New Directions for Evaluation, 2017*(153), 65-77. https://doi.org/10.1002/ev.20233

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator, 36*(1), 12. Retrieved from https://www.aft.org/ae

Ross, J. A., & Gray, P. (2006). School leadership and student achievement: The mediating effects of teacher beliefs. *Canadian Journal of Education/Revue canadienne de l'éducation*, 798-822. https://doi.org/10.2307/20054196

Rowan, B. (2002). The ecology of school improvement: Notes on the school improvement industry in the United States. *Journal of Educational Change, 3*(3), 283-314. https://doi.org/10.1023/A:1021277712833

Rubenstein-Montano, B., Liebowitz, J., Buchwalter, J., McCaw, D., Newman, B., Rebeck, K., & Team, T. K. M. M. (2001). A systems thinking framework for knowledge management. *Decision Support Systems, 31*(1), 5-16. https://doi.org/10.1016/S0167-9236(00)00116-0

Scherer, R., & Nilsen, T. (2019). Closing the gaps? Differential effectiveness and accountability as a road to school improvement. *School Effectiveness and School Improvement, 30*(3), 255-260. https://doi.org/10.1080/09243453.2019.1623450

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research, 61*(3), 257-273. https://doi.org/10.1080/00131881.2019.1625716

Schmoker, M., & Wilson, R. B. (1993). Adapting total quality doesn't mean. *Educational Leadership,* 62-63. Retrieved from https://www.ascd.org/

Schneider, M., & Somers, M. (2006). Organizations as complex adaptive systems: Implications of complexity theory for leadership research. *The Leadership Quarterly, 17*(4), 351-365. https://doi.org/10.1016/j.leaqua.2006.04.006

Schneider, A., Wickert, C., & Marti, E. (2017). Reducing complexity by creating complexity: A systems theory perspective on how organizations respond to their environments. *Journal of Management Studies, 54*(2), 182-208. https://doi.org/10.1111/joms.12206

Sebring, P. B., Allensworth, E., Bryk, A. S., Easton, J. Q., & Luppescu, S. (2006). The essential supports for school improvement. Research report. *Consortium on Chicago School Research*. Retrieved from https://consortium.uchicago.edu/publications/essential-supports-school-improvement

Shaked, H., & Schechter, C. (2020). Systems thinking leadership: New explorations for school improvement. *Management in Education, 34*(3), 107-114. https://doi.org/10.1177/0892020620907327

Shindler, J., Jones, A., Williams, A. D., Taylor, C., & Cardenas, H. (2016). The school climate-student achievement connection: If we want achievement gains, we need to begin by improving the climate. *Journal of School Administration Research and Development, 1*(1), 9-16. Retrieved from https://ojed.org/

Slavin, R. E. (2017). Evidence-based practice meets improvement science. *Éducation et didactique, 11*(11-2), 45-48. https://doi.org/10.4000/educationdidactique.2722

Solone, C. J., Thornton, B. E., Chiappe, J. C., Perez, C., Rearick, M. K., & Falvey, M. A. (2020). Creating collaborative schools in the United States: A review of best practices. *International Electronic Journal of Elementary Education, 12*(3), 283-292. Retrieved from https://iejee.com/index.php/IEJEE/article/view/1091

Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research, 88*(4), 479-507. https://doi.org/10.3102/0034654317751919

Stoll, L. (2009). Capacity building for school improvement or creating capacity for learning? A changing landscape. *Journal of Educational Change, 10*(2), 115-127. https://doi.org/10.1007/s10833-009-9104-3

Strunk, K. O., Marsh, J. A., Bush-Mecenas, S. C., & Duque, M. R. (2016). The best laid plans: An examination of school plan quality and implementation in a school improvement initiative. *Educational Administration Quarterly, 52*(2), 259-309. https://doi.org/10.1177/0013161X15616864

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education, 4*(3), 279-282. https://doi.org/10.4300/JGME-D-12-00156.1

Sun, M., Kennedy, A. I., & Loeb, S. (2021). The longitudinal effects of school improvement grants. *Educational Evaluation and Policy Analysis*, 01623737211012440. https://doi.org/10.3102%2F01623737211012440

Thompson, C. S. (2018). 'School administrators' and stakeholders' attitudes toward, and perspectives on, school improvement planning. *Educational Planning, 25*(4), 7-26. Retrieved from https://isep.info/educational-planning-journal/

Timperley, H. S. (2005). Instructional leadership challenges: The case of using student achievement information for instructional improvement. *Leadership and Policy in Schools, 4*(1), 3-22. https://doi.org/10.1080/15700760590924591

Timperley, H. S., & Robinson, V. M. (2001). Achieving school improvement through challenging and changing teachers' schema. *Journal of Educational Change, 2*(4), 281-300. https://doi.org/10.1023/A:1014646624263

