# DATA-DRIVEN DECISION-MAKING LITERATURE SUPPORT SUMMARY

## I. Definition of the Element

Data-driven decision making (DDDM) refers to the systematic collection and use of data to guide policy and practice (Lockton et al., 2019; Mandinach, 2012; Pak & Desimone, 2019; Schelling & Rubenstein, 2021). Work in DDDM has been focused primarily on K-12 settings (Piety, 2019), but examples can be found at all levels of education (Mandinach, 2012). Data has entered education as a tool for accountability, but negative consequences of this approach have been recognized over time to point the way towards a more cooperative conception of data.

### Genesis of DDDM in Education (the Pitfalls of Using Data)

The push for educational reform was revived in the K-12 setting as a response to national panic about falling behind in educational achievement (Gardner, 1983). A major legislative response to this report came almost 20 years later—the U.S. Elementary and Secondary Education Act of 2001 ("No Child Left Behind" [NCLB]), thrusting data into the forefront of educational reform efforts (Pak & Desimone, 2019; Piety, 2019). Data was to be used as the main tool for accountability (Braaten et al., 2017; Lockton et al., 2019; Piety, 2019; Schelling & Rubenstein, 2021). While NCLB was replaced by the Every Student Succeeds Act (ESSA) in 2015, granting more flexibility to states regarding specific requirements, data-driven accountability continues to be relevant in education (Haslip & Gullo, 2018).

The underlying assumption was that teachers will do a bad job unless thoroughly monitored and controlled, a fairly old notion in education management circles (Callahan, 1962). The federal government was to do the controlling with the help from state education agencies, which the law provided with funding (Urban et al., 2019). The philosophy behind NCLB is that mimicking business enterprises' approach to management was going to solve the problem of US education falling behind its peers on the world stage (Au, 2011; Ryan, 2004). Unfortunately, the management science the law was inspired by was dated and did not age well (Au, 2011; Callahan, 1962; Rebitzer & Taylor, 2011; Taylor, 1919 ). Many of the precepts of those management practices have since been challenged (Benabou & Tirole, 2003; Benabou & Tirole, 2006; Rebitzer & Taylor, 2011).

It has been theorized (Holmstrom & Milgrom, 1991) and documented empirically (Braaten et al., 2017) that focus on assessment restricts the process of instruction to what can be tested (teaching to the test). Critical thinking and the ability to formulate hypotheses are aspects of learning that suffer as a result. There is reason to worry that the factory-like approach to education is coming for pre-K (Bassok et al., 2016; Zigler & Bishop-Josef, 2004). It may have made sense to train children to become factory robots in 1911; however, in 2021 there are actual robots to do that sort of work. Rather than tying educators' hands with never-ending summative assessments, modern labor economics research shows that giving them more autonomy and support would result in higher productivity (Ariely et al., 2009; Benabou & Tirole, 2003; Rebitzer & Taylor, 2011).

PARTNERSHIP FOR PRE-K IMPROVEMENT

## Goal of DDDM in Pre-K (the Right Way to use Data)

Still, something good came out of the discussion surrounding NCLB, Race to the Top, and ESSA. Data can be a force for good (Bryk et al., 2015; Lockton et al., 2019; Schelling & Rubenstein, 2021). However, rather than using it for accountability out of distrust for educators' hard work, choking educators' intrinsic motivation in the process, it should be used for continuous quality improvement (Bryk et al, 2015; Lockton et al., 2019; Schelling & Rubenstein, 2021). What's the difference? CQI assumes that educators want to improve, but they want to lead the process, utilizing their own experience, rather than being told what to do by a bureaucrat who never set foot in the classroom. Continuous quality improvement was born as an effort to empower teachers rather than control them. In CQI, data is not used to hold someone accountable but to provide support, promote network learning communities, and give teachers the tools and insights they need to improve on their teaching journey. Pre-K was spared this drive for accountability and has embraced the drive for improvement (Zigler & Styfco, 2010). Improvement is all about intrinsic motivation, and accountability is all about extrinsic motivation. Extrinsic motivation has been linked to decreases in job satisfaction and productivity. This is why rather than collecting summative assessment data in early learning, formative assessment data is collected instead (Firestone & Gonzalez, 2007; Little et al., 2019). The state teams should continue to avoid the pitfalls of data (drive for accountability) and leverage its strengths (drive for improvement). Still, while many pre-K programs are developing integrated data systems to house program quality, formative assessment, and other sorts of data, these are used more often for administrative/compliance purposes rather than DDDM, resulting in a failure to leverage them to their full potential because of quality or linkage issues (Little et al., 2019).

## The Price of Failure—Staff Reform Fatigue

Fixsen et al. (2015) note that innovations in the human service sector often elicit fear of an uncertain future and a tendency to protect status quo within the organizations in which they are attempted. Given the record of failure compounded by the failure to learn from that failure, this does not appear to be an example of irrational intransigence, but rather a rational fear of upsetting the status quo with an ill-conceived reform effort—a situation much more common than not (Braaten et al., 2017; Lockton et al., 2019; Pak & Desimone, 2019; Schelling & Rubenstein, 2021). This implies that planning well and following the guidelines of improvement and implementation science and using a tool such as the IDM to define success, monitor progress, and provide transparency to stakeholders is required to prevent the ECE community from losing the support and faith of its frontline workforce. Given how poorly paid the ECE workforce is, faith in the cause of ECE is a major currency. Every time we fail in reform, not because the failure was impossible to anticipate but because we did not put in place the infrastructure to learn from failure and gradually improve our practice, we chip away at the goodwill the ECE community and the public at large has for the project that started with Perry Preschool, Abecedarian, and Head Start. There is substantial goodwill for ECE; however, this goodwill should not be taken for granted.

## The How of DDDM—the RPP, Improvement Science, and Implementation Science

State teams have two options when it comes to DDDM—build an in-house research team or contract with outside researchers.

DDDM's chief challenge is that it requires analysts and decision-makers to work together. Analysts know how to answer questions decision-makers may be interested in, but do not always know what are the most urgent questions facing decision-makers operating state pre-K programs. Decision-makers, for their part, do not always know which questions could be answered by data and which cannot. Finally, if the infrastructure is not in place by the time decision-makers are asking a question, an answer using data cannot be produced on time—

advance planning for infrastructure is essential. Data must exist, and data collection processes need to be in place. The right data needs to be collected in the right way at the right frequency. This capacity is rarely available to state early learning agencies. Creating a research-practice partnership where decision-makers and researchers work together has been proposed as a solution (Coburn & Penuel, 2016; Henrick et al., 2017). Researchers are interested in having access to administrative data collected by early learning agencies, while early learning agencies are sitting atop data they do not know what to do with, because they lack the specialized personnel. Hence, this matching has the potential to provide major benefits to both parties. For such a partnership to work well, both sides need to understand each other's needs. Researchers are interested in publishing original research, while early learning agencies are interested in improving practice. Both are interested in strong research designs, because the stronger the research design, the more the insight gained from data actually applies to the original question posed. Researchers care about originality, while practitioners do not. Still, arrangements can be made whereby researchers are allowed to publish some of the analysis in return for providing insights locked in the data.

If the state teams decide to navigate the challenges of DDDM on their own, it is recommended that they build up a capacity for improvement and implementation science practices within their agency. Improvement science tells us how to tinker around the edges of an existing system in a productive way, while implementation science tells us how to scale a complete system once it has been improved to satisfaction through the aforementioned improvement science tinkering (more information on implementation and improvement science can be found in Appendix B).

The DDDM Element focuses on infrastructure indicators. Equitable infrastructure indicators focus on state systems, policies, and practices that support high-quality pre-K. The infrastructure indicators are labeled as policy (e.g., established in policy and statewide standards), supports (e.g., dedicated resources), and data (e.g., data collection standards and protocols and data use). We have conceptualized the findings in the literature into the following DDDM indicators focused on infrastructure:

### Indicator 1: Program Quality Assessment

State requires programs to conduct program level assessments using reliable and valid measurements to inform program level continuous quality improvement (CQI). Assessments include evaluating the quality of the following six conditions:

- Supports for dual language learners and inclusion and individualizing for children with developmental delays and disabilities
- Classroom environments
- Teacher-child interactions
- Curriculum implementation
- Family engagement practices
- Child outcomes (e.g., kindergarten readiness)

PARTNERSHIP FOR PRE-K IMPROVEMENT

**Indicator 2: Data-Driven Decision-Making Implementation**

State engages in DDDM to ensure high quality teaching, equitable access for children and families, equitable PD for early childhood educators, and positive child outcomes. The state's DDDM efforts include these six conditions:

- Supporting programs to set annual (or more frequent) goals towards improving teaching and learning, equitable access, and child outcomes
- Monitoring programs' progress towards those goals by collecting multiple types of data including student data (e.g., enrollment, attendance, assessments), classroom observations of teaching quality, and early childhood educators, leaders, coaches, and family surveys
- Disaggregating and analyzing data by targeted populations
- Using data for improvement of policies, and supports (e.g., funding, PD, training etc.) to programs
- Supporting program leaders, early childhood educators, and other stakeholders to analyze their own data and create or modify their professional learning goals and action plans
- Improving data collection, and data analysis processes

**Indicator 3: QRIS**

State has a standardized quality rating and improvement system (QRIS) to assess program quality, or the system meets the following four conditions:

- System includes on-site program quality assessments at least once every two or three years.
- State system is differentiated so that programs rated lower in quality or with previous policy violations receive more frequent on-site program quality assessment visits.
- On-site visits include classroom observations by trained and reliable observers. Observers use research-based, valid, and reliable tools to measure quality. Observations include a focus on teacher-child interactions and instructional quality.
- The QRIS is inclusive and aligned across multiple early learning systems, including state pre-K, private, and other early learning programs (e.g., family childcare programs, childcare centers). Private programs must be rated at a high level to have state pre-K classrooms or slots.

**Indicator 4: Access to Multiple Data**

State ensures access to various kinds of data on all six of the following areas:

- Learning and development assessment data, student attendance data, including information on suspensions/expulsions
- Data on the qualifications and diversity of the ECE workforce
- Data on professional development for ECE providers, including job-embedded professional learning (JEPL) data
- Classroom quality data

- Curriculum fidelity data
- Data on family engagement efforts and early childhood educators interaction and collaboration with parents

**Indicator 5: Access to High-Quality Data**
State has formal processes for determining the relevancy and quality (i.e., reliable and valid) of data collected at the student or classroom level. These processes have been applied to all of the state's current data and data are being aggregated to the state level to use for CQI.

**Indicator 6: Data Linkages**
To make informed decisions, state has the infrastructure and data analytic capacity to connect different types of data to capture a full picture of the pre-K system and meets the following three conditions:

- Links student data to specific classrooms
- System can put program data in the context of community data (e.g., demographics, family characteristics, and health)
- System connects professional learning data with teaching quality and child assessment data

**Indicator 7: Central Data Management System**
State has a centralized data aggregation, linking and management system. Data management system meets the following four conditions:

- State collects data at all appropriate levels including classroom, program, district, and state level.
- State can link information across programs to account for all children served across various funding streams (e.g., childcare subsidy, HeadStart, Section 619 - IDEA).
- State data system collects specific demographic data including race, ethnicity, geography, socio-economic status, DLL status, and special needs status.
- State system collects and tracks longitudinal data on students to determine efficacy of pre-K efforts and collaborates with the K-12 system to ensure common usable data.

**Indicator 8: Data Use**
State conducts regular analysis and reporting for data collected and the system meets the following four conditions:

- Student data are analyzed by critical subgroups (e.g., race, ethnicity, income, DLL, and children with developmental delays and disabilities).
- System includes the analysis of trends in the data over time and relationships between key variables. Data are used by leaders to inform decision making about policies, funding, and other supports.
- Collaboration with key stakeholders to interpret data, identify key issues, and gain input on plans for improvement.

- Identification of districts, programs, or schools that have improved and processes for others to learn from their success.

### Indicator 9: DDDM to Improve Equity

State teams have access to data on all populations (children, early childhood educators, and parents) that are part of the pre-K system. State keeps track of equity differences in quality and achievement across identified subgroups and takes steps to eliminate those differences by disaggregating data in meaningful ways, such as by race, income, language or other important traits that historically predict inequalities in outcome.

