A McREL Report Prepared for

Stupski Foundation's Learning System

Systems Diagnostics

About McREL

Mid-continent Research for Education and Learning (McREL) is a nationally recognized, private, nonprofit organization dedicated to improving education for all students through applied research, product development, and service. Established in 1966, McREL now maintains a staff of around 110 in its Denver, Colorado, office.

This report is part of a larger set of reports prepared by McREL for the Stupski Foundation. The views, findings, conclusions, and recommendations expressed herein are those of the authors and do not necessarily express the viewpoint of the Foundation. Please e-mail any inquiries to Linda Brannan at info@mcrel.org.

Contact

McREL
4601 DTC Blvd., Ste. 500
Denver, CO 80237
Web site: www.mcrel.org • E-mail: info@mcrel.org
P: 303.337.0990 • F:303.337.3005

Acknowledgements

Many people contributed their time and expertise to this report. The authors express their gratitude to the following McREL staff members: Maura McGrath and Linda Brannan, who lent their library expertise to identifying and collecting the literature reviewed; Lisa Maxfield, who diligently supported our efforts to catalog and record the articles collected; Vicki Urquhart, who provided her keen editorial oversight; and Natalie Voltes, who is responsible for the striking design and layout of this report. In addition, we thank Jean Williams and Helen Apthorp for offering their expertise on research methodology, John Kendall for serving as a reviewer of the final report, and Jackie Kearns for her tireless project management efforts. We also acknowledge the support, advice, and constructive dialogue provided by staff members of the Stupski Foundation.

Authors

Charles Igel, M.A.
Helen Apthorp, Ph.D.
Gary Peterson, M.S.
Tony Davis, Ph.D.
Laurie Moore, M.P.H.
Kerry Englert, Ph.D.

20091030

Table of Contents

Executive Summary 1
Introduction 5
Findings 11
Discussion & Recommendations 15
Final Thoughts 29
References 31
References Reviewed but Not Cited in This Report 35
Appendix 37

Executive Summary

This document is one of eight reports prepared to support the development of a new learning system, an effort that is the first step in a major initiative undertaken by the Stupski Foundation. The Foundation endeavors to improve the life options of all students, especially underserved urban youth of color, whom we refer to as "Our Kids," by fundamentally redesigning the education system.

This report was created collaboratively by researchers from McREL with guidance from officers of the Stupski Foundation. Its purpose is to provide members of a "Design Collaborative" team—consisting of practitioners, parents, students, and researchers—with a review of key findings from existing literature to support their efforts to develop the systems diagnostics component of the Stupski Foundation's Learning System.

Research methodology

McREL researchers, in collaboration with Stupski Foundation staff members, generated the following research questions to guide this review:

1. To what extent have states/districts established systems diagnostics?
2. What are the components and characteristics of the ideal systems diagnostics tool?
3. Which diagnostic models from other fields have relevance to education?

While McREL researchers concentrated on these three questions to guide the systems diagnostics literature review, they did so always with an eye toward what worked for students of color and poverty. The discussion section addresses specifically how the findings inform decisions for this target population.

These questions focused an extensive review of scholarly (i.e., peer-reviewed publications) and "fugitive" literature (i.e., reports self-published by reputable foundations, associations, and other organizations). In all, the research team reviewed 80 articles and summarized 46 of these. Data and conclusions from these reports have been synthesized into a key finding: a framework for educational systems diagnostics based on elements of High Reliability Organizations (HROs).

Key findings

An assumption of this report is that diagnostics organized within the High Reliability Organizational model provide a more effective means for ensuring the success of Our Kids than the current scattershot approach utilized by most schools systems. The necessary diagnostic pieces are in place within many schools systems; they are just not organized in a coherent and complementary manner.