Tingle, E., Corrales, A., & Peters, M. L. (2019). Leadership development programs: Investing in school principals. *Educational Studies, 45*(1), 1-16. https://doi.org/10.1080/03055698.2017.1382332

Tran, N. H., Hallinger, P., & Truong, T. (2018). The heart of school improvement: A multi-site case study of leadership for teacher learning in Vietnam. *School Leadership & Management, 38*(1), 80-101. https://doi.org/10.1080/13632434.2017.1371690

VanGronigen, B. A., & Meyers, C. V. (2021). Exploring the association between short-cycle school improvement planning and student achievement in underperforming schools. *Journal of School Leadership*, 10526846211018207. https://doi.org/10.1177/10526846211018207

von Bertalanffy, L. (1972). The history and status of general systems theory. *Academy of Management Journal, 15*(4), 407-426. https://doi.org/10.5465/255139

Wallace, S. O., Acker-Hocevar, M., & Sweatt, O. (2001). Developing an open systems view for assessing educational leadership. *Journal of School Leadership, 11*(3), 241-257. https://doi.org/10.1177/105268460101100306

Wong, B. Y. (1997). Clearing hurdles in teacher adoption and sustained use of research-based instruction. *Journal of Learning Disabilities, 30*(5), 482-485. https://doi.org/10.1177/002221949703000503

## Appendix A

Sub-elements by category and correlated by elements-dimensions-themes.

*Note.* Percentages and counts are in parentheses. Dominant dimensions and themes listed.

| **C1 Assessment (7.5%)** | **C5 Interventions (4.6%)** |
|---|---|
| SE1.1 Benchmark (5), E1-D3-T2 | SE5.1 Extended Learning (8), E13-D4-T3 |
| SE1.2 Formative (13), E1-D3-T2 | SE5.99 Miscellaneous (11) |
| SE1.3 General (8) E1-D3-T2 | **C6 Professional Development (8.2%)** |
| SE1.4 Progress Monitoring (4), E9-D3-T2 | SE6.1 Coaches (12), E16-D5-T2 |
| SE1.99 Universal (1), E1-D3-T2 | SE6.2 General (3), E14-D5-T2 |
| **C2 Curriculum (4.1%)** | SE6.3 Leadership (2), E99-D5-T2 |
| SE2.1 Coaching (2), E16-D1-T1 | SE6.4 Standards (4), E3-D5-T2 |
| SE2.2 Essentials (3), E7-D1-T1 | SE6.5 Strategies (7), E10-D5-T2 |
| SE2.3 Standards Alignment (8), E3-D3-T1 | SE6.99 Miscellaneous (5) |
| SE2.99 Miscellaneous (4) | **C7 Professional Learning Communities (10.4%)** |
| **C3 Data Analysis (7.7%)** | SE7.1 General (15), E14-D6-T2 |
| SE3.1 Schoolwide (7), E5-D7-T2 | SE7.2 Data Analysis (14), E5-D6-T2 |
| SE3.2 Teacher (15), E5-D7-T2 | SE7.3 Standards Alignment (5), E3-D6-T2 |
| SE3.3 Teams/Collaboration (10), E5-D7-T2 | SE7.31 Consistency (2), E99-D6-T2 |
| **C4 Instruction (41.2%)** | SE7.32 Rigor (4), E3-D6-T2 |
| SE4.1 General (12), E6-D2-T3 | SE7.4 Norms (2), E14-D6-T2 |
| SE4.11 Guided Reading (13), E10-D2-T3 | SE7.99 Miscellaneous (1) |
| SE4.12 Independent Reading (8), E10-D2-T3 | **C8 Supervision (3.4%)** |
| SE4.13 Journaling (4), E10-D2-T3 | SE8.1 Observations (13), E15-D8-T3 |
| SE4.14 Notetaking (4), E10-D2-T3 | SE8.99 Miscellaneous (1) |
| SE4.15 Reading General (4), E10-D2-T3 | **C99 Miscellaneous (1.5%)** |
| SE4.16 Rigorous (2), E10-D2-T3 | SE99 Miscellaneous (7) |
| SE4.17 Small Groups (12), E2-D2-T3 | |
| SE4.18 Student Centered (5), E10-D2-T3 | |
| SE4.19 Vocabulary (10), E18-D2-T3 | |
| SE4.20 Essential Questions (9), E7-D2-T3 | |
| SE4.21 Writing (10), E19-D2-T3 | |
| SE4.3 Feedback (12), E4-D3-T3 | |
| SE4.4 Mini-Lessons (4), E12-D2-T3 | |
| SE4.5 On-Task (3), E12-D2-T3 | |
| SE4.6 Rigor (4), E3-D2-T3 | |
| SE4.7 Rubrics/Scales (10), E8-D3-T3 | |
| SE4.8 Standards Alignment (8), E3-D2-T3 | |
| SE4.9 Gradual Release (4), E10-D2-T3 | |
| SE4.91 Computers (2), E10-D2-T3 | |
| SE4.92 Focus (2), E10-D2-T3 | |
| SE4.93 Metacognition (2), E10-D2-T3 | |
| SE4.94 Physical Movement (2), E10-D2-T3 | |
| SE4.99 Miscellaneous (23) | |