The state's efforts to understand and address inequity include these five components:

- Ongoing data collection
- Engaging in active discussions that surface issues of inequity for targeted populations
- Action planning and implementation - creating action plans and following through with implementation
- Planning to assess and refine implementation
- Amending policies and practices that address these issues

# II. DDDM Literature Process Overview and Summary

To understand the existing literature support and identify the literature gaps and limitations for each of the IDM indicators, we conducted a systematic literature search and checked with experts for relevant sources to support the various indicators of DDDM. More details of the general review process conducted across all elements can be found in the [IDM Evidence Review Document](). For the DDDM Element, 11 key phrases were identified and explored. Out of these initial phrases, nine key phrases retrieved relevant results. The list of all sources that yielded relevant results based on the nine key phrases and expert recommendations, along with two key phrases that did not yield relevant results, can be found in Appendix A.

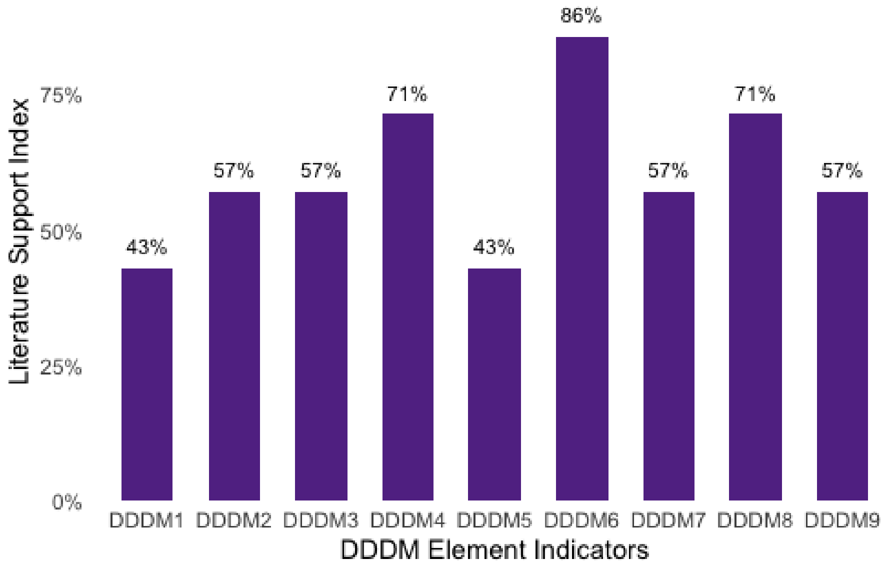
Once the literature search for the DDDM Element was completed, we reviewed the quantity and rigor of the literature supporting each indicator and computed what we termed the Literature Support Index (LSI). The LSI is calculated for each indicator based on seven components:

1. 1. at least three peer-reviewed articles;
2. 2. at least one study with no more than two limitations;
3. 3. at least one study at national or state level;
4. 4. at least one study that uses experimental or quasi-experimental design;
5. 5. at least two studies that use representative sampling;
6. 6. support from at least one national research organization; and
7. 7. support from at least one national policy organization.

The LSI is expressed as a percentage of the above seven criteria that are satisfied for a particular indicator. For more information about the rationale for the LSI and how it is calculated can be found in the IDM Evidence Review Document Figure 1 summarizes the LSI for the DDDM Element indicators.

**Figure 1**

*Overall Summary of DDDM Literature Support Index*

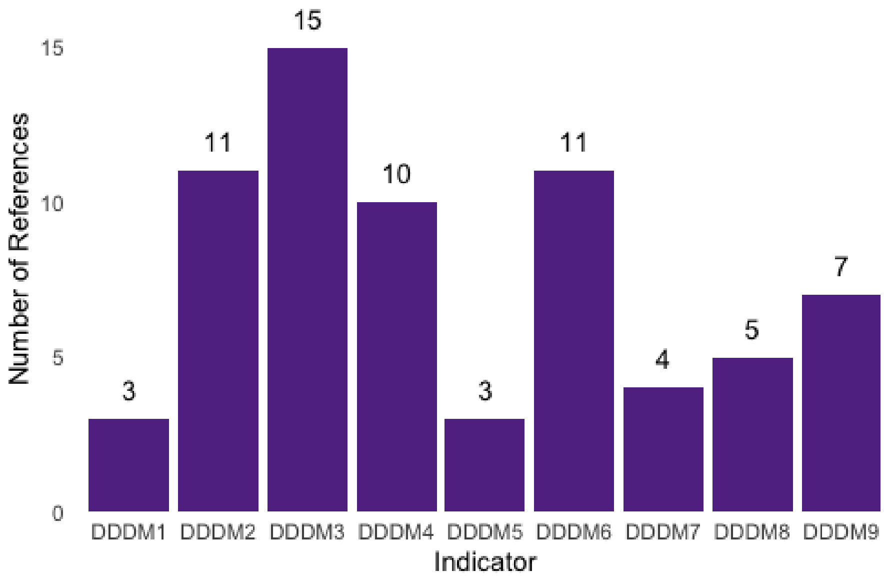



Below, we present figures describing the rigor and quantity of literature supporting each indicator, the publication types and research designs, and the child outcomes examined. In subsequent sections, we delve into each indicator in detail.

While Figure 1 combines aspects of both rigor of the literature as well as quantity supporting each indicator, Figure 2 represents solely the quantity of evidence for each indicator. Figure 2 shows that DDDM indicators 2, 3, 4, and 6 are supported by a larger number of sources than the rest of the indicators. We hope that this type of analysis can help state teams understand where there are gaps in research and potential directions for future studies (DDDM1 and DDDM5 are under-researched topics for example, hence ripe for the state teams' own research-practice partnership efforts).

**Figure 2**

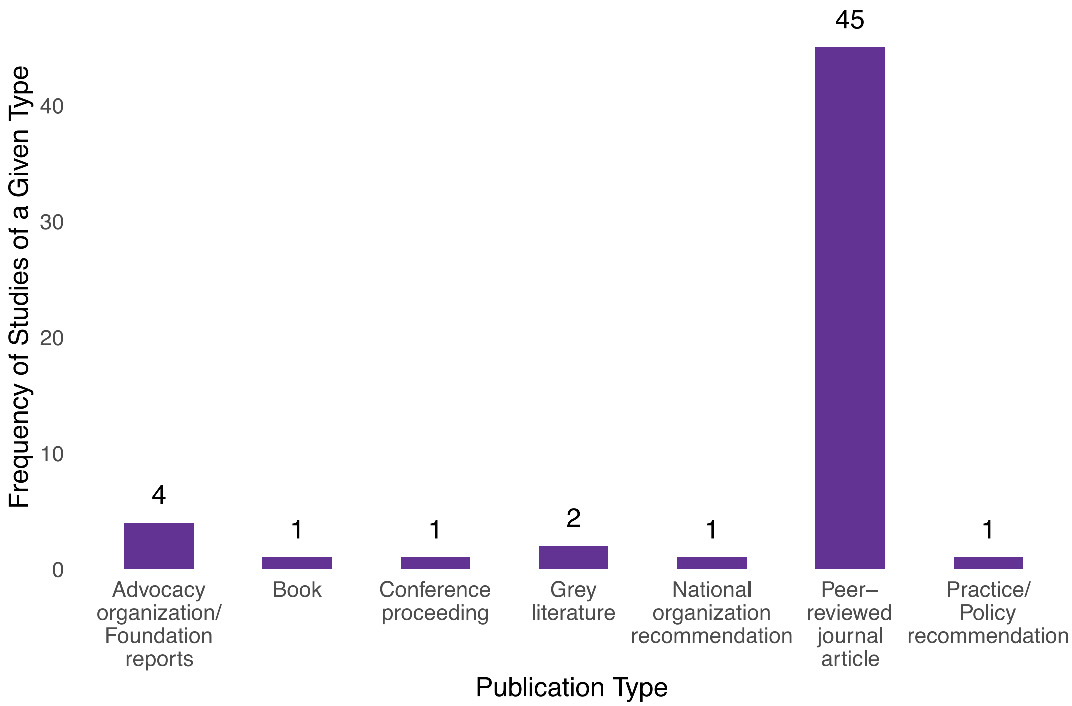
*DDDM Quantity of Evidence by Indicator*



To understand more about the nature of the literature that supports the Element, Figure 3 lists the types of publications used as evidence for it. The majority of the sources are articles from peer reviewed journals (45). These are the sources with the highest quality, conducted by researchers and reviewed by other researchers.

**Figure 3**

*DDDM Evidence by Publication Type*



In addition to types of publications, Figure 4 summarizes the research design used in the sources supporting the DDDM Element. The most common type of research designs represented in the DDDM literature scan (12) involved a pre-post study, a very common research design in the education research literature. Experimental studies and literature reviews were the second most common research designs (10).

**Figure 4**

*DDDM Summary of Research Design*

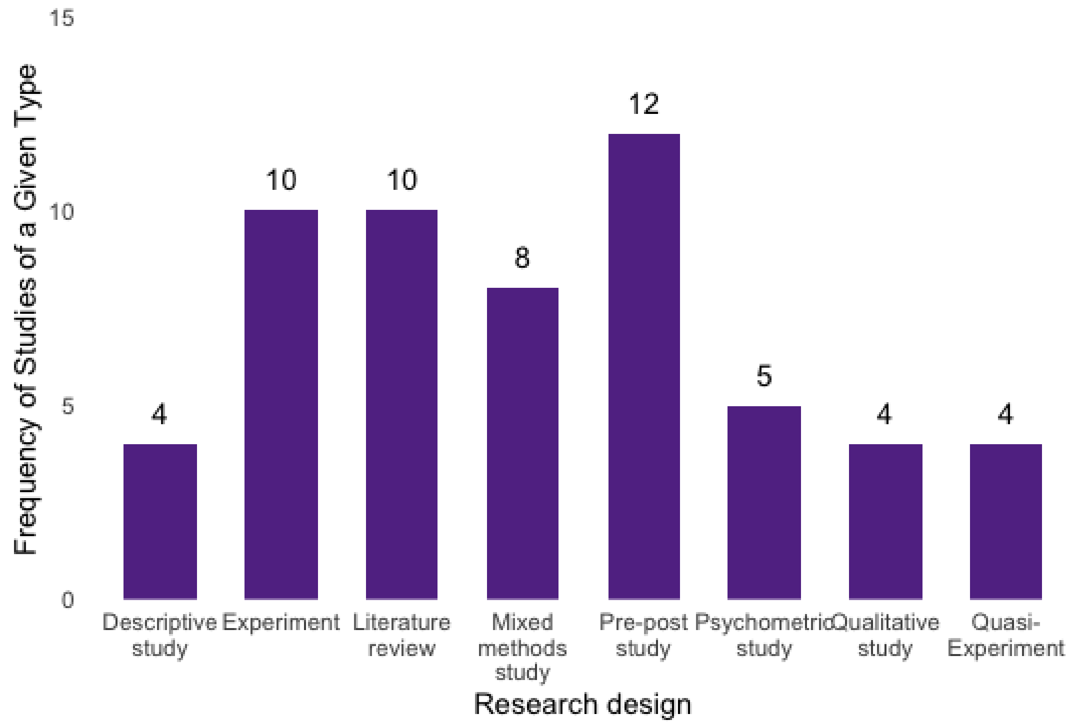


**Figure 5**

*DDDM Child Outcomes Studies Examined*

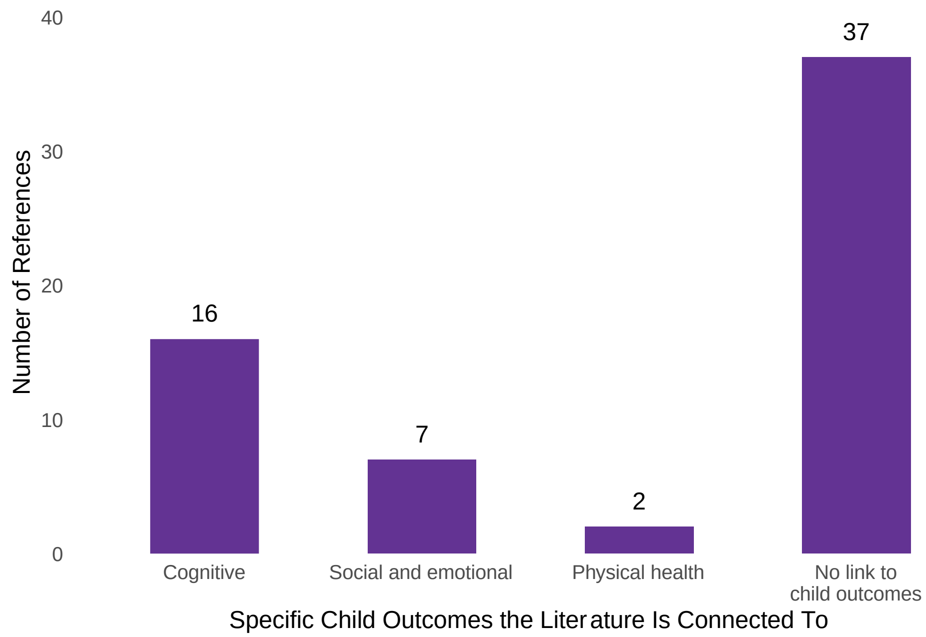



Figure 5 shows that most studies in the DDDM Element were not connected directly to child outcomes. Linking DDDM issues to child outcomes presents quite the logistical challenges in terms of data collection, linkage, scale, etc., and these numbers reflect those challenges. To the extent that studies were able to overcome these challenges, they focused on cognitive (16) and social emotional (7) outcomes.

