

**Leading Indicators of Academic Achievement: Investigating the Predictive Validity of an Observation
Instrument in A Large District**

Lindsey Devers Basileo

Merewyn Elizabeth Lyons

Michael David Toth

Applied Research Center, Instructional Empowerment, Blairsville, Pennsylvania

Author Details

Correspondence concerning this article should be addressed to Lindsey Devers Basileo, Applied Research Center, Instructional Empowerment, 175 Cornell Road, Suite 18, Blairsville, PA, 15717, United States. Email: ibasileo@instructionalempowerment.com Telephone: +1 561-421-1807

Abstract

Instructional leaders need timely, predictive data to guide systemic change post pandemic. This study investigated several educational indicators of student achievement along with the Rigor Appraisal observation instrument metrics. We used correlations analysis to assess the relationship between the Rigor Appraisal and student achievement and other indicators of school effectiveness. We then used Ordinary Least Squares to investigate the relationship between the Rigor Appraisal score, the Rigor Appraisal pillar scores, program effectiveness measures, and school demographic estimates to predict student achievement at the school level. The study sample consisted of 48 to 53 schools in Illinois during the 2021-22 school year. We found that the Rigor Appraisal score had a small to moderate and statistically significant association with student achievement, a positive moderate and statistically significant association with average weekly attendance, a negative relationship with changes in student referrals and suspensions, and a moderate positive and statistically significant association with the 5Essentials school survey. Further, one pillar of the Rigor Appraisal, Activating Student Teams to Achieve the Standard, was associated with greater achievement in schools with more low-income students. We also found that the number of classroom instructional walks by the leadership team had a positive and statistically significant impact on student achievement. We note limitations to this study including that the data was from only one school district and the findings may not be generalizable. Nevertheless, Rigor Appraisal shows promise as a predictive tool for instructional systems improvement and merits further study different contexts.

Keywords: academic achievement, instructional leadership, instructional systems, instructional improvement, observation

Leading Indicators of Academic Achievement: Investigating the Predictive Validity of an Observation Instrument in A Large District

Introduction

Elementary and secondary student performance in reading and mathematics has dropped substantially in the United States since the COVID-19 pandemic. The National Assessment of Educational Progress (NAEP) Long Term Trend assessments of reading and math for nine-year old students in 2022 revealed the largest average score decline in reading since 1990, and the first ever decline in math scores (National Assessment of Educational Progress, 2022a). Results of the main NAEP assessments for 2022 also revealed overall lower scores in reading and math for fourth- and eighth-grade students (National Assessment of Educational Progress, 2022b, 2022c, 2022d, 2022e) compared to 2019 scores. Disaggregated for race, ethnicity, and socioeconomic status, most student groups showed declines. Although preliminary studies show some rebounding of academic performance, it is still significantly lower than pre-pandemic levels (Lewis & Kuhfeld, 2022).

The median learning loss is estimated at a quarter of a year for reading and a half of year for math, with high poverty districts losing as much as a third of a year in reading and two-thirds of a year in math (Fahle et al., 2022). The depth of these declines indicate that it may take many years for students to attain full learning recovery to pre-pandemic levels (Lewis & Kuhfeld, 2022). Estimates of the long-term impact of these declines in performance include lower rates of high school completion and college enrollment leading to decreases in lifetime earnings as well as losses to national GDP (Kane et al., 2022; Hanushek, 2022). The slow pace of recovery could result in students of all races, ethnicities, and income levels graduating high school without essential skills for college and career success (Azevedo et al., 2020; Betthäuser et al., 2023; Center for Reinventing Public Education, 2022b).

These troubling findings are magnified for students of color and students from low-income households. The 2022 NAEP scores reveal a continuation of historically persistent gaps between fourth-grade and eighth-grade White students and American Indian/Alaska Native, Black, and Hispanic students in both reading and math (National Assessment of Educational Progress, 2022b, 2022c, 2022d, 2022e), and widening gaps among those groups who were already performing at the lowest deciles (Center on Reinventing Public Education, 2022a, 2022b; Kuhfeld et al., 2022; Lewis et al., 2022; Lewis & Kuhfeld, 2022). These persistent gaps also affect low-income students most negatively, widening the gap in performance between these students and more affluent peers (Hammerstein et al., 2021).

Leadership for Systemic Change

The pandemic crisis has revealed that public educational systems in the United States are not responsive to the needs of diverse learners (Center for Reinventing Public Education, 2022b). For improved outcomes for all students, these systems must change. Effective instructional leaders are critical to enacting the systemic changes needed to accelerate learning and meet the needs of all students.

Although the effect of principals on students is nearly as large as the effect of teachers (Grissom et al., 2021), there is little specificity about the most essential features of instructional leadership (Grissom et al., 2013) and sparse guidance on how to overcome persistent disparities in student outcomes based on race, ethnicity, home language, ability, and socioeconomic status (Ishimaru & Galloway, 2014). Classroom walkthroughs are frequently cited as essential leadership activities, but principals may not be collecting the right data to help them differentiate between high-and low-quality teaching practices and provide actionable feedback to teachers (Boston et al., 2017; Grissom et al., 2013).

Education leaders need better data and tools to guide the complex systemic changes needed to accelerate learning for all students and eliminate academic performance gaps. We created the Rigor Appraisal Instrument to help inform their decisions to improve instructional systems, track instructional improvement, and provide a supportive, nonevaluative method to provide meaningful feedback to teachers. In the next sections, we outline the theoretical and empirical framework of the Rigor Appraisal instrument, the methods and data used in the research study, the results of the study, and review the conclusions and limitations.

Theoretical Foundations

There are five pillars to the Rigor Appraisal Instrument: 1) Creating Conditions for Learning Rigorous Standards, 2) Using Standards-based Student Evidence, 3) Activating Student Teams to Achieve the Standard, 4) Verifying Learning to Take Action within a Lesson, and 5) Using Data to Track Student Progress toward Standards. In this section, we discuss each pillar of the Rigor Appraisal including the theoretical and empirical underpinnings.

Creating Conditions for Learning Rigorous Standards

The Creating Conditions for Learning Rigorous Standards pillar measures observable systems including the school and classroom climate, supports for student self-regulation, and high functioning teacher teams. Optimal interaction of these systems at the school level fosters a rigorous learning environment.

School and Classroom Climate

High-quality and consistent interactions between school staff and students create a positive school climate (Akman, 2021; Öngel & Tabancalı, 2022; Suldo et al., 2020). School and classroom climate affects student wellbeing, life satisfaction, ethnic and moral identity, self-efficacy, and resilience (Aldridge et al., 2018; Graham, 2022; VanLone et al., 2019). Climate is also positively associated with

student achievement (Benbenishty et al., 2016; Berkowitz et al., 2016; Davis & Warner, 2018; Konold et al., 2018; Osher et al., 2020). A positive school climate can narrow the achievement gap among student subgroups (Suldo et al., 2006).

Student Self-Regulation

Teaching strategies that build student skills of self-regulation and co-regulation support student engagement and sustain cognition, emotion, and behaviors needed to accomplish learning goals (Darling-Hammond & Cook-Harvey, 2018; Hinnant-Crawford et al., 2018; Zimmerman & Kitsantas, 2014). Many positive outcomes are associated with self-regulation, including improved student achievement (Blitz et al., 2016; Marantika, 2021; Zimmerman & Kitsantas, 2014), the development of visuomotor, mathematics, to emergent literacy, vocabulary skills in young children (Becker et al., 2014; DeFlorio et al., 2019; Williams et al., 2016), reading comprehension, vocabulary, and math in elementary school students (Day & Connor, 2017; Day et al., 2015; Lenes, et al., 2020; Skibbe et al., 2019), and improved mathematics performance of secondary and postsecondary students (Cleary et al., 2017; Musso et al., 2019). There is also evidence that students in high achieving schools use more self-regulated learning strategies than those in low-performing schools (Wenjuan et al., 2019).

Collaboration in Teacher Teams

To establish effective teacher teams, school leaders support and encourage professional growth, autonomy, and teacher ownership in shared decision-making processes, setting high expectations for academic performance (Argon & Ekinici, 2016; Conner, 2015; Hirsh & Segolsson, 2019; Katzenbach & Smith, 1993; Pierce et al., 2003; Schaap and de Bruijn, 2018; Seglem, 2017). Effective teacher team collaboration can positively impact school culture and increase student achievement (Green & Allen, 2015; Kraft & Falken, 2020; LeClerc & Moreau, 2011; Muñoz & Branham, 2016; Ratts et al., 2015; Ronfeldt et al., 2015; Tichnor-Wagner et al., 2016). The higher the quality of collaboration in teacher

teams, the more effective teachers are at raising student achievement (Park et al., 2018; Ronfeldt et al., 2015; Sun et al., 2017).

Using Standards-Based Student Evidence

Standards-based student evidence indicates the presence of structures that support consistent and systematic teaching and learning tasks with student evidence at the taxonomy level of the standard, student progress toward mastery of the standards, and the quality of tasks for collaboration among students.

Target, Task, and Taxonomy Alignment

All students benefit from rigorous and rich learning experiences. When appropriately designed, even students who were previously performing at low levels can simultaneously attain foundational skills as they engage in “authentic intellectual work” (Newmann, Bryk, & Nagaoka, 2001) that requires higher levels of cognition (Scardamalia & Bereiter, 2022; Sun, Zhang, & Scardamalia, 2010). Teachers must be able to accurately judge task complexity and design learning to enact appropriate metacognitive strategies. But the learning tasks they create often do not match the taxonomic levels of the standards and are at lower levels of cognitive complexity (Anees, 2017; Pieschl, 2009). The alignment of learning tasks to the rigor of the standards requires strategic planning of curriculum, instruction, and assessment to ensure that students demonstrate learning at the level of performance, or taxonomy, of the standards (Maye, 2013; Ziebell & Clarke, 2018).

Tasks Designed for Teams

Alignment among the target, task, and taxonomy is essential, but the task must also be suitable for teamwork in groups, or team worthy (Lotan, 2003; Pennant, 2013) and produce evidence of learning. Appropriately structured teams and tasks are observable features of student learning communities in

autonomy-supportive classrooms (Reeve & Cheon, 2021; Reeve & Su, 2014; Wehmeyer et al., 2002, 2017).

Student Work Produces Evidence of Mastery

Teachers decide on the evidence they expect students to produce that reflects mastery of the standards. Both mastery and appropriate alignment are interdependent as the level of the learning task predicts the level of student performance (Elmore, 2008). The concept of student mastery of the standards is rooted in Benjamin Bloom's (1971) mastery learning in which there must be instructional alignment between the specific learning standards that students are expected to meet and the lesson, feedback, correctives, and enrichment activities. To achieve mastery learning, instruction must provide feedback, correction for students who have not mastered the learning, and enrichment for those that have reached mastery (Guskey, 1997, 2007).

Activating Student Teams to Achieve the Standard

This pillar measures the degree to which students are in academic learning teams, that teams develop resilience through productive struggle, and that they function autonomously. The concept of academic teams demands more of students than typical small group or cooperative learning strategies. As teams, students must verify the effectiveness of their own learning processes and the quality of their collective work.

Academic Teams

Academic teams are student-led, small, diverse teams with clear protocols for engaging in standards-based academic work (Toth & Sousa, 2019). Each member of an academic team has a specific responsibility such as being a facilitator or tracking group progress towards the attainment of standards. Students develop elevated levels of autonomy and critical thinking skills through collaboration in

academic teams (Francisco, 2013). As with team-based learning (Michaelson & Sweet, 2008), academic teaming is student-centered.

Organizing students in small collaborative groups creates opportunities for interaction that lead to improved cognitive learning for all students (Valls & Kyriakides, 2013). Post-secondary empirical studies of team-based learning have shown evidence of student achievement gains, higher student satisfaction and engagement, and positive effects on team diversity (Kearney et al., 2009; Newmann et al., 1992; Roberge & Van Dick, 2010; Stahl et al., 2010; Stipek, 1996; Swanson et al., 2019).