## Appendix B

Elements

*Note.* Counts and percentages are in parentheses,

| | |
|---|---|
| E1 | Students not assessed and held accountable (27; 7.3%) |
| E2 | Cooperative learning needed (12; 3.2%) |
| E3 | Lack of standards alignment (36; 9.8%) |
| E4 | Inadequate check for understanding/feedback poor or inappropriate (13; 3.6%) |
| E5 | Lack of data-driven decisions (49; 13.4%) |
| E6 | Lack of differentiation/RtI effectiveness (17; 4.6%) |
| E7 | Lack of focused instruction/essential questions (12; 3.3%) |
| E8 | Lack of meaningful grading (10; 2.7%) |
| E9 | Lack of progress monitoring (4; 1.1%) |
| E10 | Lack of research-based strategies (83; 22.7%) |
| E11 | Lesson not paced/engaging (5; 1.4%) |
| E12 | OT-Students are off task (3; 0.8%) |
| E13 | Period of time too small (10; 2.7%) |
| E14 | PLCs fail to function (19; 5.2%) |
| E15 | Staff do not follow through/inability to produce without supervision (13; 3.6%) |
| E16 | Staff need coaching to improve (16; 4.4%) |
| E17 | Teachers lack adequate knowledge (3; 0.8%) |
| E18 | Inadequate vocabulary instruction (10; 2.7%) |
| E19 | Students do not write enough (10; 2.7%) |
| E99 | Miscellaneous (14; 3.8%) |

## Appendix C

Dimensions

*Note.* Counts and percentages in parentheses.

| | |
|---|---|
| D1 | Curriculum not aligned not proper scope and sequence. (17; 4.6%) |
| D2 | Instruction 1: Staff lack effective, research-based strategies aligned to state standards to engage students. (145; 39.6%) |
| D3 | Instruction 2: Staff members lack effective feedback, assessments procedures, and progress monitoring (53; 14.5%) |
| D4 | Interventions: There is not enough staff or time, but staff members lack the skills to correctly intervene anyway. (19; 5.2%) |
| D5 | PD: With coaches, to build missing necessary skills because staff struggle collaborating. (34; 9.3%) |
| D6 | PLCs: Schools either lack PLCs or ones which function to drive instruction by data and an essential, standards-aligned curriculum. (43; 11.7%) |
| D7 | Staff members do not collect data regularly, and when they do, they lack using the data on in classroom instruction, interventions, and whole school planning. (32; 8.7%) |
| D8 | Supervision: A lack of leadership and poor teacher skills need remedied to improve instruction. (14; 3.8%) |
| D99 | Miscellaneous (9; 2.5%) |

## Appendix D

Themes broken down by dimensions, elements, categories, who implements, and frequency

*Note.* Counts and percentages are in parentheses.

| | |
|---|---|
| T1 | School Level: School cannot deliver the necessary curriculum: lack a coherent, standards-based curriculum, an optimal schedule, and sufficient staff. (20; 5.5%)<br>Major Dimensions: D1 (17), D99 (3) Major Elements: E3, E7, E6. Major Category: C6<br>Who implements? Teachers (15), Administrators (5)<br>Frequency: Daily: 3, Weekly: 4, Bi-Weekly: 4, Monthly: 0, Quarterly: 3, Yearly: 6 |
| T2 | Behind the Scenes: Teachers need professional development and regular collaboration time to develop effective instructional techniques, intervention/differentiation, and processes for data-driven instruction. (140; 38.3%)<br>Major Dimensions: D3 (31), D5 (34), D6 (43), D7 (32). Major elements: E5, E1, E3, E4, E16. Categories: C1, C3, C6, C7.<br>Who implements? Teachers (123), Administrators (17)<br>Frequency: Daily: 36; Weekly: 53; Bi-Weekly: 5; Monthly: 17; Quarterly: 17; Yearly: 12. |
| T3 | Classroom Level: Teachers lack effective teaching strategies and interventions with proper supervision. (202; 55.2%)<br>Major Dimensions: D2 (145), D3 (22), D4 (19), D8 (14), D99 (2). Major Elements: E10, E2-4, E6, E9, E15, E18, E19. Categories: C4, C5, C8.<br>Who implements? Teachers (186), Administrators (16)<br>Frequency: Daily: 141; Weekly: 56; Bi-Weekly: 0; Monthly: 0; Quarterly: 1; Yearly: 4. |

**Appendix E**

Trends in student achievement from 2015-2018 by school