# III. Summary of DDDM Literature Supporting Indicators: Current Practices and Challenges

**Data-Driven Decision-Making**
**Infrastructure Indicators (state level)**
**Data Collection Requirements**

**Indicator 1: Program Quality Assessment**

State requires programs to conduct program level assessments using reliable and valid measurements to inform program level continuous quality improvement (CQI). Assessments include evaluating the quality of the following six conditions:

- Supports for dual language learners and inclusion and individualizing for children with developmental delays and disabilities
- Classroom environments
- Teacher-child interactions
- Curriculum implementation
- Family engagement practices
- Child outcomes (e.g. kindergarten readiness)

**Indicator 2: Data-Driven Decision-Making Implementation**

State engages in DDDM to ensure high quality teaching, equitable access for children and families, equitable PD for early childhood educators, and positive child outcomes. The state's DDDM efforts include these six conditions:

- Supporting programs to set annual (or more frequent) goals towards improving teaching and learning, equitable access, and child outcomes
- Monitoring programs' progress towards those goals by collecting multiple types of data including student data (e.g., enrollment, attendance, assessments), classroom observations of teaching quality, and early childhood educators, leaders, coaches, and family surveys
- Disaggregating and analyzing data by targeted populations
- Using data for improvement of policies, and supports (e.g. funding, PD, training etc.) to programs
- Supporting program leaders, early childhood educators, and other stakeholders to analyze their own data and create or modify their professional learning goals and action plans
- Improving data collection, and data analysis processes

**Data-Driven Decision-Making**
**Infrastructure Indicators (state level)**
**Data Collection Requirements**

**Indicator 3: QRIS**

State has a standardized quality rating and improvement system (QRIS) to assess program quality, or the system meets the following four conditions:

- System includes on-site program quality assessments at least once every two or three years
- State system is differentiated so that programs rated lower in quality or with previous policy violations receive more frequent on-site program quality assessment visits.
- On-site visits include classroom observations by trained and reliable observers. Observers use research-based, valid, and reliable tools to measure quality. Observations include a focus on teacher-child interactions and instructional quality.

**The QRIS is inclusive and aligned across multiple early learning systems, including state pre-K, private, and other early learning programs (e.g., family childcare programs, childcare centers). Private programs must be rated at a high level to have state pre-K classrooms or slots.**

Program quality assessments are often used for continuous quality improvement. Bassok and Galdo (2016) highlight the increased importance of understanding overall program quality and which children have access to high-quality programs. Measures of quality are generally categorized as either measuring structural quality or the quality of the classroom process. Structural measures include staff–child ratios, class size, full- versus half-day operation, and teacher credentials. Process quality, or how learning is happening, focuses on teacher–child interactions (Bassok & Galdo, 2016). In 2019, of 62 state-funded preschool programs surveyed (in 44 states and Washington, DC), 39 used data from CQI for program improvement (Friedman-Krauss et al., 2020).

One example of a program quality assessment instrument is the widely used Classroom Assessment Scoring System© (CLASS). CLASS was developed to rate the quality of classroom interactional processes, namely how teachers are supporting children's academic and social development (Hamre et al., 2009). The three domains of interaction assessed by CLASS are Emotional Support, Classroom Organization, and Instructional Support. Each of these three includes several dimensions (Hamre et al., 2009). CLASS is used in every Head Start program, and it is a required or optional component of many state QRISs (Bassok & Galdo, 2016). Other common instruments for program evaluation in ECE include: Early Childhood Environmental Rating Scale Revised Edition (ECERS-R), Early Childhood Classroom Observation Measure (ECCOM), Early Language and Literacy Classroom Observation Tool (ELLCO), Infant Toddler Environment Rating Scale Revised Edition (ITERS-R), School Age Care Environment Rating Scale (SACERS), Supports for Early Literacy Assessment (SELA), and National Institute for Early Education Research (NIEER; Slentz, 2008).

CQI (continuous quality improvement) is the iterative process of using data to create conditions in which there can be continuous changes over time to improve and grow programs in a way that is intentional and systematic (BUILD Initiative, 2017; Tout et al., 2015). CQI has been gaining traction in the field of QRIS because of the focus on using data to make decisions about quality improvement priorities, and on developing the capacity for ongoing assessment and improvements (Tout et al., 2015). For example, Hallam et al. (2019) demonstrates the importance of a QRIS by examining the quality of family child care programs in Delaware that participated in a QRIS. The authors compare family child care providers who participated in a project providing supplemental quality improvement supports (Stars Plus) with family child care providers who participated in QRIS but did not receive the supplemental supports. Results from this study show that within their sample of providers, those who received Stars Plus were 1.8 times more likely to improve their program quality and move up a star level, compared to providers who participated in the QRIS but did not receive Stars Plus. In conclusion, Hallam et al. (2019) reinforces the importance of state QRIS by demonstrating significant program quality improvement for providers who participated in a tailored QRIS.

The use of CQI to promote a culture of using data for goal-setting and improvements is a newer part of QRIS (Elicker & Ruprecht, 2019). While data may be a valuable tool to inform early childhood policy, it must be collected and used correctly. School leaders play a valuable role in facilitating strong "cultures of data use" (Cohen-Vogel & Harrison, 2013; Little et al., 2019). Existing research found that schools with strong cultures of data use have leaders who work to promote an atmosphere of learning and emphasize continuous improvement (Cohen-Vogel & Harrison, 2013; Firestone & Gonzalez, 2007; Little et al., 2019; Stein et al., 2013). To support cultures of data use, leaders should assess teachers' capacity for data use and plan how to build supporting structures for it (Schelling & Rubenstein, 2021).

Beyond QRIS, pre-K settings are data-rich environments that often have informal data collection through developmental screening tools and formative assessments systems (Little et al., 2019). The authors also found that engagement with and use of these data for instructional purposes varies across programs and educators, and recommend that studies of program quality are necessary if states want to achieve the goal of all students having equitable access to quality programs.

Research indicates that clear, specific goals and a clearly articulated theory of change are important determinants for whether a quality improvement initiative will succeed (Tout et al., 2015). Boller et al. (2014) catalogues intervention approaches to quality improvement (QI), emphasizing accountability systems and commitment of resources for educating the field and parents about program quality data. Lahti et al. (2015) provides a framework for maintaining high-quality CQI standards and processes, using two recent QRIS validation studies as examples to illustrate how actual QRIS systems measure up to this framework. Zellman and Fiene (2012) introduce their own framework, using QRIS systems in Indiana and Maine as examples. Their framework consists of four approaches, which include relating what ratings mean for children's development. Although there has been limited research on CQI in ECE, there is agreement in the field that it is crucial (Tout et al., 2015).

Program quality assessments are often used for continuous quality improvements. As mentioned previously under DDDM Indicator 1, it is increasingly important to the ECE field to understand overall program quality and which children have access to high-quality programs (Bassok & Galdo, 2016). As outlined above, measures of quality are generally categorized as either measuring structural quality or the quality of the classroom process.

An emphasis in ECE policy on quality, especially process quality, has led to the growth of Quality Rating and Improvement Systems (QRIS) in many states (Connors & Morris, 2015; Soderberg et al., 2016). QRISs often endeavor to both raise the quality of programs in their states and to make information on the quality of different programs available to families (Connors & Morris, 2015; Soderberg et al., 2016). Programs are rated on QRIS by a set of factors. This generally includes structural features, classroom observations, and other measures, such as use of curriculum and engagement with families (Vitiello et al., 2018). Because QRIS is fairly recent, it has not been implemented at scale in many states (Vitiello et al., 2018). Currently, QRIS vary widely across states where they have been implemented. Tout et al. (2015) argue that while QRIS should be standardized, it must also be flexible and adaptable to individual program needs. In order to standardize QRIS, stable funding is needed (Tout et al., 2015). Connors and Morris (2015) note that variation in state QRIS does offer the opportunity to compare different approaches to quality improvement.

Evaluating quality improvement in QRIS is a fairly recent undertaking. Limitations of the research literature on quality improvement include study designs and methods that are not conducive to looking at one specific part of a quality improvement initiative (Tout et al., 2015). While some research indicates that QRIS can have a meaningful impact on program quality, as described below, Vitiello and colleagues (2018) emphasize that this only happens if QRIS is consistent with our knowledge of child development.

Jeon and Buettner (2015) demonstrate the importance of QRIS through its impact on child outcomes. The authors examined the associations between QRIS, a statewide government-funded early childhood care and education policy incorporating structural quality of child care programs, and children's cognitive skills. The study found that children in the highest level of QRIS ranking demonstrated better cognitive skills, consisting of receptive vocabulary, phonological awareness, and mathematical skills. Additionally, participating in the QRIS moderated a negative association between family socioeconomic risk and children's cognitive skills. These results indicate that policymakers may expect positive returns on QRIS investments with regards to children's early cognitive development (Jeon & Buettner, 2015). As another example, Sabol and Pianta (2015) indicate associations between QRIS and child outcomes in their study focused on Virginia's QRIS. Specifically, children in programs with higher ratings had more growth in literacy skills during their preschool year compared to children in lower-rated pre-K programs. This study sheds light on the potential of QRIS to identify programs falling behind on quality. Programs so identified can then be further probed to understand the reasons behind the quality deficit and offer additional support, PD, or training. On the other hand, results from Hong et al. (2015) were inconclusive regarding a link between QRIS and child outcomes.

Looking to the future of using statewide QRIS to evaluate impact, Goodvin and Hansen (2019) describe the planned evaluation of the impact of Early Achievers, Washington state's QRIS program, by the Washington State Institute for Public Policy (WSIPP). According to the authors, this evaluation is positioned to be the most rigorous study to date of the impact of QRIS on child outcomes (Goodvin & Hansen, 2019).

| Data Driven Decision Making<br>Infrastructure Indicators (state level)<br>Access to Data (multiple and high-quality data) |
|---|
| **Indicator 4: Access to Multiple Data**<br>State ensures access to various kinds of data on all six of the following areas:<br>• Learning and development assessment data, student attendance data, including information on suspensions/expulsions<br>• Data on the qualifications and diversity of the ECE workforce<br>• Data on professional development for ECE providers, including job-embedded professional learning (JEPL) data<br>• Classroom quality data<br>• Curriculum fidelity data<br>• Data on family engagement efforts and early childhood educators interaction and collaboration with parents<br>**Indicator 5: Access to High-Quality Data**<br>State has formal processes for determining the relevancy and quality (i.e., reliable and valid) of data collected at the student or classroom level. These processes have been applied to all of the state's current data and data are being aggregated to the state level to use for CQI. |

Access to multiple data helps paint a more comprehensive picture of the workings of a pre-K program (Landry, 2009; Stein et al., 2013). Stein et al. (2013) collected “teacher surveys that asked for ratings of a child’s readiness at grade entry and exit as well as level of parent participation, grades, type of school attended, and standardized assessments” (p. 25) (quantitative data). Qualitative data in the form of parent interviews and focus groups were collected. By cross-referencing all these different datasets, they were able to identify key challenges families faced in their transitions. Educare was able to paint a more comprehensive picture of the challenges families encounter in their transitions from pre-K to kindergarten, form hypotheses about their causes, incorporate a response into their programming, and by continuing to collect data, also examine the efficacy of this response.