There is little empirical literature at the elementary or secondary level that directly evaluates hypotheses of increased student achievement from academic teaming. We have found only one study that directly assessed that relationship (Basileo, 2018). The study used propensity score matching to create a comparable control group by matching teaming students to similar students in a school district whose teachers did not receive the professional development on teaming. The results of the study revealed statistically significant and positive effects in both reading and math.

Developing Resilience through Productive Struggle

Instruction that encourages and supports productive struggle conveys the teacher's belief that all students can learn and provides opportunities for students to persist as they develop resilience and efficacy in the face of challenging tasks (Ewing et al., 2019; Livy et al., 2018; Reeve, Cheon, & Yu, 2020; Reeve & Halusic, 2009; Reeve & Su, 2014; Warshauer, 2015; Zeybek, 2016). The term *productive struggle* refers to the effort students must expend to make sense of what they are learning when answers are not readily apparent (Edwards & Beattie, 2016; Hiebert & Grouws, 2007). Although the concept of productive struggle has its roots in mathematics education, it is applicable to all academic subjects and is essential to transformative learning processes (Murdoch et al., 2020). It develops

students' skills of conceptual transfer among disparate knowledge disciplines (Sinha & Kapur, 2021; Valentine & Bolyard, 2018).

Building Student Autonomy through Teaming

Teachers develop student autonomy for learning by creating proper structures and a supportive, democratic learning environment for students to work autonomously in teams (Han, 2020; Karademir & Akgul, 2019; Stefanou et al., 2004; Reeve, 2009; Reeve & Cheon, 2021; Reeve & Jang, 2006; Reeve & Su, 2014). Cognitive autonomy leads to the deepest level of student engagement in learning (O'Brien, 2018; Wehmeyer, 1997). Students develop autonomy and self-determination as they acquire skillsets of choice-making, decision-making, problem-solving, goal setting and attainment, self-regulation, self-efficacy, self-awareness, and self-knowledge within academic teams (Kosko, 2015; Marshik et al., 2017; Ryan & Deci, 2017).

Verifying Learning to Take Action within a Lesson

This pillar measures teacher monitoring and use of instructional adjustments, actionable teacher, and peer feedback to improve learning processes, and student self-verification of learning.

Teacher Monitoring and Instructional Adjustments

During lessons, teachers recognize trouble spots and respond with micro-interventions – adjustments made during the lesson to clarify understanding and regain the flow of reasoning (Alibali et al., 2013; Bonne, 2016). Students whose teachers make these adjustments earlier perform at higher levels than students whose teachers delay instructional adjustments (Coyne et al., 2013).

Teachers monitor student progress by examining evidence of learning. When teachers use evidence to adjust instruction and provide feedback to move learning forward – and when students use it to adjust their learning strategies – it is formative assessment (Black & Wiliam, 2009; Panadero et al., 2018). A large body of evidence shows that formative assessment practices can improve student

outcomes (Black & Wiliam, 1998; Crooks, 1988; Kingston and Nash, 2011; Klute et al., 2017; Wiliam et al., 2004).

Actionable Feedback

Although teachers believe that feedback is important to improve learning and student self-confidence, they find it exceedingly difficult to do it well (Dessie & Sewagegn, 2019; Van den Bergh et al., 2014). Effective teacher feedback uses respectful, non-evaluative, flexible, and informational language that encourages student engagement while helping students understand and correct their own behavior or performance (Hattie & Yates, 2014; Reeve & Halusic, 2009). Actionable feedback clarifies expectations and standards, it occurs during learning, and it provides information about the next steps after the present learning is achieved (Brooks, et al., 2019; Fluckiger et al., 2010; Guskey, 1996, 2001, 2007; Hattie, 2023; Hattie & Timperley, 2007; Looney, 2011; Shute, 2007; Stiggins, 2008). Studies have found a positive relationship between effective feedback and student achievement (Hattie et al., 2017; Hattie & Timperley 2007; Vollmeyer & Rheinberg, 2005), but feedback is effective only if learners receive it and understand it (Hattie et al., 2017). Peer feedback, through ongoing classroom dialog among students, supports understanding as it builds student metacognitive skills and self-regulation (Braund & DeLuca, 2018; Davies, 2001). It also provides a sense of belonging and self-worth (Shin & Johnson, 2021).

Students Verify Their Learning

Student verification of their own learning is a form of self-assessment in which students evaluate the results and processes of their work (Brown & Harris, 2014). Student self-assessment is a learnable competence that is linked to improved motivation, self-efficacy, self-regulation, behavior, and the quality of relationships between students and teachers (De Smedt & Van Keer, 2018; Glaser et al., 2010; Griffiths & Davies, 1993; Munns & Woodward, 2006; Olina & Sullivan, 2002; Panadero et al., 2015; Schunk, 1996). Accurate self-assessment happens when students use exemplars or other assessment

data such as tests and graded work and participate in establishing criteria for work quality (Brown & Harris, 2014; Harrison et al., 2015; Ross, 2006).

Self-assessment works most effectively when students receive feedback from peers within a classroom climate that provides psychological safety for self-evaluation (Brown & Harris, 2014; Harrison et al., 2015; Ross, 2006). Further, when peers hold each other accountable, they feel motivated to perform at higher levels and become more self-reliant as learners (Panadero et al., 2015; Stein et al., 2016).

Using Data to Track Student Progress toward Standards

This pillar measures the effectiveness of school leaders and teachers to incorporate short-, mid-, and long-cycle data systematically to improve the quality of teaching and learning in the school and the degree to which the school leader uses data to ensure teacher accountability for student learning.

Short-, Mid- and Long-Cycle Data to Improve Student Learning

Effective school leaders and teachers use short-, mid-, and long-cycle data to continuously improve the quality of teaching and learning in the school (Wiliam, 2006). Long-cycle data spans marking periods, semesters, or years. Educators can collect medium-cycle data within and between teaching units and short-cycle data within and between lessons.

To be formative, an assessment must bring forth evidence that allows teachers to interpret the learning needs of students and use that evidence to adjust instruction to meet those needs (Black & Wiliam, 1998, 2009). Examination of short-cycle formative, mid-cycle interim, and long-cycle summative data provides essential elements for decision making for schools and districts.

Teacher Accountability

The goal of teacher accountability is for teachers to hold themselves accountable and work within their peer teams to successfully ensure students are progressing toward mastery of the standard.

The level of accountability influences teacher assessments of student performance and the cognitive processes they use to make judgements. Such judgements are either category-based or attribute-based. Teachers with low accountability tend to use category-based processing to determine the level of student performance, while teachers with high accountability use attribute-based processing. Category-based processing involves social categories or stereotypes, requires simple cognition, and can be inaccurate (Fiske, 1993; Fiske & Neuberg, 1990). Attribute-based processing considers numerous different personal attributes, requires complex cognition, and is more accurate and detailed (Fiske & Neuberg, 1990; Krolak-Schwerdt et al., 2013). Given the importance of teacher judgement in assessing student performance, accountability is an important driver of accuracy and the elimination of assessment bias.

For teachers to be accountable, they must have autonomy to control the conditions which contribute to successful student learning. The degree of professional autonomy that teachers exercise in the performance of their duties influences the quality of teaching and student learning (Ingersoll & Collins, 2017). Teachers' professional autonomy is positively associated to student autonomy and self-regulation as learners (Ahn et al., 2021; Soenens et al., 2012). But recent top-down accountability requirements for school improvement in the United States have decreased the autonomy that teachers and school leaders exercise over major decisions affecting the quality of instruction as well as school climate (Ryan & Deci, 2020; Ryan & Weinstein, 2009). Ingersoll and Collins (2017) found that schools with the most centralized decision making and least teacher autonomy are often the lowest performing.

Successful school and teacher accountability depend upon leadership. Leaders must establish a clear path to the goal, support execution through team interdependence, and measure performance (Jamal et al., 2015). This requires an open and caring climate that allows honest informational feedback about improving instructional practice as it encourages and inspires teachers to high levels of

performance (Freed et al., 2021; McCarley et al., 2016). Leaders must be able to identify students who need interventions and hold teachers accountable for the appropriate actions to ensure learning.

In summary, we outlined the theoretical research base of the Rigor Appraisal which measures five critical components of rigor: Creating Conditions for Learning Rigorous Standards, Using Standards-based Student Evidence, Activating Student Teams to Achieve the Standard, Verifying Learning to Take Action within a Lesson, and Using Data to Track Student Progress toward Standards. Next, we discuss the Methods and Data used in the investigation.

Data & Methods

The Rigor Appraisal was developed to collect data that provides an objective, non-evaluative measure of essential functions and systems of schools that contribute to optimal instructional and academic performance. The instrument has been used extensively in schools nationwide to help school leaders identify and implement measurable, systems-based pathways for whole-school improvement. The instrument has also been used as a leading indicator to improve the performance of schools with low academic scores, as well as in schools with satisfactory scores that seek to optimize the quality and rigor of instruction and student learning.

This study represents the application of a theoretically backed observational instrument in a real world setting to help gauge its use as a leading indicator of academic achievement. It represents one study of a series of studies to assess the predictability of the tool. We use the term *leading indicator* according to the definition used by the National Academies of Sciences, Engineering, and Medicine (2019). An indicator is leading if it provides early signals that allow educators at the school and district level to predict performance and make strategic and timely investments of time, effort, and resources to improve student learning (Foley et al., 2008). Leading indicators also inform productive inquiry about conditions and organizational systems that affect the quality of education students experience (Supovitz

et al., 2012). By contrast, large scale standardized assessment scores do not meet these criteria and are lagging measures of school effectiveness (May & Sanders, 2013).

Using the Rigor Appraisal metric requires its scorers to be highly calibrated prior to implementation. Each school year all third-party instructional coaches receive training and certification on the Rigor Appraisal instrument. These coaches engage in a rigorous process of understanding the metric, its intent, and the process of capturing evidence of rigorous instruction.

Third-party coaches must undergo a 6-hour Measuring Implementation training course. During this training, participants experience a deepened conversation around pedagogy, core instructional actions, and academic teaming structures. Each of the training components covers elements of the instrument, or pillars. During the training, participants practice scoring the metric using classroom curated videos. Following this, participants have up to three calibration sessions in the field. During these sessions, they walk side by side with a highly trained and certified leader coach to discuss, share, and calibrate specific scores relative to the Rigor Appraisal instrument. At the end of the training, these coaches are tested and certified for scoring the instrument for the school year.

An experienced leader coach accompanies newly certified coaches on their first Rigor Appraisal to calibrate scoring and ensure accuracy. Throughout the first year, they experience regular calibration during monthly one-on-one visits with the leader coach. These visits deepen their knowledge and understanding of rigorous instruction and how to capture it using the instrument. Leader coach visits also build the capacity of coaches to support school leadership teams in strengthening schoolwide systems observed through the Rigor Appraisal.

In this study, third-party certified coaches completed three appraisals per school building during the school year. During a Rigor Appraisal, the coach met onsite with the principal and school leadership team. Together, they walked the school to collect anonymous information from a randomly selected

sample of classrooms. Afterwards, they debriefed on their observations. They used the Rigor Appraisals to establish baseline conditions and determine next steps to implement support structures. The coach scored the schoolwide Rigor Appraisal in its entirety within 24 hours of each walk. They used the aggregated results to guide the school leadership team in identifying systems-based root causes and actions for change. The findings also informed the selection of professional development for teachers and school leaders.

To build the internal capacity of the school and district, school leadership teams received training in a similar Measuring Implementation course to conduct instructional classroom walks that are nonrandom. These aided the school leadership teams in moving the work forward in between the random Rigor Appraisals. The purpose of the layered walks was to provide focus and build internal capacity, empowering school leaders to engage students in more rigorous work.