The early learning field does not have a strong culture of data quality (Zweig et al., 2015). The push for a strong culture of data quality in education comes mostly from the federal government and private philanthropic organizations (National Center on Program Management and Fiscal Operations [NCPMFO], 2020; Urban et al., 2019; Zigler & Styfco, 2010). Hence, the only standards of high-quality data have been defined by the federal government in its push to implement the No Child Left Behind Act, the Race to the Top Act, the Every Student Succeeds Act, and the Improving Head Start for School Readiness Act of 2007 legislation (NCPMFO, 2020; Urban et al., 2019; U.S. Department of Education [US DOE], 2006). The DOE and DHHS standards are in line with the high-quality data concepts considered in other fields, such as medicine (Richesson & Krischer, 2007).

Research-practice partnerships (RPPs) offer a way for states to obtain access to researchers who are knowledgeable about aspects of high-quality data without having to hire their own teams (Henrick et al., 2017). Interviewing members of three types of RPPs, including research alliances, design research partnerships, and networked improvement communities (NICs), Henrick et al. (2017) suggest that high-quality data comes from research that is methodologically thorough while also being relevant and applicable for both practitioners and policymakers. The Office of Head Start criteria for high-quality data (i.e., relevance, timeliness, accuracy, completeness, validity, and reliability) is also recommended for this DDDM indicator. The report notes that RPP research should be practical and comprehensive to include multiple perspectives from researchers, practitioners, and stakeholders.

### Examples of Benefits of High-Quality Data

A great example of high-quality data being put to work in service to families is a study by Buzhardt et al. (2011) involving the use of a web-based support system called MOD (Making Online Decisions) within a home-visit intervention. The MOD is a clinical decision-making support system (CDSS) for home visitors, designed to support evidence-based practices. Similar to CDSSs successfully employed in the medical field, the MOD was integrated with an online data system and the activities of home visitors. The authors put forth that the MOD can be an answer to a number of challenges in utilizing DDDM in early childhood settings, and provides a system to bring together different elements of evidence-based practice to guide decision making. This study checks multiple of the above criteria regarding high-quality data.

Setodji et al. (2018) is another excellent example of the kind of insights that high-quality data can reveal. Their findings suggest that while there has been a growing trend of states using ECERS–R to make important decisions, ECERS-R may only reliably show associations with children's outcomes within certain score ranges or subscales. Their work would not have succeeded if it were not for their using a valid and reliable tool (ECERS-R).

| Data Driven Decision Making<br>Infrastructure Indicators (state level)<br>Data Management |
|---|
| **Indicator 6: Data Linkages**<br>To make informed decisions, state has the infrastructure and data analytic capacity to connect different types of data to capture a full picture of the pre-K system and meets the following three conditions:<br>• Links student data to specific classrooms<br>• System can put program data in the context of community data (e.g., demographics, family characteristics, and health)<br>• System connects professional learning data with teaching quality and child assessment data<br>**Indicator 7: Central Data Management System**<br>State has a centralized data aggregation, linking and management system. Data management system meets the following four conditions: |

| **Data Driven Decision Making**<br>**Infrastructure Indicators (state level)**<br>**Data Management** |
|---|
| • State collects data at all appropriate levels including classroom, program, district, and state level.<br>• State can link information across programs to account for all children served across various funding streams (e.g., childcare subsidy, HeadStart, Section 619 - IDEA).<br>• State data system collects specific demographic data including race, ethnicity, geography, socio-economic status, DLL status, and special needs status.<br>• State system collects and tracks longitudinal data on students to determine efficacy of pre-K efforts and collaborates with K-12 system to ensure common usable data.<br>**Indicator 8: Data Use**<br>State conducts regular analysis and reporting for data collected and the system meets the following four conditions:<br>• Student data are analyzed by critical subgroups (e.g., race, ethnicity, income, DLL, and children with developmental delays and disabilities).<br>• System includes the analysis of trends in the data over time and relationships between key variables. Data are used by leaders to inform decision making about policies, funding, and other supports.<br>• Collaboration with key stakeholders to interpret data, identify key issues, and gain input on plans for improvement.<br>• Identification of districts, programs, or schools that have improved and processes for others to learn from their success. |

According to King et al. (2018), at the time of publication, only 22 states currently linked child-level data to a comprehensive understanding of early childhood education in their state.

Regarding data linkages, another study (King et al., 2018) showed 22 states link child-level data with program-level data and focus on a comprehensive understanding of early learning.

In contrast to the child-level data, states were less likely to link workforce-level data with program-level data in the 2018 report (King et al., 2018).

There is a focus in the literature around formal processes for each state to collect data.

For the organization of data collecting, there were fewer defined data governance bodies conducting early childhood education related data work in 2018 than in 2013, which decreased from 32 to 22. Furthermore, in 2018, fewer states had a data governance body as compared with 2013 (King et al., 2018).

Integrated Data Systems (IDS) are promising mechanisms for analyzing complex patterns of information across systems (Limlingan et al., 2015). Already, many IDS have been utilized in research that requires linked data. In one example from the K-12 level, researchers working with Pittsburgh Public Schools used an IDS to find a link between a number of students who were struggling academically and use of social service providers, such as child welfare and homeless services.

State-level decision making needs to be supported by state-level data collection and analysis. The needs for state-level data collection and analysis are myriad, including state-representativeness of the data being analyzed, the need to collect cross-system data to perform cross-system analysis, a need for a large sample to achieve sufficient statistical power to be able to perform a conclusive analysis, or the need to perform complex analyses that require trained researchers. Formative assessment is an example of data collection and use that can happen entirely at the classroom level. This is the most effective DDDM cycle, since the data collector, data analyst, and policy maker are the same person—the educator themself. This person also intimately knows the subject for which the data provide information—the child. Unfortunately, not all processes can be analyzed in such a simple manner.

There are two kinds of questions data can answer (Hsu, 2005). One is in general: “What is?” What is the state of quality? What is the state of children’s learning? Questions like these are the easiest to answer (even though to obtain answers representative of all communities in the state, a proper sampling methodology is still required). The second type of question involves a mechanism of action such as: “Under what conditions is PD an effective way to improve teacher–child interactions?” Research designs to address this type of question generally fall into two categories—using administrative data or organizing a new data collection with a specific research question in mind. When using administrative data, one is forced to use data that are by definition not the best data for answering a given research question (since data was collected for administrative rather than research purposes). In such a scenario, one is limited as to the types of research designs this type of data allows: nonequivalent groups, post-test only, pre- and-post-test, propensity score matching, or regression discontinuity design (Riley-Ayers et al., 2011). Data collected with the express intention of answering a specific research question will by definition be much better suited to answering that question and will allow researchers to rule out alternative explanations for the conclusions reached (experimental design, sampling designs to understand descriptive statistics of subjects in a population from a sample of individuals).

The following are examples of ways states can utilize data to answer specific questions to inform decision making. Le et al. (2015) is an example of both using a sophisticated analysis that requires trained research staff and using a type of analysis that requires a lot of data. The authors used nonlinear modeling (i.e., splines; generalized additive model [GAM]) to explore which range of quality on ECERS-R and other quality measures from Colorado’s QRIS were associated with improvement in child outcomes. This is an important question, as ECERS-R is a widely used quality tool, and state efforts to improve quality need to be guided by what constitutes a reasonable goal in improvement as measured by this tool. Threshold analysis such as Le et al. (2015) can provide an answer to that question. However, since GAM is designed to find possibly different associations between ECERS-R and child outcomes along the range of ECERS-R, it requires a great deal of data covering the whole range (i.e., larger range of scores and standard deviation), a condition that tends to be difficult to satisfy, absent of having access to data from the whole state. Another study discussed the relationship between ECERS-R data and child outcomes (Setodji et al., 2018). Results indicate that "once classrooms achieved a score of 3.4 on the overall ECERS-R composite score, there was a leveling off effect, such that no additional improvements to children's social, cognitive, or language outcomes were observed" (Setodji et al., 2018, p. 1). The study yielded the result that ECERS-R has limitations for medium ratings in detecting factors associated with child outcomes. This initial variation in results from research studies indicates that more research is needed to establish whether these thresholds really exist in general and where they tend to be found.

An example of an analysis that is statewide by intent is Reid et al. (2019). The authors found significant variation in classroom quality by program setting using data from NYC's UPK program. While this study focused on one city-wide program, variations in quality caused by systemic obstacles, such as a divided governance structure and a lack of administrative cohesion, are faced by many other state-funded pre-K programs. This sort of analysis sheds light on how administrative organization of programs could stand in the way of quality improvement. States need to monitor program quality across settings to ensure all families are getting services at the intended quality level.

Klawiter and Sheng (2019) provide an example of an analysis that requires statewide data to provide representative understanding of the situation across the state. The authors conducted a systematic review of a multitude of such analyses in the research literature. This allowed them to investigate whether teachers who took part in a professional development program and received individualized coaching on emergent literacy instructional skills demonstrated improved pedagogical practice and teacher–child interactions. In another example of the use of statewide data, Huang et al. (2012) looked at the effects of the Virginia Preschool Initiative (VPI), Virginia's state-funded pre-K program, over time. The authors analyzed data from over 60,000 students from the beginning of kindergarten to the end of first grade using two-level hierarchical logistic regression models. Findings indicate that attending a VPI program was associated with a lower likelihood of repeating kindergarten and greater likelihood of meeting or exceeding minimum literacy competencies.

Early et al. (2005) examines data on state-funded pre-K programs in 11 states by combining data from two major studies: The Multi-State Study of Pre-Kindergarten that included six states, and the StateWide Early Education Programs Study that included five states.

Looking across states allowed this study to identify a larger pattern among state-funded pre-K programs: that while many of the programs met the existing professional guidelines for structural features of quality, the average process quality was lower than expected.

In summary, data used in the DDDM process may be collected by the state at the program, classroom, teacher, or child level to work towards improving quality. Some studies, such as Le et al. (2015) require more data-intensive analysis, which could only be achieved by gathering data at the state level. Looking at data can help states make informed decisions on promising practices.

| **Data Driven Decision Making**<br>**Infrastructure Indicators (state level)**<br>**DDDM and Equity** |
|---|
| **Indicator 9: DDDM to Improve Equity**<br><br>State teams have access to data on all populations (children, early childhood educators, and parents) that are part of the pre-K system. State keeps track of equity differences in quality and achievement across identified subgroups and takes steps to eliminate those differences by disaggregating data in meaningful ways, such as by race, income, language or other important traits that historically predict inequalities in outcome. |

| Data Driven Decision Making<br>Infrastructure Indicators (state level)<br>DDDM and Equity |
|---|
| The state's efforts to understand and address inequity include these five components:<br>• Ongoing data collection,<br>• Engaging in active discussions that surface issues of inequity for targeted populations,<br>• Action planning and implementation - creating action plans and following through with implementation,<br>• Planning to assess and refine implementation<br>• Amending policies and practices that address these issues |

In line with the framework of targeted universalism (Powell et al., 2019) used to guide the development of the IDM, equity indicators in each Element highlight the importance of ongoing data collection, the disaggregation of data, and the use of data for decision making, action planning, and assessing implementation. This supports the five steps of targeted universalism (Powell et al., 2019), where once a universal goal is established (Step 1), and there is information about the performance of the general population relative to the universal goal (Step 2), the performance of different groups can be identified (Step 3), further analysis can be done to understand the structures that support or impede each group for achieving the universal goal (Step 4), and targeted strategies for each group can be developed and implemented to reach the universal goal (Step 5).

The importance of disaggregation of data is also reflected in NAEYC's recommendations (NAEYC Recommendations for Public Policymakers #3), which emphasize the need for early learning standards to reflect culturally diverse settings. Only through data disaggregation can it be revealed to state decision-makers what the different needs and situations of different communities are in order to devise tailored interventions and strategies of engagement for those communities. NAEYC further recommends providing "ongoing, in-depth staff development on how to use standards in diverse classrooms" (NAEYC Recommendations for Public Policymakers #3). Such staff development could be made more equitable through the use of disaggregated data on PD practices and their effects in the classroom. Instead of pooling data from all communities and looking for the overall effect, disaggregation will allow state teams to understand that what works for one community may not work for another and vice versa. Similarly, NAEYC argues in favor of quality rating and improvement systems furthering "the principles of equity across all aspects of education, including curriculum, instruction, full inclusion, family engagement, program design, and delivery" (NAEYC Recommendations for Public Policymakers #3). All of these aspects would likewise benefit from monitoring progress community by community through the use of disaggregated data and design of intervention with focus on the needs of specific communities.

There are numerous ways states can implement efforts to understand and address inequity in their early childhood education systems. These strategies include conducting ongoing data collection, linking data and disaggregating them by community, and creating and implementing action plans. To promote systemwide equity, states need to keep track of equity in their pre-K programs and take steps to eliminate disparities. A fine example of a study that could be replicated by state teams once they have the necessary DDDM infrastructure in place is Bassok and Galdo (2016), which shows state pre-K programs in low-income and high-minority communities in Georgia having significantly lower ratings on CLASS (a widely used and validated measure of classroom quality, as previously described in this paper). In another study, Hatfield et al. (2015) used multiple statewide data sources to better understand how characteristics of communities and programs were related to points awarded in North Carolina's Tiered Quality Rating and Improvement System (TQRIS). The study found inequities in the availability of highest-quality programs based on differences in community context, including socioeconomic differences among communities. In their study of the Virginia Preschool Initiative programs (see above for more details), Huang et al. (2012) found that Hispanic, Black, and disabled children who attended VPI programs experienced academic benefits until the end of the first grade despite following up their pre-K education in elementary schools with high concentrations of poverty.

Other research has used cross-state analysis to examine issues of equity.  For example, a study by Early et al. (2010) analyzed classroom observations of children in pre-K programs in 11 states. Their findings indicated that at the classroom-level, classes with lower percentages of Latino and African American children, and children from economically disadvantaged households, were generally spending more time engaged in more meaningful learning experiences. In another excellent example of equity-focused analysis, Valentino (2018) found notable gaps in quality between public pre-K programs attended by more low-income students or students of color compared to programs attended by higher-income and fewer students of color. Valentino (2018) ascribes these state-level gaps in pre-K quality to state-level residential segregation, showing that states with higher residential segregation have greater gaps between low income, students of color, and the non-poor, non-minority children. Valentino (2018) points out that pre-K cannot accomplish its mission of narrowing the achievement gap in education and adult outcomes if it itself suffers from inequitable access.

A 2015 report by Reid and Kagan offers another example of how data analysis may be used to reveal equity issues. This report discussed concerning trends in the demographic data, such as low-income children being more likely to attend low-quality preschool programs, and that the majority of public preschool programs are economically segregated, and often also segregated by race and/or ethnicity (Reid & Kagan, 2015).

These examples show the importance of data linkages in equity work. Findings of the kind discussed above can immediately alert state teams to this issue. Being aware helps them focus on mapping these disparities, looking for their causes, reviewing the way the state has been engaging with those communities, studying any prior interventions in those communities, analyzing which approaches worked and which did not, and proposing new interventions to address the disparities.

**Figure 6**

*DDDM Inequities of Focus in the Literature*

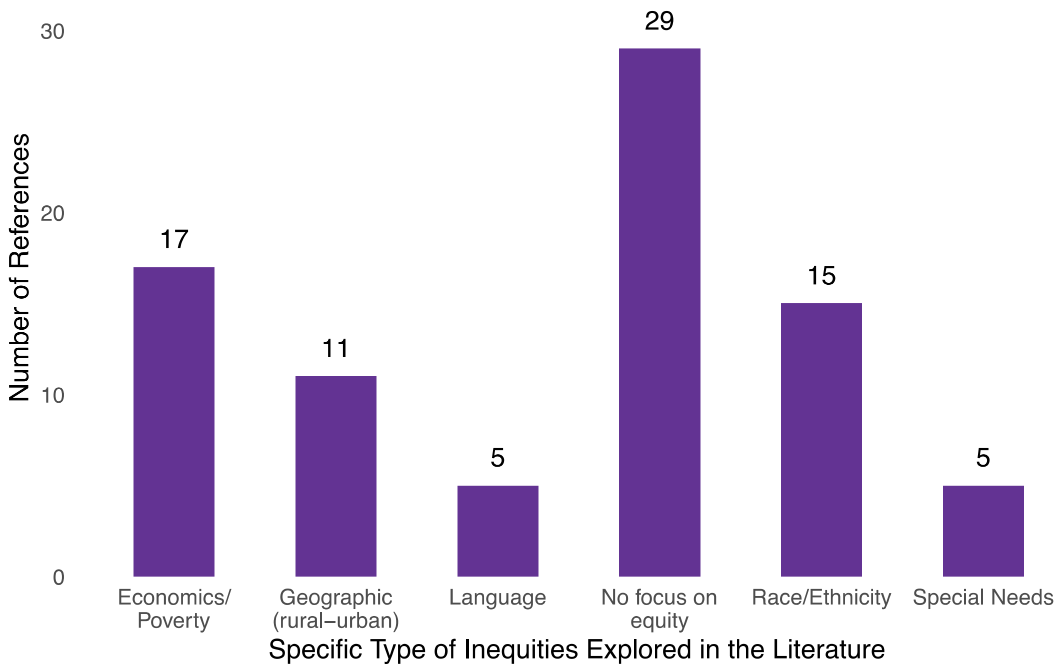



Figure 6 provides a summary of the studies underpinning the DDDM Element regarding their focus on equity. The figure shows that many studies included in the scan focused on equity and multiple types of equity. Some studies did not contain direct links to child outcomes with an equity focus (29). Of those that did, most focused on economic equity (17), followed by racial (15), and geographic (11). Special needs was the category that was focused on by the fewest studies (5). This provides state teams a good understanding of which equity categories the literature focused on, and if this does not align with the states' own equity focus, putting such issues on the equity research-practice partnerships (RPP) agenda of the states' team would be warranted.

# IV. Future Directions and Limitations

As mentioned in the introduction, DDDM as a field of inquiry within the education sciences was born out of the No Child Left Behind legislation introduced in 2001 (Piety, 2019; Urban et al., 2019). The urgent need for this approach has been communicated top-down from the federal government (through the Department of Education) to the states, school districts, and schools. Early learning services, mostly operated from the Department of Health and Human Services, have been spared this drive for accountability by the federal government (even if minor pressure was applied to it via the Improving Head Start Act of 2007). As a result, as of this writing in 2021, the adoption of the DDDM philosophy has been lukewarm at best in the early learning community. Presently, it is treated as a luxury one cannot afford; however, the example of K-12 shows that if the field continues to treat it that way, DDDM will end up being imposed upon it by the federal government. Moreover, it will come in the spirit of accountability rather than the spirit of improvement (Firestone & Gonzalez, 2007; Urban et al., 2019).

As a result of this slow pace of cultural change, few studies analyze the impact of DDDM on early learning, as one cannot study empirically what one cannot observe. The gap becomes obvious when one compares the way DDDM has progressed within K-12 with the progress made in the early learning space. A similarly illuminating comparison could be made within the academic literature on the impact of DDDM. While data is available within pre-K, the infrastructure and capacity to use it are very low at all levels, from teacher to state early learning agencies (Little et al., 2019; Loeb, 2012; Tout et al., 2015). Partnering with an RPP has been suggested as a solution to this problem (Coburn & Penuel, 2016; Henrick et al., 2017). As a result of this paucity of progress at the level of state early learning agencies, most of our evidence regarding DDDM comes from small pilot studies (Stein et al., 2013). We cite this evidence above even for indicators that recommend action at the state level because we believe that these small pilot studies have demonstrated that the benefits are tangible and that it is time to scale these efforts to the next level. Experts in the field agree (Tout et al., 2015).

While data is an important source of insights, in the end, decisions are best made by melding insights originated from data analysis with experience of the decision-maker and local context (Loeb, 2012). There is also a question in the field regarding the role of equity in CQI, including unequal access to resources and support, determining metrics of success, establishing standards, and who is given a voice to make decisions (BUILD Initiative, 2017). Hence, while the field agrees that this is the way forward, given the lack of evidence, to some extent education agencies will have to blaze their own way forward. The IDM provides as much guidance as possible given the evidence available. It is up to the state teams to fill in the rest via their engagement in CQI, because CQI is about learning to learn (Bryk et al., 2015).

# References

2018 Early Childhood Data Systems Survey, King, C., Perkins, V., Nugent, C., & Jordan, E. (2018). 2018 State of state early childhood data systems. https://www.childtrends.org/wp-content/uploads/2018/09/ECDC-50-state-survey-9.25.pdf

Abbott, M., Beecher, C., Petersen, S., Greenwood, C. R., & Atwater, J. (2017). A team approach to data-driven decision-making literacy instruction in preschool classrooms: Child assessment and intervention through classroom team self-reflection. *Young Exceptional Children*, *20*(3), 117–132.

Alexander, M. (2020). The new Jim Crow: Mass incarceration in the age of colorblindness. The New Press.

Apple, P. L. (2006). A developmental approach to early childhood program quality improvement: The relation between state regulation and NAEYC accreditation. *Early Education and Development*, *17*(4), 535–552. https://doi.org/10.1207/s15566935eed1704_2

Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review*, *99*(1), 544–55.

Au, W. (2011). Teaching under the new Taylorism: High-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, *43*(1), 25–45.

Balfour, L. (2003). Unreconstructed democracy: WEB Du Bois and the case for reparations. *American Political Science Review, 97*(1), 33–44. https://doi.org/10.1017/S0003055403000509

Barnett, W. S., Jung, K., Youn, M., & Frede, E. C. (2013). Abbott preschool program longitudinal effects study: Fifth grade follow-up. *New Brunswick, NJ: National Institute for Early Education Research*, *10*, 2001–2004.

Bassok, D. & Galdo, E. (2016). Inequality in preschool quality? Community-level disparities in access to high-quality learning environments. *Early Education and Development, 27*(1), 128–144. doi.org/10.1080/10409289.2015.1057463

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, *2*(1), 23.

Beatty, B. (1995). *Preschool education in America: The culture of young children from the colonial era to the present*. Yale University Press.

Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, *70*(3), 489–520.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652–1678.

Boller, K., Tarrant, K. & Schaack, D. D. (2014). Early care and education quality improvement: A typology of intervention approaches. OPRE Research Report #2014-36. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Braaten, M., Bradford, C., Kirchgasler, K. L., & Barocas, S. F. (2017). How data use for accountability undermines equitable science education. *Journal of Educational Administration, 55*(4), 427–446.

Brawley, S., & Stormont, M. A. (2014). Investigating reported data practices in early childhood: An exploratory study. *Journal of Positive Behavior Interventions*, 16(2), 102–111. https://doi.org/10.1177/1098300713480838

Brock, M. E., & Beaman-Diglia, L. E. (2018). Efficacy of coaching preschool teachers to manage challenging behavior. *Education & Treatment of Children*, 41(1), 31–48. https://doi.org/10.1353/etc.2018.0001

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). Learning to improve: How America's schools can get better at getting better. Harvard Education Press.

BUILD Initiative. (2017). Continuous quality improvement in early childhood and school age programs: An update from the field. Build Initiative.

Buzhardt, J., Greenwood, C. R., Walker, D., Anderson, R., Howard, W., & Carta, J. J. (2011). Effects of web-based support on Early Head Start home visitors' use of evidence-based intervention decision making and growth in children's expressive communication. *NHSA Dialog*, 14(3), 121–146.

Callahan, R. E. (1962). Education and the cult of efficiency. University of Chicago Press.

Casbergue, R. M., Bedford, A. W., & Burstein, K. (2014). CLASS reliability training as professional development for preschool teachers. *Journal of Research in Childhood Education*, 28(4), 426–440. https://doi.org/10.1080/02568543.2014.944724

Chetty, R., & Friedman, J. N. (2011). Does local tax financing of public schools perpetuate inequality? In National Tax Association Proceedings: Vol. 103, pp. 112–118.

Clifford, R. M., Reszka, S. S., & Rossback, H. (2010). Reliability and validity of the Early Childhood Environment Rating Scale. Retrieved May 30, 2010, from http://ers.fpg.unc.edu/sites/ers.fpg.unc.edu/files/ReliabilityEcers.pdf.