Research Questions

The following research questions guide this study and are all based at the school level:

1. Are school rates of learning correlated with student proficiency?
2. Is the Rigor Appraisal correlated with student achievement?
3. Are the Rigor Appraisal pillars related to student achievement?
4. Is the Rigor Appraisal correlated with other school effectiveness measures?
5. Are program effectiveness measures correlated with student achievement?
6. Are school wellness measures correlated with the Rigor Appraisal and student achievement?
7. Does the Rigor Appraisal predict student achievement?

We would expect that a positive change in the Rigor Appraisal scores over time would be positively correlated with student achievement and other leading indicators of school effectiveness such as student attendance and teacher retention. Additionally, we would expect scores to have a negative

relationship with student referrals and student suspensions. In this section, we first describe the study sample, dependent and predictor variables including how they are calculated and operationalized, and the methodology used for the analysis.

Study Sample

The study sample consisted of 48 to 53 schools in one district in Illinois that received Rigor Appraisals within the 2021-22 school year and had student achievement scores. The district served a diverse student population of approximately 35,000. The Illinois State Board of Education (2023) classified 46.4 percent of students as Low Income, 39.1 percent as English Learners, and 18.6 percent as Students with Disabilities. Table 1 shows the demographic distribution of students by race and ethnicity.

Table 1

Racial and Ethnic Distribution of Students Within the District

Race/ Ethnicity	Percentage
American Indian	1.1%
Asian	8.5%
Black	6.2%
Hispanic	56.4%
Two or More Races	3.3%
Pacific Islander	0.1%
White	24.4%

Dependent Variables: Student Achievement

We calculated two student achievement measures, student proficiency and school learning rates, for English Language Arts (ELA) and math to assess correlation levels with school and program effectiveness measures. Table 2 shows how student achievement was operationalized. Each will be described in more detail next.

Table 2*Outcome Variables*

Student Achievement Measures	Operational Definition
ELA Proficiency	The percent difference between the number of students who met or exceeded proficiency on the Illinois Assessment of Readiness (IAR) in ELA from 2021 to 2022
Math Proficiency	The percent difference between the number of students who met or exceeded proficiency on the IAR in math from 2021 to 2022
ELA Learning Rate	Weighted Hedges' g effect size of the mean ELA scale score Spring 2021 to Spring 2022 accounting for student sample size within grade levels
Math Learning Rate	Weighted Hedges' g effect size of the mean math scale score Spring 2021 to Spring 2022 accounting for student sample size within grade levels

We used the Illinois Assessment of Readiness (IAR) ELA and math scores to gauge change in proficiency, and calculated change in proficiency by taking the difference between the percent of students who met or exceeded proficiency on the IAR in ELA and math from 2021 to 2022. The change in proficiency is used in this analysis because it is a common state assessment measure that is easily available, and many states report on it. Additionally, the main use for the Rigor Appraisal instrument is to be a leading indicator of school change, thus, change will always be investigated where possible. While proficiency is one measure used in the investigations, it is only one proxy of student achievement.

While the percentage of proficient students is a common measure of student achievement, it purely measures the number of students below, at, or above a certain scale score level. Proficiency rates do not exist outside of the value systems of those setting the cut scores (Cizek, 1993; Jaeger, 1989; McClarty et al., 2013; Shepard, 1979). Interpretations of the percent of proficient students are unpredictable and can lead to incorrect or incomplete inferences about distributional change (Ho, 2008). We purposely move away from proficiency due to the known statistical and substantive costs of making comparisons due to their limited and unrepresentative depictions of test score trends.

Furthermore, when operating in lower performing schools, the percent proficient is less likely to have variability compared to investigating growth on scale scores across two points in time. Lower performing students can make growth but still fall below proficiency depending on how far they are away from the cut score at baseline. Thus, we include two additional proxies of student achievement for this study: learning rates in ELA and math.

We refer to this as a school level learning rate. Using Bloom and colleagues (2008) as a guideline, we calculated effect sizes for each grade-to-grade pair (third to fourth, fourth to fifth, etc.) for each school in the district. We use this approach because Bloom and colleagues (2008) found that regardless of the assessment, student growth in achievement is greatest in elementary grades (K to first grade is about 1.5 standard deviations), then declines throughout middle and high school (11th to 12th graders is about .05 standard deviations). If the number of students within each grade level are not accounted for, schools with a larger number of younger students will appear to achieve more. Thus, learning rates need to account for the average growth and the number of students within each grade level.

We used the Hedges' g effect size formula for each grade level pair as it is the most appropriate effect size when accounting for small samples (What Works Clearinghouse, 2022). This effect size metric represents the mean difference between two groups of observations in two points in time on a standard deviation scale. More specifically, we calculated effect sizes for each grade level pair using all students in the district who were tested in Spring 2021 and Spring 2022 for ELA and math. Then, we aggregated the pairs to the school by using weighted averages that accounted for the number of students within each grade to create an overall school learning rate for each subject area. Consequently, the school learning rates reflect both the effects of schooling plus other developmental influences on students during the year. Hedges' g formula is presented as Equation 1 below.

Hedges' g Formula

(1)

$$g = \frac{\omega(y_i - y_c)}{\sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}}}$$

- y_i is the adjusted (or unadjusted) mean post intervention group
- y_c is the adjusted (or unadjusted) mean pre intervention group
- n_i is the sample size of the post intervention group
- n_c is the sample size of the pre intervention group
- s_i is the unadjusted standard deviation for the post intervention group
- s_c is the unadjusted standard deviation for the pre intervention group
- Omega (ω) is the small sample size correction
 - It is equal to $1 - \frac{3}{4N-9}$ where N is the total sample size (or $n_i + n_c$)

Since the IAR is not a vertically scaled instrument and there were no other Fall to Spring or Spring to Spring district assessment data available, we tested the premise that Hedges' g could still be calculated with a nonvertically scaled assessment. If we do not assess this premise, effect sizes based on a nonvertically scaled assessment could produce erroneous ordering of schools and distort correlation coefficients.

Following Briggs and Domingue's (2013) investigation into the impact of gain scores, we assessed each grade-to-grade pair of effect sizes prior to calculating the weighted school averages. They found that non-vertical scaling impacted grade-to-grade comparisons only when there was the presence of a large degree of scale shrinkage which restricts the range of gain scores (Briggs and Dominique, 2013). Furthermore, correlation levels dropped from .87 to .57 when grade-to-grade standard deviations were -.30 or less. Consequently, anything equal or less than -.30 could have an impact on the ordering of schools as a function of average gain scores.

In our study, we found that only 3% (4 out of 139) of our grade-to-grade pairs had standard deviations less than or equal to -.30. Consequently, with such a small impact, we concluded that this

would not have a significant effect on the ordering of schools and the effect sizes should represent a more precise measure for the ordering of schools than proficiency gains. The next step in the investigation was to test this assumption by examining the correlations between learning rates and student proficiency.

One limitation to outcome variables during the pandemic school years relates to missing test scores for students. In the calculation of learning rates, we kept all tested students in distribution to reflect the true population of students more accurately in 2022 as we did not want to inadvertently create testing bias by removing students that were not tested in 2021. About 35 to 48 percent of scores in 2021 were missing for grade level pairs district wide.

We used several models to investigate the hypothesis that students not tested in 2021 were different than the population tested in 2022. We found statistically significant differences: those who did not take the 2021 assessment were significantly more likely to score 2.31 scale score points lower on the 2022 ELA assessment and 4.9 scale score points lower in math, controlling for student characteristics and accounting for the clustering of students within schools. Additionally, Black, Hispanic, and low-income students were significantly more likely not to take the 2021 ELA state assessment, and students identified by the district as gifted were significantly more likely to take it (these findings also held true for math). Consequently, our learning rate calculations do not account for missing student data for the baseline year. Furthermore, changes in proficiency rates also do not consider missing test scores.

Predictor Variables

Table 3 shows all predictor variables used in the correlation analysis. The first set of indicators measures the Rigor Appraisal, followed by traditional school effectiveness measures, program effectiveness measures, and school wellness as measured by the Illinois State Board of Education, 5Essentials survey. Each will be described in more detail next.

Table 3*Predictor Variables Operationalized*

Variable	Operational Definition
<i>Rigor Appraisal</i>	
Rigor Appraisal	The raw Rigor Appraisal change score from Fall to Spring 2021 to 2022
Rigor Appraisal Beta	Ordinary Least Squares Beta coefficient of the change in Rigor Appraisal scores across the number of school days
Creating Conditions for Learning Rigorous Standards	The raw pillar change score from Fall to Spring 2021 to 2022
Using Standards-based Student Evidence	The raw pillar change score from Fall to Spring 2021 to 2022
Activating Student Teams to Achieve the Standard	The raw pillar change score from Fall to Spring 2021 to 2022
Using Data to Track Student Progress toward Standards	The raw pillar change score from Fall to Spring 2021 to 2022
<i>School Effectiveness</i>	
Weekly Attendance	The average weekly percent of students who attended school from 2021-2022
Total Referrals	The total number of referrals at a school from 2021-2022
Total Suspensions	The total number of suspensions at a school from 2021-2022
Teacher Retention	The difference in the rate of teacher retention from 2022 to 2021
<i>Program Effectiveness</i>	
Instructional Walks	The total number of school leadership instructional walks conducted at the school from Fall to Spring 2021 to 2022
ELA Instructional Walks	The total number of school leadership instructional walks conducted in ELA classrooms at the school from Fall to Spring 2021 to 2022
Math Instructional Walks	The total number of school leadership instructional walks conducted in math classrooms at the school from Fall to Spring 2021 to 2022
Coaching Days	The total number of third-party school leadership and faculty coaching days the school received from Fall to Spring 2021 to 2022
<i>School Wellness</i>	
5Essentials (5E)	The average difference across pillars of the 5 Essentials from 2021-2022
Ambitious Instruction	The percent difference of Ambition Instruction from 2022 to 2021
Supportive Environment	The percent difference of Supportive Environment from 2022 to 2021
Involved Families	The percent difference of Involved Families from 2022 to 2021
Collaborative Teachers	The percent difference of Collaborative Teachers from 2022 to 2021
Effective Leaders	The percent difference of Effective Leaders from 2021 to 2022

The Rigor Appraisal

We calculated the raw score change from the first baseline observations from Fall of 2021 to Spring of 2022. Table 4 shows the scale of the Rigor Appraisal instrument. It consists of 12 options, ranging from strongly disagree to strongly agree. Each pillar contains three to four items, each receiving a score between one and 12. Scores for each item within a pillar are averaged, then each of the five pillar scores are averaged to determine the overall Rigor Appraisal score for the observation. The software tool includes a scoring rubric of the criteria within each agreement category.

Table 4

The Rigor Appraisal Scale

Strongly Disagree			Disagree			Agree			Strongly Agree		
1	2	3	4	5	6	7	8	9	10	11	12

We also investigated the reliability of the instrument. The sample was larger than the current study sample as it included all 116 schools in the United States that received three Rigor Appraisals during the school year. The reliability standard aims to set principles for maximum allowable measurement error. Higher reliability indicates lower measurement error. Cronbach's alpha assesses internal consistency and can capture measurement errors that result from poor question wording. Cronbach's alpha of .6 and higher is considered a reliable measure (What Works Clearinghouse, 2022). Table 5 shows that Cronbach's alpha for all pillars deemed these measures as reliable with the lowest at .734.

Table 5*Cronbach Alpha of Rigor Appraisal Pillars*

Pillar	Fall	Interim	Spring	Average
Creating Conditions for Learning Rigorous Standards	.734	.851	.794	.793
Using Standards-based Student Evidence	.911	.932	.870	.904
Activating Student Teams to Achieve the Standard	.915	.946	.854	.905
Verifying Learning to Take Action within a Lesson	.869	.898	.840	.869
Using Data to Track Student Progress toward Standards	.953	.956	.909	.939

We also calculated the Rigor Appraisal Beta score using Ordinary Least Squares to measure the predicted growth trend rather than the raw growth over time. The Beta coefficient estimates the magnitude of change over the course of Rigor Appraisals using the number of days in which the Appraisals take place. It reduces measurement error of coders as it smooths out the trendline over time.

School Effectiveness

School effectiveness measures included all traditional leading indicators of student achievement. We calculated the average weekly attendance by taking the average percent of students that attended the school across the school year. We also calculated the total number of behavior referrals at each school for the 2021-2022 school year as well as the the sum of the number of out of school suspensions for the same school year. Typically, we would investigate the change over time but because of the pandemic and virtual schooling the rate change was unavailable. Additionally, we calculated teacher retention by using the rate change from 2022 to 2021.

Program Effectiveness

Program effectiveness measures included four measures, the total number of instructional walks conducted by the principal and school leadership team, the number of instructional walks they conducted in ELA and math, and the number of third-party leadership and faculty coaching days focused on instructional systems improvement during the school year. The walks are non-evaluative observations conducted by the principal, assistant principals, and/or coaches to look for certain aspects of rigorous instruction in the classroom. They differed from the Rigor Appraisal as they were quick non-evaluative observations by school administrators or coaches whereas a Rigor Appraisal randomly conducted by a certified third-party consultant. Instructional observation walks are scored in a software tool that analyzes them and provides trend data. The principal's actions should lead to observable changes in classroom instruction, so these data are important indicators to assess program effectiveness. We wanted to evaluate the impact of the program effectiveness measures and included them in the final OLS regression models.

School Wellness

Chicago Impact at the University of Chicago administers The Illinois 5Essentials Survey (Illinois State Board of Education, n.d.) online on behalf of the Illinois State Board of Education. The State Board has recognized that test scores alone do not provide a full picture of teaching and learning in any one school. Under the Illinois Compiled Statutes (2018), the State Board must implement a learning conditions survey annually. The 5Essentials Survey identifies five indicators that lead to improved outcomes: Effective Leaders; Collaborative Teachers; Involved Families; Supportive Environments; and Ambitious Instruction. Schools across Illinois participate in this survey.

We included the 5Essentials survey data in our analysis to assess whether the state indicators of school wellness are correlated, as the survey has shown to be predictive of school achievement in

elementary and high schools (Hart et al., 2020). We measure the change in time for each of the 5Essentials indicators in addition to a change score which averages the change over time.

Methodology

Correlations assess the magnitude and direction of the relationship between two variables. The correlation coefficients can fall between -1 and +1. Correlation analyses are necessary to assess predictability, strength, and direction of observational instruments. They are also helpful in identifying inputs for a more sophisticated analysis, or for understanding the relationship between variables. Cohen's (1988) conventions to interpret effect sizes of .10, .50, and .80 standard deviations as small, moderate, and large. As Bloom and colleagues (2008) point out, these guidelines are not always relevant to intervention effects in education; however, they may be used when there is no better basis for estimating the magnitude of impact across studies (Cohen, 1988).

We also used OLS regression to investigate the relationship between the Rigor Appraisal score, the Rigor Appraisal pillar scores, program effectiveness measures, and school demographic estimates to predict student achievement. School demographic measures included the percent of students with disabilities (SWD) and the percent of low-income students. We also investigated multicollinearity and removed the percent of English Learners from the model due to its high correlation to low income ($r > .8$). Multicollinearity occurs when two or more independent variables have a high correlation with one another, which makes it difficult to determine the individual effect of each independent variable on the dependent variable. We used OLS regression to achieve this goal because the Rigor Appraisal is a school level indicator to assess change at the system level; thus, we wanted to keep the analysis limited to this hierarchy as it is at this level district administrators can track changes in school trends.

Results

Table 6 shows the descriptive statistics of the sample. The number of schools included in the correlation analysis ranged from 48 to 53 depending upon the measures examined. The change in proficiency varied from -25% to an increase of 14% proficiency. The learning rates noted a drop of -.85 standard deviations to an increase of .72 for ELA. Rigor Appraisal change scores also fluctuated from -.45 scale score point decrease to a 3.27 increase in the 12-point scale.

Table 6

Descriptive Statistics

	N	Min.	Max.	Mean	Std. D.
ELA Proficiency	48	-.25	.14	.00	.08
Math Proficiency	48	-.25	.14	.01	.07
ELA Learning Rate	48	-.85	.72	.22	.28
Math Learning Rate	48	-.29	.35	.05	.15
Rigor Appraisal	53	-.10	3.27	.98	.66
Rigor Appraisal Beta	53	-.79	1.00	.83	.32
Conditions for Learning	53	-1.75	2.75	.49	.82
Standards-Based Student Evidence	53	-1.25	4.75	1.31	1.10
Activating Student Teams	53	-1.75	4.00	1.31	1.11
Verifying Student Learning	53	-2.00	4.00	.91	1.07
Tracking Student Progress	53	-1.00	4.25	.88	.99
Weekly Attendance	49	.83	.94	.91	.02
Total Referrals	49	45.00	7387.00	580.92	1098.64
Total Suspensions	49	.00	390.00	28.92	69.28
Teacher Retention	53	-9.00	5.00	.28	2.75
Total Classroom Visits	53	125.00	1282.00	356.45	183.95
ELA Classroom Visits	53	44.00	483.00	118.70	69.93
Math Classroom Visits	53	38.00	568.00	99.98	78.66
Total Coach Days	49	4.50	40.00	11.34	9.53
5 Essentials (5E)	49	-20.00	15.00	-4.46	6.58
Ambitious Instruction	49	-9.00	27.00	9.49	7.46
Supportive Environment	49	-31.00	10.00	-5.31	10.04
Involved Families	47	-25.00	3.00	-10.57	6.96
Collaborative Teachers	47	-33.00	15.00	-6.96	9.92
Effective Leaders	47	-29.00	7.00	-10.96	8.80
Percent Students with Disabilities	53	10.50	28.70	17.28	4.02
Percent Low-Income Students	53	20.80	95.20	58.67	19.77

Next, we will present the results from each of the research questions posed.

Are school learning rates correlated with student proficiency?

Table 7 shows that all correlations were statistically significant. Learning rates had a moderate to large association with ELA and math proficiency. This remained true even after removing the four schools that had grade-to-grade standard deviations that were $\geq .30$ or greater (not shown). Removing these schools from the analysis did not affect the correlation coefficients for ELA and math and they remained statistically significant. It should be noted that we calculated the learning rates using students who tested both years only and they had weaker correlation levels to proficiency. Consequently, school learning rates that included all tested students in both years were reliable predictors of student achievement at the school level.

Table 7

Proficiency and Learning Rate Correlation Coefficients

		ELA Proficiency	Math Proficiency	ELA Learning Rate	Math Learning Rate
ELA Proficiency	r	1	.696**	.582**	.311*
	p-value		<.001	<.001	.032
	N	48	48	48	48
Math Proficiency	r	.696**	1	.449**	.424**
	p-value	<.001		.001	.003
	N	48	48	48	48
ELA Learning Rate	r	.582**	.449**	1	.675**
	p-value	<.001	.001		<.001
	N	48	48	48	48
Math Learning Rate	r	.311*	.424**	.675**	1
	p-value	.032	.003	<.001	
	N	48	48	48	48

** Correlation is significant at the 0.01 level

* Correlation is significant at the 0.05 level

† Correlation is approaching statistical significance at the 0.10 level

Is the Rigor Appraisal Correlated with Student Achievement?

Table 8 shows that the Rigor Appraisal had a small to moderate and statistically significant association with ELA and approached statistical significance with math learning rates, meaning that as the change in the Rigor Appraisal increased, so did student achievement. The Rigor Appraisal and proficiency had a small relationship, and it was not a statistically significant association. Furthermore, the Beta coefficient had a slightly negative relationship with student achievement, and it was not statistically significant.

Table 8

Rigor Appraisal, Beta & Student Achievement Correlation Coefficients

		Rigor Appraisal	Rigor Appraisal Beta	ELA Proficiency	Math Proficiency	ELA Learning Rate	Math Learning Rate
Rigor Appraisal	r	1	.578**	.225	.176	.388**	.267 [†]
	p-value	-	<.001	.123	.231	.006	.067
	N	53	53	48	48	48	48
Rigor Appraisal Beta	r	.578**	1	-.049	-.013	.009	-.083
	p-value	<.001	-	.74	.928	.954	.573
	N	53	53	48	48	48	48

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

[†] Correlation is approaching statistical significance at the 0.10 level (2-tailed).

Are the Rigor Appraisal Pillars Related to Student Achievement?

Table 9 shows that the Rigor Appraisal pillars had a small to moderate and statistically significant association with two pillars: Using Standards-based Student Evidence and Activating Student Teams to Achieve the Standard with the latter having a large correlation to student achievement. Because this was the first year of implementation and these were the areas on which the schools focused, this is not surprising. However, one surprising finding is the negative and statistically significant relationship with

Using Data to Track Student Progress toward Standards and math proficiency. This may have happened by chance; it may have been due to the decline in overall student achievement due to the pandemic.

Additionally, few schools were implementing strategies associated with this pillar and none were implementing the data systems addressed by this pillar. More investigation is necessary.

Table 9

Rigor Appraisal Pillars and Student Achievement

		ELA Proficiency	Math Proficiency	ELA Learning Rate	Math Learning Rate
Creating Conditions for Learning Rigorous Standards	r	.187	.157	.253 [†]	.105
	p-value	.204	.287	.083	.478
	N	48	48	48	48
Using Standards- based Student Evidence	r	.239	.287*	.276 [†]	.295*
	p-value	.102	.048	.058	.042
	N	48	48	48	48
Activating Student Teams to Achieve the Standard	r	.439**	.468**	.589**	.433**
	p-value	.002	<.001	<.001	.002
	N	48	48	48	48
Verifying Learning to Take Action within a Lesson	r	.048	-.022	.151	.135
	p-value	.748	.882	.306	.362
	N	48	48	48	48
Using Data to Track Student Progress toward Standards	r	-.220	-.347*	-.094	-.182
	p-value	.132	.016	.523	.217
	N	48	48	48	48

** Correlation is significant at the 0.01 level

* Correlation is significant at the 0.05 level

[†] Correlation is approaching statistical significance at the .10 level

Is the Rigor Appraisal Correlated with School Effectiveness Measures?

Table 10 shows that the Rigor Appraisal had a positive moderate and statistically significant association with student's average weekly attendance. Further, the Rigor Appraisal had a negative relationship with the change in student referrals and suspensions, meaning that as it increased, student

referrals and suspensions decreased. The Rigor Appraisal was not significantly correlated with the change in teacher retention. Interestingly, while the Beta did not have a relationship with student achievement it had a moderate to large statistically significant relationship with all school effectiveness measures except for teacher retention.

Table 10

Rigor Appraisal, Beta & School Effectiveness

		Avg. Weekly Attend	Referrals	Suspensions	Teacher Retention
Rigor Appraisal	r	.259 [†]	-.388**	-.431**	.106
	p-value	.073	.006	.002	.451
	N	49	49	49	53
Rigor Appraisal Beta	r	.453**	-.730**	-.636**	.141
	p-value	.001	<.001	<.001	.314
	N	49	49	49	53
** Correlation is significant at the 0.01 level					
* Correlation is significant at the 0.05 level					
[†] Correlation is approaching statistical significance at the .10 level					

Are Program Effectiveness Measures Correlated with Student Achievement?

As shown in Table 11, the total number of classrooms observed during instructional walks and the number of ELA and math classrooms walked at the school had a small to moderate statistically significant association with Rigor Appraisal, the change in math proficiency, ELA and math learning rates. In other words, the more the school leadership team observed classrooms, the higher the association with student achievement. Additionally, the number of leadership and faculty coaching days a school received had a small to moderate statistically significant association with ELA and math learning rates. Meaning the more coaching days a school had, the larger the achievement on the state assessment.