Clotfelter, C. T. (2011). After Brown: The rise and retreat of school desegregation. Princeton University Press.

Coburn, C. E., & Penuel, W. R. (2016). Research-practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, 45(1), 48–54.

Cohen-Vogel, L. (2011). "Staffing to the test": Are today's school personnel practices evidence based? *Educational Evaluation and Policy Analysis*, 33(4), 483–505. https://doi.org/10.3102/0162373711419845

Cohen-Vogel, L., & Harrison, C. (2013). Leading with data: Evidence from the National Center on Scaling Up Effective Schools. *Leadership and Policy in Schools*, 12(2), 122–145. https://doi.org/10.1080/15700763.2013.792934

Connors, M. C., & Morris, P. A. (2015). Comparing state policy approaches to early care and education quality: A multidimensional assessment of quality rating and improvement systems and child care licensing regulations. *Early Childhood Research Quarterly, 30*(Part B), 266–279. https://doi.org/10.1016/j.ecresq.2014.05.006

Conroy, M. A., Sutherland, K. S., Algina, J., Ladwig, C., Werch, B., Martinez, J., Jessee, G., & Gyure, M. (2019). Outcomes of the best in class intervention on teachers' use of effective practices, self-efficacy, and classroom quality. *School Psychology Review*, *48*(1), 31–45.

Crawford, A. D., Zucker, T. A., Williams, J. M., Bhavsar, V., & Landry, S. H. (2013). Initial validation of the prekindergarten classroom observation tool and goal setting system for data-based coaching. *School Psychology Quarterly*, *28*(4), 277–300.

Early, D., Barbarin, O., Bryant, B., Burchinal, M., Chang, F., Clifford, R., & Weaver, W. (2005). Pre-Kindergarten in eleven states: NCEDL's multi-state study of pre-kindergarten and state-wide early educational programs (SWEEP) study. *National Institute for Early Education Research, Washington, DC*.

Early, D. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. https://www.sciencedirect.com/science/article/abs/pii/S0885200609000726

Easterly, W. (2017). *The white man's burden: Why the West's efforts to aid the rest have done so much ill and so little good* [Audiobook]. Tantor Media.

Elicker, J., & Ruprecht, K. (2019). Child care quality rating and improvement systems (QRIS). In *The Wiley Handbook of Early Childhood Care and Education* (pp. 515–536). Wiley.

Firestone, W. A., & González, R. A. (2007). Chapter 6 culture and processes affecting data use in school districts. *Yearbook of the National Society for the Study of Education*, *106*(1), 132–154.

Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2015). Implementation science. *International encyclopedia of the social and behavioral sciences*, *11*, 695–702.

Friedman, M. (1997). Public schools: Make them private. *Education Economics*, *5*(3), 341–344.

Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & Gardiner, B. (2020). *The state of preschool 2019: State preschool yearbook.* National Institute for Early Education Research.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2.

Fryer Jr, R. G., Levitt, S. D., & List, J. A. (2015). *Parental incentives and early childhood achievement: A field experiment in Chicago heights* (No. w21477). National Bureau of Economic Research.

Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *American Economic Review*, *92*(4), 999–1012.

Gardner, D. P. (1983). A nation at risk: The imperative for educational reform. An open letter to the American people. *A Report to the Nation and the Secretary of Education.*

Goldstein, H., McKenna, M., Barker, R. M., & Brown, T. H. (2019). Research-practice partnership: Application to implementation of multi-tiered system of supports in early childhood education. *Grantee Submission*, *4*, 38–50.

Goodvin, R., & Hansen, J. (2019). Early Achievers evaluation report one: Background and research design (Document Number 19-12-2202). Olympia: Washington State Institute for Public Policy.

Goodvin, R., Rashid, A., & He, L. (2020). Early Achievers evaluation report two: Prekindergarten quality and child outcomes in kindergarten (Document Number 20-12-2203). Olympia: Washington State Institute for Public Policy.

Hallam, R. A., Hooper, A., Buell, M., Ziegler, M., & Han, M. (2019). Boosting family child care success in quality rating and improvement systems. *Early Childhood Research Quarterly, 47*, 239–247. https://doi.org/10.1016/j.ecresq.2018.12.008

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23.

Hamad, M. M. and Jihad, A. A. (2011, December 6–8). An enhanced technique to clean data in the data warehouse, *2011 Developments in e-Systems Engineering*, pp. 306–311. Dubai, 2011. doi: 10.1109/DeSE.2011.32.

Hamre, B. K., Goffin, S. G., & Kraft-Sayre, M. (2009). *Classroom Assessment Scoring System* (CLASS) *implementation guide: Measuring and improving classroom interactions in early childhood settings*. Center for Advanced Study of Teaching and Learning.

Hanno, E. C., & Gonzalez, K. E. (2020). The effects of teacher professional development on children's attendance in preschool. *Journal of Research on Educational Effectiveness, 13*(1), 3–28.

Haslip, M. J., & Gullo, D. F. (2018). The changing landscape of early childhood education: Implications for policy and practice. *Early Childhood Education Journal, 46*(3), 249–264. https://doi.org/10.1007/s10643-017-0865-7

Hatfield. (2015). Inequities in access to early quality care and education: Associations with funding and community context. *Early Childhood Research Quarterly, 30*(Part B), 316–326. https://www.sciencedirect.com/science/article/abs/pii/S0885200614000039

Heckman, J. J. (2013). *Giving kids a fair chance*. MIT Press.

Hedges, L. V. (2018). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, *11*(1), 1–21.

Henrick, E. C., Cobb, P., Penuel, W. R., Jackson, K., & Clark, T. (2017). *Assessing research-practice partnerships: Five dimensions of effectiveness.* William T. Grant Foundation.

Hestenes, L. L., Kintner-Duffy, V., Wang, Y. C., La Paro, K., Mims, S. U., Crosby, D., Scott-Little, C., & Cassidy, D. J. (2015). Comparisons among quality measures in child care settings: Understanding the use of multiple measures in North Carolina's QRIS and their links to social-emotional development in preschool children. *Early Childhood Research Quarterly, 30*(Part B), 199–214. https://doi.org/10.1016/j.ecresq.2014.06.003

Hindman, A. H., Snell, E. K., Wasik, B. A., Lewis, K. N., Hammer, C. S., & Iannone-Campbell, C. (2015). Research and practice partnerships for professional development in early childhood: Lessons from ExCELL-e. *Journal of Education for Students Placed at Risk*, *20*(1–2), 12–28. https://doi.org/10.1080/10824669.2014.984036

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, *7*, 24.

Hong, S. L. S., Howes, C., Marcella, J., Zucker, E., & Huang, Y. (2015). Quality rating and improvement systems: Validation of a local implementation in LA county and children's school-readiness. *Early Childhood Research Quarterly, 30*(Part B), 227–240. https://doi.org/10.1016/j.ecresq.2014.05.001

Hooks, L. M., Scott-Little, C., Marshall, B. J., & Brown, G. (2006). Accountability for quality: One state's experience in improving practice. *Early Childhood Education Journal*, *33*(6), 399–403. https://doi.org/10.1007/s10643-006-0065-3

Hsu, T. C. (2005). Research methods and data analysis procedures used by educational researchers. *International Journal of Research & Method in Education, 28*(2), 109–133.

Huang, F. L., Invernizzi, M. A., & Drake, E. A. (2012). The differential effects of preschool: Evidence from Virginia. *Early Childhood Research Quarterly*, *27*(1), 33–45.

Jeon, L., & Buettner, C. K. (2015). Quality rating and improvement systems and children's cognitive development. *Child & Youth Care Forum, 44*(2), 191–207. https://doi.org/10.1007/s10566-014-9277-7

Johnson, A. D., Martin, A., & Schochet, O. N. (2020). Inside the classroom door: Understanding early care and education workforce and classroom characteristics experienced by children in subsidized center-based care. *Early Childhood Research Quarterly*, *51*, 462–472. https://doi.org/10.1016/j.ecresq.2020.01.006

Johnson, A. D., Martin, A., & Schochet, O. N. (2019). How do early care and education workforce and classroom characteristics differ between subsidized centers and available center-based alternatives for low-income children? *Children and Youth Services Review*, *107*. https://doi.org/10.1016/j.childyouth.2019.104567

Karoly, L. A., Zellman, G. L., & Perlman, M. (2013). Understanding variation in classroom quality within early childhood centers: Evidence from Colorado's quality rating and improvement system. *Early Childhood Research Quarterly*, *28*(4), 645–657. https://doi.org/10.1016/j.ecresq.2013.05.001

Kenney, C. (2008). *The best practice: How the new quality movement is transforming medicine*. Public Affairs.

Klawiter, A. & Sheng, L. (2019). Efficacy of professional development with individualized coaching to enhance educator knowledge and practice of emergent literacy skills. *Evidence-Based Practice Briefs, 13*(6).

Lahti, M., Elicker, J., Zellman, G., & Fiene, R. (2015). Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly, 30*(Part B), 280–290. https://doi.org/10.1016/j.ecresq.2014.04.005

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology*, *101*(2), 448–465. https://doi.org/10.1037/a0013842

Landry, S. H., Swank, P. R., Anthony, J. L., & Assel, M. A. (2011). An experimental study evaluating professional development activities within a state funded pre-kindergarten program. *Reading and Writing: An Interdisciplinary Journal*, *24*(8), 971–1010.

Le, V.-N., Schaack, D. D., & Setodji, C. M. (2015). Identifying baseline and ceiling thresholds within the Qualistar Early Learning Quality Rating and Improvement System. *Early Childhood Research Quarterly, 30* (Part B), 215–226. https://doi.org/10.1016/j.ecresq.2014.03.003

Limlingan, M., Grindal, T., Lopez, M., Blocklin, M., & Bumgarner, E. (2015). Integrated data systems: An emerging tool to support services for low-income Hispanic families with young children. National Research Center on Hispanic Children & Families. https://www.hispanicresearchcenter.org/wp-content/uploads/2019/08/Integrated-Data-Systems-Brief-8.19-V21.pdf

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, *45*, 155–176.

Little, M., Cohen-Vogel, L., Sadler, J., & Merrill, B. (2019). Data-driven decision making in early education: Evidence from North Carolina's pre-k program. *Education Policy Analysis Archives, 27*(18). http://offcampus.lib.washington.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1210325&site=ehost-live

Lockton, M., Weddle, H., & Datnow, A. (2020). When data don't drive: teacher agency in data use efforts in low-performing schools. *School Effectiveness and School Improvement*, *31*(2), 243–265.

Loeb, S. (2012). In light of the limitations of data-driven decision making. *Education Finance and Policy*, 7(1), 1–7.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71–85.

Massey, D., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.

McGinty, A. S., Justice, L., & Rimm-Kaufman, S. E. (2008). Sense of school community for preschool teachers serving at-risk children. *Early Education and Development*, *19*(2), 361–384. https://doi.org/10.1080/10409280801964036

Mischel, W. (2014). *The marshmallow test: Understanding self-control and how to master it*. Random House.

Moyo, D. (2009). *Dead aid: Why aid is not working and how there is a better way for Africa*. Macmillan.

National Center on Program Management and Fiscal Operations. (2020). *What is quality data for programs serving infants and toddlers.* Office of Head Start National Centers, Office of Head Start, Administration for Children and Families, U.S. Department of Health and Human Services. U.S. Government Printing Office. Retrieved October 30, 2020 from https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/what-is-quality-data-infants-toddlers.pdf

Pak, K., & Desimone, L. M. (2019). Developing principals' data-driven decision-making capacity: Lessons from one urban district. *Phi Delta Kappan*, *100*(7), 37–42.

Pearman, F. A. (2020). The moderating effect of neighborhood poverty on preschool effectiveness: Evidence from the Tennessee voluntary prekindergarten experiment. *American Educational Research Journal*, *57*(3), 1323–1357.