Table 11*Program Effectiveness Measures and Student Achievement*

		Rigor Appraisal	ELA Proficiency	Math Proficiency	ELA Learning Rate	Math Learning Rate
Total Instructional Walk Classrooms	r	.287*	.188	.277 [†]	.336*	.432**
	p-value	.037	.201	.057	.019	.002
	N	53	48	48	48	48
ELA Instructional Walk Classrooms	r	.190	.348*	.435**	.435**	.405**
	p-value	.173	.015	.002	.002	.004
	N	53	48	48	48	48
Math Instructional Walk Classrooms	r	.179	.207	.315*	.394**	.376**
	p-value	.201	.158	.029	.006	.009
	N	53	48	48	48	48
Coaching Days	r	.142	.223	.114	.356*	.280 [†]
	p-value	.319	.128	.439	.013	.054
	N	51	48	48	48	48

** Correlation is significant at the 0.01 level

* Correlation is significant at the 0.05 level

[†] Correlation is approaching statistical significance at the .10 level

Are School Wellness Measures Correlated with Rigor Appraisal and Student Achievement?

The average change in the 5Essentials had a moderate and statistically significant association with Rigor Appraisal and all measures of student achievement. In other words, as the 5Essentials increased year over year, so did the Rigor Appraisal and student achievement. Almost all the 5Essentials categories had statistically significant associations to the Rigor Appraisal except for Involved Families which did not correlate well with any other measure of student achievement. Table 12 shows these associations.

Table 12*The 5Essentials, Rigor Appraisal and Student Achievement*

		Rigor Appraisal	ELA Proficiency	Math Proficiency	ELA Learning Rate	Math Learning Rate
5Essentials Average Change	r	.459**	.415**	.386**	.403**	.331*
	p-value	<.001	.003	.007	.005	.022
	N	51	48	48	48	48
Ambitious Instruction	r	.361**	.328*	.067	.298*	.214
	p-value	.009	.023	.653	.040	.144
	N	51	48	48	48	48
Supportive Environment	r	.270 [†]	.526**	.419**	.589**	.457**
	p-value	.055	<.001	.003	<.001	.001
	N	51	48	48	48	48
Involved Families	r	.136	.153	.119	.157	.071
	p-value	.351	.309	.431	.298	.638
	N	49	46	46	46	46
Collaborative Teachers	r	.425**	.179	.338*	.221	.199
	p-value	.002	.233	.022	.140	.185
	N	49	46	46	46	46
Effective Leaders	r	.284*	.115	.254 [†]	.271 [†]	.237
	p-value	.048	.449	.088	.068	.113
	N	49	46	46	46	46

** Correlation is significant at the 0.01 level

* Correlation is significant at the 0.05 level

[†] Correlation is approaching statistical significance at the .10 level**Does the Rigor Appraisal Predict Student Achievement?**

In this section we will review the results from two models which used OLS regression to predict student achievement. Three different models are shown in Table 13. The first uses the program effectiveness measures and predicts the ELA learning rate in addition to the two school demographic measures. All measures besides the school demographic measures are statistically significant and ELA Instructional Walk Classrooms is the largest predictor in the model with the Adjusted R-squared explaining 30% of the variance for the ELA learning rate that is explained by the predictor variables.

The second model removes the full Rigor Appraisal measure and adds the pillars. Interestingly, all program effectiveness measures besides ELA Instructional Walk Classrooms drop out of statistical significance leaving Activating Student Teams to Achieve the Standard as the largest predictor of ELA achievement. Because we saw large correlations between Activating Student Teams to Achieve the Standard and ELA and math achievement, we created an interaction effect between that pillar and the percent of low-income students. Student teaming in low-income schools may increase student learning (Toth & Sousa, 2019). The third model shown only includes the interaction term with coaching days, and ELA Instructional Walk Classrooms. Interestingly, when there is a higher percentage of low-income students at a school and classrooms are showing an increase in the quality of academic teaming, it has a greater impact on ELA achievement. Thus, academic teaming has a greater impact on student achievement in schools that have a higher percentage of low-income students.

Table 13.

The Effects of Rigor Appraisal and ELA Achievement

ELA Learning Rate (N=48)	Beta	SE	P-Value	Beta	SE	P-Value	Beta	SE	P-Value
Coaching Days	.277*	.004	.036	.221	.004	.113	.146	.004	.258
ELA Instructional Walk Classrooms	.327*	.001	.015	.328*	.001	.018	.210 [†]	.000	.096
Rigor Appraisal	.283*	.064	.031	-	-	-	-	-	-
SWD	.050	.009	.698	.039	.009	.752	.065	.008	.583
Low Income	.105	.002	.427	.101	.002	.446	-	-	-
Conditions	-	-	-	.103	.050	.436	-	-	-
Standards	-	-	-	-.075	.042	.616	-	-	-
Activating	-	-	-	.486**	.045	.003	-	-	-
Verifying	-	-	-	-.166	.040	.235	-	-	-
Tracking	-	-	-	.173	.039	.216	-	-	-
Activating Teaming*Low Income	-	-	-	-	-	-	.501**	.001	<.001
Adjusted r ²			.305			.378			.406

**Significant at the .01 level; *Significant at the .05 level; [†]Approaching statistical significance at the .10 level

Table 14 demonstrates that same impact on the math learning rate. The only differences in this case are that Low Income is statistically significant in the first model and Verifying was not statistically significant in the second model. However, the combined interaction effect between Activating Student Teams to Achieve the Standard and math achievement remained with the Adjusted R-squared values increasing from model 1 to 3. These models also demonstrate the need for instructional leaders to non-evaluatively observe teacher classrooms. In these instances, the number of classrooms visited had an impact on student achievement regardless of all the other variables in the model.

Table 14

The Effects of Rigor Appraisal and Math Achievement

Math Learning Rate (n=48)	Beta	SE	P-Value	Beta	SE	P-Value	Beta	SE	P-Value
Coaching Days	.218	.002	.108	.089	.002	.548	.129	.002	.348
Math Instructional Walk Classrooms	.269*	.000	.049	.253 [†]	.000	.077	.228 [†]	.000	.086
Rigor Appraisal	.146	.035	.273	-	-	-	-	-	-
SWD	-.090	.005	.507	-.070	.005	.602	-.112	.005	.376
Low Income	.280*	.001	.048	.343*	.001	.021	-	-	-
Conditions	-	-	-	-.078	.029	.584	-	-	-
Standards	-	-	-	.197	.024	.229	-	-	-
Activating	-	-	-	.241	.026	.149	-	-	-
Verifying	-	-	-	-.139	.023	.355	-	-	-
Tracking	-	-	-	-.003	.022	.983	-	-	-
Activating *Low Income	-	-	-	-	-	-	.416**	.000	.006
Adjusted r ²			.252			.271			.309

** Significant at the .01 level; * Significant at the .05 level; [†]Approaching statistical significance at the .10 level

Conclusion

In summary, we found that the Rigor Appraisal score had a small to moderate and statistically significant association with student achievement, a positive moderate and statistically significant

association with average weekly attendance, a negative relationship with changes in student referrals and suspensions, and a moderate positive and statistically significant association with the 5Essentials school survey. Further, one pillar of the Rigor Appraisal, Activating Student Teams to Achieve the Standard, was associated with greater achievement in schools with more low-income students. We also found that the number of classroom instructional walks conducted by the leadership team had a positive and statistically significant impact on student achievement.

More importantly, we obtained these findings during a time when there was a national decline in student achievement, and we did not see the same kind of regression in our analysis that we saw nationally. The Rigor Appraisal process could have helped to slow the decline in achievement that occurred elsewhere across the nation and/or the interventions themselves could have had an impact. Regardless of the exact cause, the instrument's predictability appears strong in this context.

Despite these strengths, there are limitations. First, there were disruptions to in-person instruction which limited the times during which observations could occur. There were also unanticipated changes in personnel assigned to the teaching teams which resulted in fewer classrooms observed consistently throughout the period of this study. The pandemic also impacted the amount of missing state assessment data for the baseline year in 2021. Further, the statistically significant negative correlation between the pillar Using Data to Track Student Progress toward Standards and math proficiency was in the unexpected direction. More investigation is needed into why that association was negative. Finally, the data used in this study was from only one school district, therefore the findings may not be generalizable.

While the analysis was based on correlational analyses to determine the validity of the Rigor Appraisal instrument as a leading indicator of achievement, correlation does not infer causality. The OLS models attempted to go beyond correlational analysis and demonstrated the impact of program

effectiveness and Rigor Appraisal measures on student achievement. However, more investigation is warranted. Areas of future work should include replicating these findings using path analysis for each of the Rigor Appraisal pillars and their associations with student achievement. In this study, the pillar Activating Student Teams to Achieve the Standard had the largest impact on student achievement, but does this always hold true or only in certain circumstances? One explanation for this finding is a direct link between Activating Student Teams to Achieve the Standard and observed student learning. If the other pillars do not directly impact student achievement, in what contexts do they matter?

Another area of future work will include a much more in-depth assessment of interrater reliability between the third-party certified coaches in addition to the school leadership teams who complete the instructional classroom walks. Developing this knowledge within the district is essential for reform. Furthermore, building the internal capacity of school and district leaders is essential to closing achievement gaps and making up for lost learning.

The Rigor Appraisal can be a valuable tool to aid schools in recovering from the effects of the pandemic, improving educational equity, and eliminating academic performance disparities for students from low-income backgrounds. It also has the potential of reducing the need to interrupt instructional time for interim or benchmark testing to assess student progress. Using its leading indicators can help educators make better informed, proactive, real-time changes to instruction that accelerate learning, improve the quality of teaching, and improve academic outcomes for all students.

References

- Ahn, I., Chiu, M. M., & Patrick, H. (2021). Connecting teacher and student motivation: Student-perceived need-supportive practices and student need satisfaction. *Contemporary Educational Psychology*, 64, Article 101950. <https://doi.org/10.1016/j.cedpsych.2021.101950>
- Akman, Y., (2021). The relationship between school climate and students' aggressive behaviors. *International Journal of Progressive Education*, 17(2), 430-448. <https://doi.org/10.29329/ijpe.2021.332.26>
- Aldridge, J., Fraser, B. J., Fozdar, F., Ala'i, K., Earnest, J., & Afari, E. (2018). Students' perceptions of school climate as determinants of wellbeing, resilience and identity. *Improving Schools*, 19(1), 5-26. <https://doi.org/10.1177/1365480215612616>
- Alibali, M. W., Nathan, M. J., Church, R. B., Wolfgram, M. S., Kim, S., & Knuth, E. J. (2013). Teachers' gestures and speech in mathematics lessons: Forging common ground by resolving trouble spots. *ZDM Mathematics Education*, 45, 425-440. <https://doi.org/10.1007/s11858-012-0476-0>
- Anees, S. (2017). Analysis of assessment levels of students' learning according to cognitive domain of Bloom's taxonomy. Online submission ED586762. <https://eric.ed.gov/?q=Analysis+of+Assessment+Levels+of+Students%e2%80%99+Learning+according+to+Cognitive+Domain+of+Bloom%e2%80%99s+Taxonomy&id=ED586762>
- Argon, T., & Ekinci, S. (2016). Teachers' view on organizational deviance, psychological ownership and social innovation. *Universal Journal of Educational Research*, 4(12A), 133-139. <https://doi.org/10.13189/ujer.2016.041317>
- Azevedo, J. P., Hasan, A., Goldemberg, D., Iqbal, S. A., & Geven, K. (2020). *Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global*

- estimates*. The World Bank. <https://thedocs.worldbank.org/en/doc/798061592482682799-0090022020/original/covidandeducationJune17r6.pdf>
- Basileo, L.D. (2018). *How a great city school district is improving student achievement, increasing equity and closing achievement gaps*. Instructional Empowerment. <https://vb8e8a.p3cdn1.secureserver.net/wp-content/uploads/2023/03/IE01-121-Research-Report-Des-Moines-03-23.pdf>
- Becker, D. R., Miao, A., Duncan, R., & McClelland, M. M. (2014). Behavioral self-regulation and executive function both predict visuomotor skills and early academic achievement. *Early Childhood Research Quarterly*, 29, 411-424. <http://dx.doi.org/10.1016/j.ecresq.2014.04.014>
- Benbenishty, R., Astor, R. A., Roziner, I., & Wrabel, S. (2016). Testing the causal links between school climate, school violence, and school academic performance; A cross-lagged panel autoregressive model. *Educational Researcher*, 45(3), 197-206. <https://doi.org/10.3102/0013189X16644603>
- Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2016). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research*, 87(2), 425-469. <https://doi.org/10.3102/0034654316669821>
- Betthäuser, B. A., Bach-Mortensen, A. M., Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behavior*, 7, 375-385. <https://doi.org/10.1038/s41562-022-01506-4>
- Briggs, D. C., & Domingue, B. (2013). The gains from vertical scaling. *Journal of Educational and Behavioral Statistics*, 38(6), 551-576. <https://doi.org/10.3102/1076998613508317>
- Black, P., & Wiliam, D. (1998) *Inside the Black Box: raising standards through classroom assessment*. School of Education, King's College.

- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Blitz, L. V., Yull, D., & Clauhs, M. (2016). Bringing sanctuary to school: Assessing school climate as a foundation for culturally responsive trauma-informed approaches for urban schools. *Urban Education*, 55(1), 95-124. <https://doi.org/10.1177/0042085916651323>
- Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47-63). New York: Holt, Rinehart & Winston.
- Bloom, H.S., Hill, C., Black, A.R., & Lipsey, M.W. (2008). *Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions*. MDRC. https://www.mdrc.org/sites/default/files/full_473.pdf.
- Bonne, L. (2016). New Zealand students' mathematics-related beliefs and attitudes: Recent evidence. *New Zealand Journal of Educational Studies*, 51(1), 69-82. <https://doi.org/10.1007/s40841-016-0035-2>
- Boston, M. D., Henrick, E. C., Gibbons, L. K, Berebitsky, D., & Colby, G. T. (2017). Investigating how to support principals as instructional leaders in mathematics. *Journal of Research on Leadership Education*, 12(3), 184-214. <https://doi.org/10.1177/1942775116640254>
- Braund, H., & DeLuca, C. (2018). Elementary students as active agents in their learning: An empirical study of the connections between assessment practices and student metacognition. *Australian Educational Researcher*, 45(1), 65-85. <https://doi.org/10.1007/s13384-018-0265-z>
- Brooks, C., Carroll, A., Gillies, R. M., & Hattie, J. (2019). A matrix of feedback for learning. *Australian Journal of Teacher Education*, 44(4), 14-32. <https://doi.org/10.14221/ajte.2018v44n4.2>

Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research, 3*, 22-30.

<https://doi.org/10.14786/flr.v2i1.24>

Center on Reinventing Public Education. (2022a). *Student achievement gaps and the pandemic: A new review of evidence from 2021-2022*. Center on Reinventing Public Education.

<https://crpe.org/student-achievement-gaps-and-the-pandemic-a-new-review-of-evidence-from-2021-2022/>

Center on Reinventing Public Education. (2022b). *The state of the American student: Fall 2022*. Center on Reinventing Public Education. <https://crpe.org/the-state-of-the-american-student/>

Cizek, G. J. (2012). *Setting performance standards: Foundations, methods, and innovations* (2nd ed.). Routledge.

Cleary, T. J., Velardi, B., & Schnaidman, B. (2017). Effects of the Self-Regulation Empowerment Program, (SREP) on middle school students' strategic skills, self-efficacy, and mathematics achievement.

Journal of School Psychology, 64, 28-42. <https://doi.org/10.1016/j.jsp.2017.04.004>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Conner, T. (2015). Relationships and authentic collaboration: Perceptions of a building leadership team.

Leadership and Research in Education: The Journal of the Ohio Council of Professors of Educational Administration (OCPEA), 2(1), 12-24. <https://files.eric.ed.gov/fulltext/EJ1088557.pdf>

Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Simmons, L. E., Kwok, O., Kim, M., & Rawlinson, D. M.

(2013). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children, 80*(1), 25-44.

<https://doi.org/10.1177/001440291308000101>

- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.
- Darling-Hammond, L., & Cook-Harvey, C. M. (2018). *Educating the whole child: Improving school climate to support student success*. Learning Policy Institute. <https://doi.org/10.54300/145.655>
- Davies, A. (2001). Involving students in communicating about their learning. *National Association of Secondary School Principals. NASSP Bulletin, 85*(621), 47-52.
<https://doi.org/10.1177/019263650108562106>
- Davis, J. R., & Warner, N. (2018). Schools matter: The positive relationship between New York City high schools' student academic progress and school climate. *Urban Education, 53*(8), 959-980.
<https://doi.org/10.1177/0042085915613544>
- Day, S. L., & Connor, C. M. (2017). Examining the relations between self-regulation and achievement in third-grade students. *Assessment for Effective Intervention, 42*(2), 97-109.
<https://doi.org/10.1177/1534508416670367>
- Day, S. L., Connor, C. M., & McClelland, M. M. (2015). Children's behavioral regulation and literacy: The impact of the first grade environment. *Journal of School Psychology, 53*, 409-428.
<https://doi.org/10.1016/j.jsp.2015.07.004>
- DeFlorio, L., Klein, A., Starkey, P., Swank, P. R., Taylor, H. B., Halliday, S. E., Beliakoff, A., & Mulcahy, C. (2019). A study of the developing relations between self-regulation and mathematical knowledge in the context of an early math intervention. *Early Childhood Research Quarterly, 46*, 33-48. <https://doi.org/10.1016/j.ecresq.2018.06.008>

Dessie, A. A., & Sewagegn, A. A. (2019). Moving beyond a sign of judgment: Primary school teachers' perception and practice of feedback. *International Journal of Instruction*, 12(2), 51- 66.

<https://doi.org/10.29333/iji.2019.1224a>

De Smedt, F., & Van Keer, H., (2018). Fostering writing in upper primary grades: A study into the distinct and combined impact of explicit instruction and peer assistance. *Reading and Writing*, 31(2), 325-354. <https://psycnet.apa.org/doi/10.1007/s11145-017-9787-4>

Edwards, A. R., & Beattie, R. L., (2016). Promoting student learning and productive persistence in developmental mathematics: Research frameworks informing the Carnegie Pathways. *NADE Digest*, 9(1), 30-39. <http://files.eric.ed.gov/fulltext/EJ1097458.pdf>

Elmore, R. F. (2008). *Improving the instructional core*. Achieve the Core.

https://achievethecore.org/content/upload/Improving%20The%20Instructional%20Core_Elmor e%20Article.pdf

Ewing, J., Gresham, G. J., & Dickey, B. (2019). Pre-service teachers learning to engage all students, including English language learners, in productive struggle. *Issues in the Undergraduate Mathematics Preparation of School Teachers: The Journal*, 2, 1-11.

<http://files.eric.ed.gov/fulltext/EJ1206251.pdf>

Fahle, E., Kane, T. J., Patterson, T., Reardon, S. F., & Staiger, D. O. (2022). *Local achievement impacts of the pandemic*. [https://educationrecoverycorecard.org/wp-](https://educationrecoverycorecard.org/wp-content/uploads/2022/10/Education-Recovery-Scorecard_Key-Findings_102822.pdf)

[content/uploads/2022/10/Education-Recovery-Scorecard_Key-Findings_102822.pdf](https://educationrecoverycorecard.org/wp-content/uploads/2022/10/Education-Recovery-Scorecard_Key-Findings_102822.pdf)

Fiske, S. T. (1993). Social cognition and social perception. *Annual Review of Psychology*, 44, 155–194.

<https://psycnet.apa.org/doi/10.1146/annurev.ps.44.020193.001103>

- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, 23, 1–74.
[https://doi.org/10.1016/S0065-2601\(08\)60317-2](https://doi.org/10.1016/S0065-2601(08)60317-2)
- Fluckiger, J., Tixier y Vigil, Y., Pasco, R., & Danielson, K. (2010). Formative feedback: Involving students as partners in assessment to enhance learning. *College Teaching*, 58, 136-140.
<https://doi.org/10.1080/87567555.2010.484031>
- Foley, E., Mishook, J., Thompson, J., Kubiak, M., Supovitz, J., & Rhude-Faust, M. K. (2008). *Beyond the test scores: Leading indicators for Education*. Annenberg Institute for School Reform at Brown University. <http://files.eric.ed.gov/fulltext/ED533117.pdf>
- Francisco, J. M. (2013). Learning in collaborative settings: Students building on each other's ideas to promote their mathematical understanding. *Educational Studies in Mathematics*, 82(3), 417-438. <https://doi.org/10.1007/s10649-012-9437-3>
- Freed, D., Sims, P., Tagaris, A., Hornberger, R., & Safer, A. (2021). International school principals' insights and experiences with teacher motivation. *International Journal of Educational Leadership Preparation*, 16(1), 60-73. <http://files.eric.ed.gov/fulltext/EJ1313056.pdf>
- Glaser, C., Kessler, C., Palm, D., & Brunstein, J. C. (2010). Improving fourth graders' self-regulated writing skills: Specialized and shared effects of process-oriented and outcome-related self-regulation procedures on students' task performance, strategy use, and self-evaluation. *Zeitschrift Für Pädagogische Psychologie*, 24(3-4), 177–190. <https://doi.org/10.1024/1010-0652/a000015>
- Graham, J. (2022). Explaining the racial school climate gap: Evidence from Georgia. *AERA Open*, 8(1), 1-19. <https://doi.org/10.1177/23328584221131>

- Green, T., & Allen, M. (2015). Professional development in urban schools: What do teachers say? *Journal of Inquiry & Action in Education*, 6(2), 53-79. <http://files.eric.ed.gov/fulltext/EJ1133585.pdf>
- Griffiths, M., & Davies, C. (1993). Learning to learn: Action research from an equal opportunities perspective in a junior school. *British Educational Research Journal*, 19(1), 43–58.
<https://doi.org/10.1080/0141192930190104>
- Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). *How principals affect students and schools: A systematic synthesis of two decades of research*. The Wallace Foundation.
<http://www.wallacefoundation.org/principalsynthesis>
- Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, 42(8), 433-444.
<https://doi.org/10.3102/0013189X13510020>
- Guskey, T. R. (1996). Reporting on student learning: Lessons from the past—Prescriptions for the future. In *Communicating student learning: 1996 Yearbook of the Association for Supervision and Curriculum Development*, T. R. Guskey (Ed.), (pp. 13–24). Association of Supervision and Curriculum and Development.
- Guskey, T. R. (1997). *Implementing mastery learning* (2nd ed.). Wadsworth.
- Guskey, T.R. (2001). Helping standards make the grade. *Educational Leadership*, 59(1): 20–27.
<https://www.ascd.org/el/articles/helping-standards-make-the-grade>
- Guskey, T. R. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom’s “Learning for Mastery.” *Journal of Advanced Academics*, 19(1), 8-31. <https://doi.org/10.4219/jaa-2007-704>

Hammerstein, S., König, C., Dreisöner, T., & Frey, A. (2021). Effects of COVID-19-related school closures on student achievement – A systematic review. *Frontiers in Psychology*, 12, Article 746289.

<https://doi.org/10.3389/fpsyg.2021.746289>

Han, L. (2020). On the relationship between teacher autonomy and learner autonomy. *International Education Studies*, 13(6), 153-162. <https://doi.org/10.5539/ies.v13n6p153>

Hanushek, E. A. (2022). *The economic cost of the pandemic: State by state*. Hoover Education Success Initiative.