Piasta, S. B., Farley, K. S., Mauck, S. A., Soto Ramirez, P., Schachter, R. E., O'Connell, A. A., Justice, L. M., Spear, C. F., & Weber-Mayrer, M. (2020). At-scale, state-sponsored language and literacy professional development: Impacts on early childhood classroom practices and children's outcomes. *Journal of Educational Psychology*, *112*(2), 329–343. https://doi.org/10.1037/edu0000380

Piety, P. J. (2019). Components, infrastructures, and capacity: The quest for the impact of actionable data use on P–20 educator practice. *Review of Research in Education, 43*(1), 394–421. https://doi.org/10.3102/0091732X18821116

Powell. (2008). Post-racialism or targeted universalism? *Denver Law Review*, *86*(3). https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1432&context=dlr

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, *23*(1), 10–26. https://doi.org/10.1016/j.ecresq.2007.09.001

Reardon, S. F., & Owens, A. (2014). 60 years after Brown: Trends and consequences of school segregation. *Annual Review of Sociology, 40*.

Rebitzer, J. B., & Taylor, L. J. (2011). Extrinsic rewards and intrinsic motives: Standard and behavioral approaches to agency and labor markets. In *Handbook of Labor Economics* (Vol. 4, pp. 701–772). Elsevier.

Reid, J. L., & Kagan, S. L. (2015). *A better start: Why classroom diversity matters in early education.* Poverty & Race Research Action Council.

Reid, J. L., Melvin, S. A., Kagan, S. L., & Brooks-Gunn, J. (2019). Building a unified system for universal pre-k: The case of New York City. *Children and Youth Services Review, 100*, 191–205. https://doi.org/10.1016/j.childyouth.2019.02.030

Richesson, R. L., Krischer, J. (2007). Data standards in clinical research: Gaps, overlaps, challenges and future directions, *Journal of the American Medical Informatics Association*, 14(6) 687–696. https://doi.org/10.1197/jamia.M2470

Riley-Ayers, S., Frede, E., Barnett, S. W., & Brenneman, K. (2011). *Improving early education programs through data-based decision making.* National Institute for Early Education Research.

Ryan, J. E. (2004). The perverse incentives of the No Child Left Behind Act. *New York University Law Review*, *79*(3), 932.

Sabol, T. J., & Pianta, R. C. (2015). Validating Virginia's quality rating and improvement system among state-funded pre-kindergarten programs. *Early Childhood Research Quarterly, 30*(Part B), 183–198. https://doi.org/10.1016/j.ecresq.2014.03.004

Schelling, N., & Rubenstein, L. D. (2021). Elementary teachers' perceptions of data-driven decision-making. *Educational Assessment, Evaluation and Accountability*, *33*(2), 317–344.

Setodji, C. M., Schaack, D., & Le, V.-N. (2018). Using the Early Childhood Environment Rating Scale-Revised in high stakes contexts: Does evidence warrant the practice? *Early Childhood Research Quarterly, 42*, 158–169. https://doi.org/10.1016/j.ecresq.2017.10.001

Sheridan, S. M., Koziol, N., Witte, A. L., Iruka, I., & Knoche, L. L. (2020). Longitudinal and geographic trends in family engagement during the pre-kindergarten to kindergarten transition. *Early Childhood Education Journal*, *48*(3), 365–377. https://doi.org/10.1007/s10643-019-01008-5

Slentz, K. L. (2008). A guide to assessment in early childhood: Infancy to age eight. Washington State Office of Superintendent of Public Instruction.

Smith, P. C., Budzeika, K. A., Edwards, N. A., Johnson, S. M., & Bearse, L. N. (1986). Guidelines for clean data: Detection of common mistakes. *Journal of Applied Psychology, 71*(3), 457–460. https://doi.org/10.1037/0021-9010.71.3.457

Soderberg, J., Joseph, G. E., Stull, S., & Hassairi, N. (2016). Early Achievers standards validation study. Childcare Quality and Early Learning Center for Research & Professional Development.

Solari, E. J., Zucker, T. A., Landry, S. H., & Williams, J. M. (2016). Relative effects of a comprehensive versus reduced training for Head Start teachers who serve Spanish-speaking English learners. *Early Education and Development*, *27*(7), 1060–1076.

Stein, A., & Connors, M. C. (2016, March 3rd). *Measuring data use beliefs and practices in early education settings* [Conference abstract]. Society for Research on Educational Effectiveness, Washington, DC.

Stein, A., Freel, K., Hanson, A., Pacchiano, D., & Eiland-Williford, B. (2013). The Educare Chicago research-program partnership and follow-up study: Using data on program graduates to enhance quality improvement efforts. *Early Education & Development, 24*, 19–41. https://doi.org/10.1080/10409289.2013.739542

Taylor, F. W. (1919). *The principles of scientific management*. Harper & Brothers.

Thomas, J., Cook, T. D., Klein, A., Starkey, P., & DeFlorio, L. (2018). The sequential scale-up of an evidence-based intervention: A case study. *Evaluation Review*, *42*(3), 318–357.

Tout, K., Epstein, D., Soli, M., & Lowe, C. (2015). *A blueprint for early care and education quality improvement initiatives: Final report.* Child Trends.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Urban, W. J., Wagoner Jr, J. L., & Gaither, M. (2019). *American education: A history.* Routledge.

U.S. Department of Education. (2006). Improving data quality for Title I standards, assessments, and accountability reporting. U.S. Department of Education.

Valentino, R. (2018). Will public pre-k really close achievement gaps? Gaps in prekindergarten quality between students and across states. *American Educational Research Journal, 55*(1), 79–116.

Vitiello, V. E., Bassok, D., Hamre, B. K., Player, D., & Williford, A. P. (2018). Measuring the quality of teacher–child interactions at scale: Comparing research-based and state observation approaches. *Early Childhood Research Quarterly, 44*, 161–169. https://doi.org/10.1016/j.ecresq.2018.03.003

Woodward, C. V., & McFeely, W. S. (2002). *The strange career of Jim Crow*. Oxford University Press.

Yazejian, N., & Iruka, I. U. (2015). Associations among tiered quality rating and improvement system supports and quality improvement. *Early Childhood Research Quarterly*, *30*(Part B), 255–265. https://doi.org/10.1016/j.ecresq.2014.05.005

Yoder, M. L., & Williford, A. P. (2019). Teacher perception of preschool disruptive vehavior: Prevalence and contributing factors. *Early Education and Development*, *30*(7), 835–853.

Yu, S. (2019). Head Start teachers' attitudes and perceived competence toward inclusion. *Journal of Early Intervention, 41*(1), 30–43.

Zamarro, G., Engberg, J., Saavedra, J. E., & Steele, J. (2015). Disentangling disadvantage: Can we distinguish good teaching from classroom composition? *Journal of Research on Educational Effectiveness*, *8*(1), 84–111.

Zhang, L. (2014). The Abdul Latif Jameel Poverty Action Lab: Bringing evidence-based policy into international development. *Harvard International Review*, *35*(4), 4.

Zigler, E., & Styfco, S. J. (2010). *The hidden history of Head Start*. Oxford University Press.

Zigler, E. F., Singer, D. G., & Bishop-Josef, S. J. (2004). *Children's play: The roots of reading*. ZERO TO THREE/National Center for Infants, Toddlers and Families.

Zweig, J., Irwin, C. W., Kook, J. F., & Cox, J. (2015). Data collection and use in early childhood education programs: Evidence from the Northeast Region (REL 2015–084). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory North East & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

# Appendix A

**Table 1 Data-Driven Decision Key Words and Reference Summary**

| Key word or phrase | # Articles for initial abstract review based on inclusion criteria | # Articles for 2nd abstract review with exclusion criteria | # Articles passed full article review | Article citation |
|---|---|---|---|---|
| Data-driven decision-making<br>(new and cross key phrase) | 48 | 10 | 6 | Abbott et al., 2017 (DDDM2); Brawley & Stormont, 2014 (DDDM2); Little et al., 2019 (DDDM2); Piety, 2019 (DDDM1); Valentino, 2018 (DDDM1, 2, & 9); Vitiello et al., 2018 (DDDM2, 3) |
| Continuous quality improvement (CQI)<br>(continuous improvement new search) | 30 | 6 | 6 | Stein & Connors, 2016 (DDDM2); Johnson et al., 2019 (DDDM6); Johnson et al., 2020 (DDDM4, 6); Piasta et al., 2020 (DDDM4); Reid et al., 2019 (DDDM7); Sheridan et al., 2020 (DDDM4) |
| Quality rating & improvement system (QRIS) | 18 | 13 | 11 | Connors & Morris, 2015 (DDDM3); Hallam et al., 2019 (DDDM3); Hestenes et al., 2015 (DDDM3); Hong et al., 2015 (DDDM3); Jeon & Buettner, 2015 (DDDM3); Karoly et al., 2013 (DDDM3); Lahti et al., 2015 (DDDM3); Le et al., 2015 (DDDM3); Sabol & Pianta, 2015 (DDDM3); Setodji et al., 2018 (DDDM5); Yazejian & Iruka, 2015 (DDDM3) |
| Research-practice partnership (RPP) | 4 | 2 | 3 | Goldstein et al., 2019 (DDDM2); Hindman et al., 2015 (DDDM4); Henrick et al., 2017 (DDDM5, 8) |
| Network improvement communities (NIC) | 109 | 6 | 5 | Brock & Beaman-Diglia, 2018 (DDDM6); Casbergue et al., 2014 (DDDM3); Landry et al., 2009 (DDDM8); McGinty et al., 2008 (DDDM4); Raver et al., 2008 (DDDM6) |

| Key word or phrase | # Articles for initial abstract review based on inclusion criteria | # Articles for 2nd abstract review with exclusion criteria | # Articles passed full article review | Article citation |
|---|---|---|---|---|
| Progress monitoring | 20 | 10 | 5 | Buzhardt et al., 2011 (DDDM8); Crawford et al., 2013 (DDDM6); Haslip & Gullo, 2018 (DDDM4, 9); Landry et al., 2011 (DDDM4); Solari et al., 2016 (DDDM6) |
| Performance management | 60 | 11 | 5 | Conroy et al., 2019 (DDDM4); Hanno & Gonzalez, 2020 (DDDM2); Klawiter et al., 2019 (DDDM6); Yoder & Williford, 2019 (DDDM6); Yu, 2019 (DDDM8) |
| Education improvement | 2 | 2 | 1 | Apple, 2006 (DDDM2) |
| State accountability | 1 | 1 | 1 | Hooks et al., 2006 (DDDM2) |
| Expert recommendation | NA | NA | 14 | Slentz, 2008 (DDDM1); Riley-Ayers, 2011 (DDDM8); Boller et al., 2014 (DDDM2, 3, & 7); Bassok & Galdo, 2016 (DDDM3, 7, & 9); Tout et al., 2015 (DDDM3, 6); King et al., 2018 (DDDM4, 5, 6, & 7); Coburn & Penuel, 2016 (DDDM6); Stein et al., 2013 (DDDM4); Early et al., 2010 (DDDM9); Hatfield, 2015 (DDDM9); Powell, 2008 (DDDM9); Reid & Kagan, 2015 (DDDM9); Huang et al., 2012 (DDDM8); Goodvin et al., 2020 (DDDM3) |
| **Total** | 292 | 61 | 57 | |

**Data Driven Decision-Making Literature Review Summary (excluded articles)**

| Key word or phrase | # Articles for initial abstract review based on inclusion criteria | # Articles for 2nd abstract review with exclusion criteria | # Articles passed full article review |
|---|---|---|---|
| Evidence-based decision-making | 45 | 3 | 0 |
| Program improvement | 17 | 1 | 0 |
| Total | 62 | 4 | 0 |

# Appendix B

## Overview of Improvement and Implementation Science

### Pre-K: Pilots to scale-ups (implementation science with application to pre-K)

What are the foundations of the pre-K industry? How do we know that it works? It all started with Perry preschool and the Abecedarian project (Heckman, 2013). Thanks to those rigorous research studies, we have evidence beyond reasonable doubt that early intervention could change the lives of children and families for the better. Another well-known example was the Abbot preschool program (Barnett et al., 2013). Perry intervention ran from 1962 to 1965. Head Start was established in 1965 in the context of President Johnson's War on Poverty. The amount of trust in the promise of preschool is evidenced in the fact that a formal evaluation of the Head Start program's effectiveness was only commissioned in 1998, 33 years after its founding. The goodwill for pre-K generated by Perry and Abecedarian eventually ran out, and policymakers once again required evidence of effectiveness. It did not help that the Head Start Impact Study commissioned by Congress in 1998 showed evidence of early gains and subsequent fade-out of those gains (Garces & Currie, 2002). The word fade-out has appeared again in scientific literature with the rigorous evaluation of the Tennessee Voluntary Pre-K program (VPK) (Lipsey et al., 2018). Subsequent studies partially rehabilitated VPK's effectiveness for literacy achievement among children living in high-poverty neighborhoods (Pearman, 2020). Some researchers say fade-out is only temporary and has not prevented participating children from reaping gains in adulthood (Garces & Currie, 2002; Heckman, 2013).