<http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202022%20HESI%20EconomicCost.pdf>

Harrison, K., O'Hara, J., & McNamara, G. (2015). Re-thinking assessment: Self- and peer-assessment as drivers of self-direction in learning. *Eurasian Journal of Educational Research*, 60, 75-88.

<http://files.eric.ed.gov/fulltext/EJ1076698.pdf>

Hart, H., Young, C., Chen, A., Zou, A., & Allensworth, E.M. (2020). *Supporting school improvement: Early findings from reexamination of the 5Essentials survey*. University of Chicago Consortium on School Research. <https://consortium.uchicago.edu/sites/default/files/2020-08/Supporting%20School%20Improvement%205Essentials%20Survey%20ES-Aug2020-Consortium.pdf>

Hattie, J. (2023). *Visible learning, the sequel: a synthesis of over 2,100 meta-analyses relating to achievement*. Routledge.

Hattie, J., Gan, M., & Brooks, C. (2017). Instruction based on feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., pp. 290-324. Routledge.

Hattie, J., & Timperly, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.

<https://doi.org/10.3102/003465430298487>

Hattie, J., & Yates, G. C. R. (2014). *Visible learning and the science of how we learn*. Routledge.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Information Age.

Hinnant-Crawford, B., Faison, M., & Chang, M. (2018). Culture as mediator. Co-regulation, self-regulation, and middle school mathematics achievement. *Journal for Multicultural Education*, 10(3), 274-293. <https://doi.org/10.1108/JME-05-2016-0032>

Hirsh, A., & Segolsson, M. (2019). Enabling teacher-driven school-development and collaborative learning: An activity theory-based study of leadership as an overarching practice. *Educational Management Administration & Leadership*, 47(3), 400-420.

<https://doi.org/10.1177/1741143217739363>

Ho, A. D. (2008). The problem with “proficiency”: Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*, 37(6), 351-360. <https://doi.org/10.3102/0013189X08323842>

Illinois Compiled Statutes 105 ILCS 100-1046/2-3.153 (2018).

<https://www.ilga.gov/legislation/ilcs/ilcs4.asp?DocName=010500050HArt%2E+2&ActID=1005&ChapterID=17&SeqStart=14200000&SeqEnd=39000000>

Illinois State Board of Education. (n.d.). *5Essentials survey*. <https://www.isbe.net/Pages/5Essentials-Survey.aspx>

Illinois State Board of Education. (2023). *Illinois report card*. Retrieved March 23, 2023, from <https://www.illinoisreportcard.com/>

- Ingersoll, R. M., & Collins, G. J. (2017). Accountability and control in American schools. *Journal of Curriculum Studies*, 1, 75-95. <https://doi.org/10.1080/00220272.2016.1205142>
- Ishimaru, A. M., & Galloway, M. K. (2014). Beyond individual effectiveness: Conceptualizing organizational leadership for equity. *Leadership and Policy in Schools*, 13(1), 93-146. <https://doi.org/10.1080/15700763.2014.890733>
- Jamal, A., Tilchin, O., & Essawi, M. (2015). A teacher accountability model for overcoming self-exclusion of pupils. *International Education Studies*, 8(9), 58-64. <https://www.ccsenet.org/journal/index.php/ies/article/view/46288>
- Lenes, R., McClelland, M. M., ten Braak, D., Idsøe, T., & Størksen, I. (2020). Direct and indirect pathways from children' early self-regulation to academic achievement in fifth grade in Norway. *Early Childhood Research Quarterly*, 53, 612-624. <https://doi.org/10.1016/j.ecresq.2020.07.005>
- Kane, T., Doty E., Patterson, T., & Staiger, D., (2022) *What Do Changes in State Test Scores Imply for Later Life Outcomes?* Center for Education Policy Research, Harvard University. https://cepr.harvard.edu/files/cepr/files/long_term_outcomes.pdf?m=1666806941
- Karademir, C. A. & Akgul, A. (2019). Secondary school students' social studies-oriented academic risk-taking behaviours and autonomous learning skills. *Cypriot Journal of Educational Sciences*. 14(1), 056-068. <https://files.eric.ed.gov/fulltext/EJ1211747.pdf>
- Katzenbach, J. R., & Smith, D. K. (1993). The discipline of teams. *Harvard Business Review*, 71(2), 111-120.
- Kearney, E., Gebert, D., & Voelpel, S. C. (2009). When and how diversity benefits teams: The importance of team members' need for cognition. *The Academy of Management Journal*, 52(3), 581-596. <https://doi.org/10.5465/amj.2009.41331431>

Kingston, N., & Nash, B., (2011). Formative Assessment: A meta-analysis and a call for research.

Educational Measurement: Issues and Practice, 30(4), 28-37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>

Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). *Formative assessment and elementary school student academic achievement: A review of the evidence* (REL 2017–259). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. <http://ies.ed.gov/ncee/edlabs>.

Konold, T., Cornell, D., Jia, Y., & Malone, M. (2018). School climate, student engagement, and academic achievement: A latent variable, multilevel multi-informant examination. *AERA Open*, 4(4), 1-17. <https://doi.org/10.1177/2332858418815661>

Kosko, K. W. (2015). Geometry students' self-determination and their engagement in mathematical whole class discussion. *Investigations in Mathematics Learning*, 8(2), 17-36. <https://doi.org/10.1080/24727466.2015.11790349>

Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2013). The impact of accountability on teachers' assessments of student performance: A social cognitive analysis. *Social Psychology of Education: An International Journal*, 16, 215-239. <https://doi.org/10.1007/s11218-013-9215-9>

Kraft, M. A., & Falken, G. T. (2020). Why school climate matters for teachers and students. *State Education Standard*, 20(2), 33-35. https://nasbe.nyc3.digitaloceanspaces.com/2020/05/Kraft-Falken_May-2020-Standard.pdf

Kuhfeld, M., Soland, J., & Lewis, K. (2022). *Test Score Patterns Across Three COVID-19-impacted School Years*. (EdWorkingPaper: 22-521). Annenberg Institute at Brown University. <https://doi.org/10.26300/ga82-6v47>

Jaeger, R. M. (1989). Certification of student competence. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 485–514). American Council on Education.

LeClerc, M., & Moreau, A. C. (2011). Communautés d'apprentissage professionnelles dans huit écoles inclusives de l'Ontario. *Education et Francophonie*, 39(2), 189-206.
<https://doi.org/10.7202/1007734ar>

Lewis, K., & Kuhfeld, M. (2022). *Progress toward pandemic recovery: Continued signs of rebounding achievement at the start of the 2022-23 school year*. Center for School and Student Progress.
<https://www.nwea.org/research/publication/progress-towards-pandemic-recovery-continued-signs-of-rebounding-achievement-at-the-start-of-the-2022-23-school-year/>

Lewis, K., Kuhfeld, M., Langi, M., Peters, S., & Fahle, E. (2022). *The widening achievement divide during COVID-19*. Center for School and Student Progress.
<https://www.nwea.org/research/publication/the-widening-achievement-divide-during-covid-19/>

Livy, S., Muir, T., & Sullivan, P. (2018). Challenging tasks lead to productive struggle. *Australian Primary Mathematics Classroom*, 23(1), 19-24. <https://eprints.utas.edu.au/27642/2/125774%20-%20Challenging%20tasks%20lead%20to%20productive%20struggle.pdf>

Looney, J. W. (2011). *Integrating formative and summative assessment: Progress toward a seamless system?* OECD Education Working Papers, No. 58, OECD Publishing.
<https://doi.org/10.1787/5kghx3kbl734-en>

Lotan, R. (2003). Group-worthy tasks. *Educational Leadership*, 60(6), 72-75.
<https://www.ascd.org/el/articles/group-worthy-tasks>

- Marantika, J. E. R. (2021). Metacognitive ability and autonomous learning strategy in improving learning outcomes. *Journal of Education and Learning* 15(1), 88-96.
<https://doi.org/10.11591/edulearn.v15i1.17392>
- Marshik, T., Ashton, P. T., & Algina, J. (2017). Teachers' and students' needs for autonomy, competence, and relatedness as predictors of students' achievement. *Social Psychology of Education: An International Journal*, 20(1), 39-67. <https://doi.org/10.1007/s11218-016-9360-z>
- May, J. J., & Sanders, E. T. W. (2013). Beyond standardized test scores: An examination of leadership and climate as leading indicators of future success in the transformation of turnaround schools. *Journal of Urban Learning, Teaching, and Research*, 9, 42-54.
<http://files.eric.ed.gov/fulltext/EJ1028857.pdf>
- Maye, D. (2013). Hitting the mark: Strategic planning for academic rigor. *Delta Kappa Gamma Bulletin*, 79(4), 29-36.
- McCarley, T. A., Peters, M. L., & Decman, J. M. (2016). Transformational leadership related to school climate: A multi-level analysis. *Educational Management Administration & Leadership*, 44(2), 322-342. <https://doi.org/10.1177/1741143214549966>
- McClarty, K. L., Way, W. D., Porter, A. C., Beimers, J. N., & Miles, J. A. (2013). Evidence-Based Standard Setting: Establishing a Validity Framework for Cut Scores. *Educational Researcher*, 42(2), 78-88. <https://doi.org/10.3102/0013189X12470855>.
- Michaelsen, L. K., & Sweet, M. (2008). The essential elements of team-based learning. *New Directions for Teaching and Learning*, 116, 7-27. <https://doi.org/10.1002/tl.330>

- Munns, G., & Woodward, H. (2006). Student engagement and student self-assessment: The REAL Framework. *Assessment in Education: Principles, Policy and Practice*, 13(2), 193–213.
<https://doi.org/10.1080/09695940600703969>
- Muñoz, M. A., & Branham, K. E. (2016). Professional learning communities focusing on results and data-use to improve learning: The right implementation matters. *Planning and Changing*, 47(1/2), 37-46. <https://education.illinoisstate.edu/planning/articles/vol47.php>
- Murdoch, D., English, A. R., Hintz, A., & Tyson, K. (2020). Feeling heard: Inclusive education, transformative learning, and productive struggle. *Educational Theory*, 70(5), 653-679.
<https://doi.org/10.1111/edth.12449>
- Musso, M. F., Boekaerts, M., Segers, M., & Cascallar, E. C. (2019). Individual differences in basic cognitive processes and self-regulated learning: Their interaction effects on math performance. *Learning and Individual Differences*, 71, 58-70. <https://doi.org/10.1016/j.lindif.2019.03.003>
- National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring Educational Equity*. The National Academies Press. <https://doi.org/10.17226/25389>
- National Assessment of Educational Progress. (2022a, September 1). *NAEP long-term trend assessment results: Reading and mathematics*. <https://www.nationsreportcard.gov/highlights/ltr/2022/>
- National Assessment of Educational Progress. (2022b, October 24). *NAEP report card: Reading*. <https://www.nationsreportcard.gov/reading/nation/groups/?grade=4>
- National Assessment of Educational Progress. (2022c, October 24). *NAEP report card: Reading*. <https://www.nationsreportcard.gov/reading/nation/groups/?grade=8>
- National Assessment of Educational Progress. (2022d, October 24). *NAEP report card: Math*. <https://www.nationsreportcard.gov/mathematics/nation/groups/?grade=4>

National Assessment of Educational Progress. (2022e, October 24). *NAEP report card: Math*.

<https://www.nationsreportcard.gov/mathematics/nation/groups/?grade=8>

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests:*

Conflict or coexistence? Consortium on Chicago School Reform.

<https://consortium.uchicago.edu/sites/default/files/2018-10/p0a02.pdf>

Newmann, F. M., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student achievement. In F. M Newmann (Ed.), *Student engagement and achievement in American secondary schools*. Teachers College Press.

O'Brien, P. (2018). Self-determination for primary school children: Theory and practice. *REACH Journal of Special Needs Education in Ireland*, 31(2), 155-168.

<https://www.reachjournal.ie/index.php/reach/article/view/21>

Olina, Z., & Sullivan, H. J. (2002). Effects of classroom evaluation strategies on student achievement and attitudes. *Educational Technology Research and Development*, 50(3), 61–75.

<https://doi.org/10.1007/BF02505025>

Öngel, G., & Tabancalı, E. (2022). Teacher Enthusiasm and Collaborative School Climate. *Education Quarterly Reviews*, 5(2), 347-356. <https://doi.org/10.31014/aior.1993.05.02.494>

Osher, D., Neiman, S., & Williamson, S. (2020). School climate and measurement. *State Education Standard*, 20(2), 23-27. https://nasbe.nyc3.digitaloceanspaces.com/2020/05/Osher-Neiman-Williamson_May-2020-Standard.pdf

Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. The *Australian Educational Researcher*, 45(1), 13-31. <https://doi.org/10.1007/s13384-018-0258-y>

- Panadero, E., Brown, G. T. L., & Strijbos, J. (2015). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803-830.
<https://doi.org/10.1007/s10648-015-9350-2>
- Park, J., Lee, I. H., & Cooc, N. (2018). The role of school-level mechanisms: How principal support, professional learning communities, collective responsibility, and group-level teacher expectations affect student achievement. *Educational Administration Quarterly*, 0(0), 1-39.
<https://doi.org/10.1177/0013161X18821355>
- Pennant, J. (2018). *Group-worthy tasks and their potential to support children to develop independent problem-solving skills*. University of Cambridge NRICH. <https://nrich.maths.org/9935>
- Pierce, J. L., Kostova, T., & Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of General Psychology*, 7(1), 84-107.
<https://psycnet.apa.org/doi/10.1037/1089-2680.7.1.84>
- Pieschl, S. (2009). Metacognitive calibration – an extended conceptualization and potential applications. *Metacognition Learning*, 4, 3-31. <https://doi.org/10.1007/s11409-008-9030-4>
- Ratts, R. F., Pate, J. L., Archibald, J. G., Andrews, S. P., Ballard, C. C., & Lowney, K. S. (2015). The influence of professional learning communities on student achievement in elementary schools. *Journal of Education & Social Policy*, 2(4), 57-61
https://www.jespnet.com/journals/Vol_2_No_4_October_2015/5.pdf
- Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist*, 44(3), 159-175.
<https://doi.org/10.1080/00461520903028990>

Reeve, J., & Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational Psychologist, 56*(1), 54-77.

<https://doi.org/10.1080/00461520.2020.1862657>

Reeve, J., Cheon, S. H., & Yu, T. H. (2020). An autonomy-supportive intervention to develop students' resilience by boosting authentic engagement. *International Journal of Behavioral Development, 44*(4), 325-338. <https://doi.org/10.1177/0165025420911103>

Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology, 28*(1), 209-218.

<https://psycnet.apa.org/doi/10.1037/0022-0663.98.1.209>

Reeve, M., & Halusic, M. (2009). How K-12 teachers can put self-determination theory principles into practice. *Theory and Research in Education, 7*(2), 145-154.

<https://doi.org/10.1177/1477878509104319>

Reeve, J., & Su, Y. (2014). Teacher Motivation. In M. Gagné (Ed.), *The Oxford Handbook of Work Engagement, Motivation, and Self-Determination Theory* (pp. 349–362). Oxford University Press.

Roberge, M. E., & Van Dick, R. (2010, December). Recognizing the benefits of diversity: When and how does diversity increase group performance? *Human Resource Management Review, 20*(4), 295–

308. <https://doi.org/10.1016/j.hrmr.2009.09.002>

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52*(3), 475-514.

<https://doi.org/10.3102/0002831215585562>

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation, 11*(10), 1-13. <http://pareonline.net/getvn.asp?v=11&n=10>

- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, Article 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education*, 7(2), 224-233. <https://doi.org/10.1177%2F1477878509104327>
- Scardamalia, M., & Bereiter, C. (2022). Knowledge building and knowledge creation. In R. K. Sawyer (Ed.). *The Cambridge Handbook of the Learning Sciences* (3rd ed., pp. 385-405). Cambridge University Press. <https://doi.org/10.1017/9781108888295>
- Schaap, H., & De Bruijn, E. (2018). Elements affecting the development of professional learning communities in schools. *Learning Environments Research*, 21(1), 109-134. <https://doi.org/10.1007/s10984-017-9244-y>
- Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal*, 33(2), 359-382. <https://doi.org/10.3102/00028312033002359>
- Seglem, R. (2017). Creating a circle of learning: Teachers taking ownership through professional communities. *Voices from the Middle*, 25(1), 56-60. <https://library.ncte.org/journals/vm/issues/v16-4/7156>
- Shepard, L. A. (1979). Setting standards. In M. A. Bunda & J. R. Sanders (Eds.), *Practices and problems in competency-based measurement* (pp. 72-88). National Council on Measurement on Education.

- Shin, M., & Johnson, Z. D. (2021). From student-to-student confirmation to students' self-determination: An integrated peer-centered model of self-determination theory in the classroom. *Communication Education, 70*(4), 365-383. <https://doi.org/10.1080/03634523.2021.1912372>
- Shute, V. J. (2007). *Focus on formative feedback. Research Report*. Educational Testing Service. <https://files.eric.ed.gov/fulltext/EJ1111586.pdf>
- Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 92*(5), 761-798. <https://doi.org/10.3102/00346543211019105>
- Skibbe, L. E., Montroy, J. J., Bowles, R. P., & Morrison, F. J. (2019). Self-regulation and the development of literacy and language achievement from preschool through second grade. *Early Childhood Research Quarterly, 46*, 240-251. <https://doi.org/10.1016/j.ecresq.2018.02.005>
- Soenens, B., Sierens, E., Vansteenkiste, M., Dochy, F., & Goossens, L. (2012). Psychologically controlling teaching: Examining outcomes, antecedents, and mediators. *Journal of Educational Psychology, 104*(1), 108-120. <https://doi.org/10.1037/a0025742>
- Stahl, G. K., Maznevski, M. L., Voigt, A., & Jonsen, K. (2010). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies, 41*(4), 690-709. <https://doi.org/10.1057/s41267-020-00389-9>
- Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychologist, 39*(2), 97-110. https://doi.org/10.1207/s15326985ep3902_2
- Stein, R. E., Colyer, C. J., & Manning, J. (2016). Student accountability in team-based learning classes. *Teaching Sociology, 44*(1), 28-38. <https://doi.org/10.1177/0092055X15603429>

Stiggins, R. J. (2008). *Student-involved assessment for learning*. Merrill Prentice Hall.

Stipek, D. (1996). Motivation and instruction. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology*. Macmillan.

Suldo, S. M., Riley, K. N., & Shaffer, E. J. (2006). Academic Correlates of Children and Adolescents' Life Satisfaction. *School Psychology International*, 27(5), 567–582.

<https://doi.org/10.1177/0143034306073411>

Sun, M., Loeb, S., & Grissom, J. A. (2017). Building teacher teams: Evidence of positive spillovers from more effective colleagues. *Educational Evaluation and Policy Analysis*, 39(1), 104-125.

<https://doi.org/10.3102/0162373716665698>

Sun, Y., Zhang, J. & Scardamalia, M. (2010). Developing deep understanding and literacy while addressing a gender-based literacy gap. *Canadian Journal of Learning and Technology*, 36(1), 1-20. <https://doi.org/10.21432/T20P4D>

Supovitz, J. A., Foley, E., & Mishook, J. (2012). In search of leading indicators in education. *Education Policy Analysis Archives*, 20 (19), 1-27. <http://dx.doi.org/10.14507/epaa.v20n19.2012>

Swanson, E., McCulley, L. V., Osman, D. J., Lewis, N. S., & Solis, M. (2019). The effect of team-based learning on content knowledge: A meta-analysis. *Active Learning in Higher Education*, 20(1), 39-50. <https://doi.org/10.1177/1469787417731201>

Tichnor-Wagner, A., Harrison, C., & Cohen-Vogel, L. (2016). Cultures of learning in effective high schools. *Educational Administration Quarterly*, 52(4), 602-642.

<https://doi.org/10.1177/0013161X16644957>

Toth, M., & Sousa, D. (2019). *The power of student teams: Achieving the social, emotional, and cognitive learning in every classroom through academic teams*. Learning Sciences International.

- Valentine, K., & Bolyard, J. (2018). *Creating a classroom culture that supports productive struggle: Pre-service teachers' reflections on teaching mathematics*. Paper presented at the annual conference of the American Educational Research Association, New York, NY.
- Valls, R., & Kyriakides, L. (2013). The power of interactive groups: How diversity of adults volunteering in classroom groups can promote inclusion and success for children of vulnerable minority ethnic populations. *Cambridge Journal of Education*, 43(1), 17-33.
<https://doi.org/10.1080/0305764X.2012.749213>
- Van den Bergh, L., Ros, A., & Beijaard, D. (2014). Improving teacher feedback during active learning: Effects of a professional development program. *American Educational Research Journal*, 51(4), 772-809. <https://doi.org/10.3102/0002831214531322>
- VanLone, J., Freeman, J., LaSalle, T., Gordon, L., Polk, T., & Rocha Neves, J. (2019). A Practical Guide to Improving School Climate in High Schools. *Intervention in School and Clinic*, 55(1), 39-45.
<https://doi.org/10.1177/1053451219832988>
- Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, 15, 589-602. <https://doi.org/10.1016/j.learninstruc.2005.08.001>
- Warshauer, H. K. (2015). Strategies to support productive struggle. *Mathematics Teaching in the Middle School*, 20(7), 390-393. <https://doi.org/10.5951/mathteacmidscho.20.7.0390>
- Wehmeyer, M. L. (1997) Self-determination as an educational outcome. A definitional framework and implications for intervention. *Journal of Developmental and Physical Disabilities*, 9, 175-209.
<https://doi.org/10.1023/A:1024981820074>
- Wehmeyer, M. L., Sands, D. J., Knowlton, H. E., & Kozleski, E. B. (2002). *Teaching students with mental retardation: Providing access to the general curriculum*. Paul H. Brookes.

- Wehmeyer, M. L., Shogren, K. A., Toste, J. R., & Mahal, S. (2017). Self-determined learning to motivate struggling learners in reading and writing. *Intervention in School and Clinic, 52*(5), 295-303.
<https://doi.org/10.1177/1053451216676800>
- Wenjuan, G., Ling, L. K., & Jun, W. (2019). Teacher feedback and students' self-regulated learning in mathematics: A comparison between a high achieving and a low-achieving secondary school. *Studies in Educational Evaluation, 63*, 48-58. <https://doi.org/10.1016/j.stueduc.2019.07.001>
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). <https://ies.ed.gov/ncee/wwc/Handbooks>
- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment, 11*(3 & 4), 283-289. <https://doi.org/10.1080/10627197.2006.9652993>
- Wiliam, D, Lee, C.; Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles Policy and Practice, 11*(1), 49–65.
<https://doi.org/10.1080/0969594042000208994>
- Williams, K. E., White, S. L. J., & MacDonald, A. (2016). Early mathematics achievement of boys and girls: Do differences in early self-regulation pathways explain later achievement? *Learning and Individual Differences, 51*, 199-209. <https://doi.org/10.1016/j.lindif.2016.09.006>
- Zeybek, Z. (2016). Productive struggle in a geometry class. *International Journal of Research in Education and Science (IJRES), 2*(2), 396-415. <https://files.eric.ed.gov/fulltext/EJ1110272.pdf>
- Ziebell, N., & Clarke, D. (2018). Curriculum alignment: Performance types in the intended, enacted, and assessed curriculum in primary mathematics and science classrooms. *Studia Paedagogica, 23*(2), 175-203. <https://doi.org/10.5817/SP2018-2-10>

Zimmerman, B. J., & Kitsantas, A. (2014). Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement. *Contemporary Educational Psychology*, 39, 145-155. <https://doi.org/10.1016/j.cedpsych.2014.03.004>