Still, the setback of fade-out has left some policymakers with a bitter after-taste and politicians with a healthy amount of skepticism. It does not help that the field has conducted very few impact evaluations that are methodologically sound. How is it that small pilots demonstrated unquestionable benefits, yet scale-up large programs failed to generate the same enthusiasm? What is the difference between Perry, VPK, and Head Start? Researchers suspect it has something to do with the transition from pilots to large-scale programs and fidelity to implementation (Fixsen et al., 2015). Thomas et al. (2018) set out to study this hypothesis. They distinguished among the following stages of implementation:

8. Pilot
9. Efficacy research
10. Effectiveness studies
11. Scale-up studies

Using the example of the pre-K mathematics program introduced in California, they have demonstrated that effect sizes indeed tend to fall between pilot and implementation. Effect size is a ratio of intervention's effect and the standard deviation of the control group and was designed by statisticians to put studies on equal footing in cases they use measures with different scales (Fritz et al., 2012). Thomas et al. (2018) note that falling effect sizes from pilot to scale-up would then be either due to:

1. Lower intervention effect
2. Higher standard deviation of the control group

The former would take place if fidelity to implementation (or evaluation) was lower for some reason, while the latter would happen if the larger sample of children and classrooms was more heterogeneous than the pilot group (for example, if it included rural, urban, and suburban programs rather than just one of those categories, as would be most likely the case for a pilot program). Thomas et al. (2018) then propose that the following could lead to falling effect sizes (either because of increasing heterogeneity or decreasing effect of the intervention):

1. Number and heterogeneity of study sites goes up.
2. Number and heterogeneity of participants (children) goes up.
3. Quality of program delivery goes down.
4. Quality of program content goes down.
5. The quality of the control group's education goes up (if the intervention has become well-known, it is possible that other programs are implementing some features even without being prodded to do so).
6. Quality of measurement goes down (introducing more noise and hence increasing standard deviation in the control group).
7. Quality of assessment goes down (because the assessment tool does not capture the dimensions of education that the intervention is targeting).
8. Retention of children participants may become harder in the scale-up evaluation.
9. Study rigor may fall in the scale-up.

Thomas et al. (2018) point out that the whole idea of a scale-up is to spend less money per child on the intervention compared to the pilot. So, the failed effectiveness of scale-up may be due to money being cut in places that are crucial to the intervention's effectiveness. To escape the trap of implementation gap, we need the following:

1. A tool that can help keep track of all the aspects of successful preschool implementation (IDM).
2. The ability to learn from mistakes and to take action in such a way that if success is not achieved, learning can be obtained as a silver lining. This is the subject of the next section, and the IDM CQI Element helps state early education agencies chart the way forward to acquire this ability.

### Improvement Science

Bryk et al. (2015) provide an anatomy of an educational reform failure. Their insight is that the reform process usually proceeds in the following steps:

1. Observe
2. Make a hypothesis
3. Do a pilot
4. Scale up
5. Question the mounting evidence of failure
6. Admit failure and find a scapegoat
7. Make a new hypothesis

We will call the above "the impatient approach." Both proponents of the improvement science approach (Bryk et al., 2015) and the implementation science approach (Fixsen et al., 2015) argue that the following process would be a much better use of everyone's time and resources:

1. Observe
2. Make a hypothesis
3. Test
4. Reject the hypothesis
5. Form a new hypothesis

The reason why the former approach rather than the latter is often employed is, of course, because policymakers and the public are acting like children and do not pass the marshmallow test (Mischel, 2014). The marshmallow test offers the child a choice between having one marshmallow right away or two marshmallows one hour from now. Children who are patient enough to wait one hour to get two marshmallows grow into more successful adults than children who opt for one marshmallow right away (Mischel, 2014).

It turns out that policymakers who cannot pass the marshmallow test lead us often to failure when it comes to education reform or reform of any kind (Bryk et al., 2015; Zamarro et al., 2015). This is not the case only in education but also in health care (Kenney, 2005) or the area of providing economic aid to less developed countries (Easterly, 2017; Moyo, 2009). The latter approach takes more time and resources in the planning stage and delivers results. The former approach takes less time and resources in the planning stage and delivers a failure.

The government and private foundations often justify the impatience by referring to the private sector. But the private sector has the benefit of competition and bankruptcy, the survival of the fittest process that selects for the best implementers. The best companies are often not those that have the process but the companies that offer the right product under the right circumstances, and this often happens by accident rather than by deliberate design. Since the public sector doesn't have a bankruptcy process or competition to aid them (and neither have the private foundations), they need to plan better than commercial organizations. While the private sector rides the wave of "move quickly and break things," the public sector doesn't have this luxury. If a private company comes up with a terrible product, it simply goes under and the founders create a new company. A government entity does not have the same opportunities for reinventing themselves. Then they need to plan better and fail less or at a smaller scale or incorporate failure more explicitly into their production process. They need to test their products more thoroughly prior to rollout or implementation. This is decidedly not the case either in K-12 or in pre-K (Little et al., 2019).

## Using the IDM in Line with the "patient Approach"

### *Observe*

Step 1 in the patient approach is to observe. This could involve a literature review of existing research or a survey of best practices by practitioners in the field or policymakers with executive experience. This is what the Gates Foundation has done in coming up with the 15 Essential Elements. This is also what drives the entire IDM process, the careful collection of existing evidence and categorizing into the main aspects of pre-K provision. Every single indicator on the IDM is backed up by evidence either in the form of research journal articles or recommendations by experienced policymakers.

### *Make a hypothesis*

The whole IDM is a hypothesis, as well as every single indicator, and this is how it should be used in the field. As Bryk et al. (2015) observe, the field of education needs much more research and the field of pre-K as well. Every single indicator on the IDM is a hypothesis about how progression of improvement takes place. Bryk et al. (2015) advocate the use of Network Improvement Communities to "learn to improve." What this means in practice with the IDM is that ideally every single state would have their own research partners that would help verify the efficacy of every single IDM indicator and its impact on child outcomes in a given state under given local circumstances. This would be done through rigorous methodological procedures that involve experimental or quasi-experimental methods. IDM represents all we know as yet about pre-K; however, it is not the final word on the subject, since the research in early education is ongoing. Every state needs to take this and "run with it," continue the research, and continue to refine what the IDM has to say. Let us take for example Indicator 4 on High-Quality Teaching. The four steps and their definition imply that equal improvement is achieved with each step. This is our best opinion based on the literature and opinion of experts. However, this needs not to be the case in every single state and in every single setting. States are encouraged to follow the implied recommendations of the IDM in practice while simultaneously researching them and refining them. Some states may, for example, find that a cut-off of 5.5 on the CLASS emotional support rubric separates two meaningful levels of quality rather than 5.0. This implies a need for massive increase in research funding and if need be, for states to cooperate to lower this research burden. This is why Bryk et al. (2015) encourage the creation of network improvement communities.

### *Test*

Once a hypothesis is formed, it is time to test. The best approach to testing a hypothesis is conducting an experiment. There are ethical considerations to be kept in mind when conducting an experiment; however, the health care field is conducting them very often, and they manage to navigate the ethical quandary. When experiments are not possible, quasi-experimental approaches are the next best thing. A well-designed experiment may tell us not only whether some approach or idea works, but also why or why not. As is often the case, the devil is in the details—the implementation is as important as the idea. In the same way that sometimes good ideas for education reform die because of poor execution, sometimes experiments are seen as not useful because they are not designed or executed well. A well-designed experiment can incorporate specific context as well as general hypotheses (Hedges, 2018; Zhang, 2014).

To continue with the example above, the IDM tells us that CLASS Emotional Support of 5 is an important cut-off, generally speaking. Is that the case for family child care, center-based child care, etc., or just some of them? How is this all related to structural aspects of pre-K experience? The classroom size? Is this the case for language and ethnic minorities? None of these questions have been asked in the literature. They need to be asked and researched, and if the academic community fails to take them up, it is up to the state agencies to do so, for the sake of the children in their own states. While a cut-off of 5 is our best guess as of right now for all groups that the state can put into practice immediately, a state would be wise to continue to research the impact on various subgroups and flesh out this kernel of knowledge in more detail. This can be done using experiments or quasi-experimental approaches.

Another example of an experiment in the education setting is Fryer et al. (2015), who paid students to improve their grades. The students did not improve and the idea failed, but at least money was not wasted in statewide implementation of such policy. Unfortunately, this particular study did not explore the reason for such failure, but a follow-up study would have been able to do so (one may suspect crowding out of intrinsic motivation as we indicated in the introduction). This would be in line with the approach of reject-propose a new hypothesis. But as Bryk et al. (2015) point out, often academics are not incentivized to conduct such iterative continuous quality improvement types of research. Follow-ups after a failed finding are rare in academia.

*Reject the hypothesis*

Once a rigorous research study is done on a reform idea under consideration and the data is collected and analyzed, we obtain a conclusion regarding whether the idea works and under what conditions. We may learn that CLASS ES of 5.0 is a valid threshold for Spanish-speaking children but not for rural children, etc. This prevents the policymaker from frivolously introducing a uniform rule when specific rules are called for. All these are called moderator variables, and often policy failures are made because moderator variables fail to be considered. It may so happen that an idea (intervention) that worked in a certain environment with a specific value of the moderator is scaled to a wider set of contexts with differing levels of this moderator variable. In such a scenario, the policy may fail as it is introduced into environments with different values of such moderator variables, for which efficacy has not been investigated. A well-designed research process carefully tests for whether a given hypothesis works for different values of such moderator variables. These moderator variables may include demographics, school setting, geographic context, etc. Business investors know that 90% of ideas fail, so oftentimes it is a matter of rejecting it as quickly as possible (after conducting the relevant research) and moving on to the next one. Being able to reject ideas is very important. The problem with failed educational reform initiatives is that often we do not even know (in terms of solid evidence) why they failed (speculation is, of course, always available), which prevents us from proceeding to the next logical step in the process. If we know why an idea failed or are confident it does not work, we can go back to the drawing board and continue to explore an alternative. In such a manner, instead of stumbling in the dark, we proceed according to a rational plan that will sooner or later yield results.

*Do not change anything unless there is an alternative with solid evidence supporting its superiority over status quo*

The implication of the approach we have outlined here is that no action should be taken unless this is part of the process outlined above. If leaders implement arbitrary reform ideas that are not sufficiently researched, they are running a serious risk of failure. This is a fact born out again and again, and Bryk et al. (2015) take us through the graveyard of well-intentioned under-planned, under-researched ideas for reform that anyone who is feeling lucky should read.

*Do not equate "bad research" with "research"*

Many studies do not follow best practices in terms of methodology. Someone may conduct a study using a non-representative sample of children (middle class for example) and then present the conclusions of the study as being generally applicable. This is decidedly not the case. The study only applies to children from similar circumstances as the children recruited for the study. We should not judge the whole field of science by a few bad apples; this is a human bias that we need to put some effort into avoiding (Tversky & Kahneman, 1974). High-quality research is the only guarantee of success of reform implementation. Research should be involved through implementation, in planning as well as execution.

*Start small and work your way up*

IDM, by the virtue of being divided into Elements and indicators, also supports another important notion in the CQI process that could be summarized as "start small and work your way up." Complex systems are hard to study because they have many potentially interacting components. In their study, it's important to have the ability to "hold other things constant," tweak one component and observe an effect on the whole system. By making a complete survey of preexisting knowledge, the IDM allows states to put best practices into action and then pick one or two indicators and study the effect of tweaking them. Implementing a complex system all at once also creates undue burden and implementation challenges on the teachers and instructional leaders (Braaten et al., 2017).

## Acknowledgement

This work was created as part of the Partnership for Pre-K Improvement, a collaboration between the Alliance for Early Success, Cultivate Learning, and Start Early, with support from the Bill & Melinda Gates Foundation.