HROs are defined by several common features:

- Preoccupation with the operational breakdowns that lead to systemic failure
- Reluctance to simplify operations and interpretations of lessons learned
- Sensitivity to the importance of ground-level decision making
- Commitment to organizational resilience in the face of systemic failures
- Deference to situational expertise over hierarchy or rank (Weick & Sutcliffe, 2001)

As it applies to this report, HROs are organized around four simultaneous and overlapping levels: proactive operations, responsive monitoring, adaptive interventions, and ongoing feedback. The following options are elements within each of these phases.

Recommendations

The options presented here represent four areas for improving educational opportunities for Our Kids by leveraging promising practices, supported by the literature, at a systems level.

Option 1: Actively recruit good teachers & support their professional development

Effective teachers are among a district's most important assests, and successful systems establish mechanisms for recruiting the right people. To do so, most successful school systems use a more proactive approach than the conventional method of recruitment. For example, schools in Singapore proactively screen candidates for a limited number of positions within their teacher preparation programs. Matriculates become employees of the Ministry of Education, receive a salary, and are placed into teaching positions upon graduation. This practice of screening candidates before they begin training is also exhibited in fellows and residency programs in U.S. cities such as Chicago, Boston, and New York, albeit on a much smaller scale. Once good teachers are recruited, they must be developed into better teachers.

Option 2: Build a cohesive culture focused on data-driven decision making with defined site-level responsibility

While there is little empirical literature on the culture of education in terms of cohesiveness and data-driven decision making, an examination of common problems facing large urban districts laid the foundation for our recommendations. District bureaucracies should be as small as possible, and efficiency should be the coin of the realm. Alignment between district policies and site-level practice is the key. To facilitate the effective use of data and reduce resistance to new ideas, school-level personnel should be included in planning sessions on data use. This provides needed information to teachers about the proper use of data, and it also allows district-level personnel to better understand school-level needs.

Option 3: Develop & utilize longitudinal data collection systems for critical areas

The appropriate and timely use of data is a challenge, yet both dropout and academic monitoring is vitally important to the success of Our Kids. Accurately tracking dropout rates is difficult due to high mobility rates of students and unreliable data sources. In addition to more accurate dropout tracking, states, districts, and schools should focus on preventative measures. There are a number of successful and targeted programs from which to choose (e.g., Achievement for Latinos through Academic Success, Talent Search, Check & Connect, Accelerated Middle Schools).

For tracking academic performance, most states use cross-cohort models. With these models, student performance is measured against a set of fixed standards. However, many states are moving toward gain or growth models that adjust for a student's prior experience by substituting or statistically adjusting for entering achievement levels (Ladd & Walsh, 2002). Because each model has its strengths and weaknesses, a combination of cross-cohort and growth models is recommended.

Option 4:
Enact timely & evidence-based interventions where systems have failed

When schools or districts fail, timely and evidence-based interventions are needed to deal with problems before they become endemic. A review of successful turnarounds across different organizational types reveals several common features (Hassel & Hassel, 2009). First, successful turnaround campaigns focused on a "few early wins." Rather than attempting to reform the entire system, interventions focused on a limited number of high-priority goals with equally high visibility. Secondly, effective turnarounds purge operating norms that stand in the way of high performance. Finally, get the right people into key leadership positions quickly. All turnaround campaigns face resistance—some from individuals who hold differing opinions about new policies but genuinely work for best interest of the organization, and others who are simply protecting turf—and having leadership in a position to help win over the former and eliminate the latter is crucial to the success of an intervention.

Final thoughts

Within urban districts that serve Our Kids, many of the factors needed for HRO-modeled systems diagnostics are already in place. Toward this point, some of the recommendations are probably not new to the reader who is familiar with urban education. What this report offers is a *comprehensive* framework for systems diagnostics organized around the proven HRO model. The authors make no claim to the exclusivity of the HRO model as a comprehensive tool; however, the current scattershot approach has not worked. The purpose of this report is to begin conversation around a comprehensive approach to systems diagnostics rather than to provide a final answer.

Introduction

Purpose of this document

This document is one of eight reports prepared to support the development of a new learning system, a development effort that is the first step in a major initiative undertaken by the Stupski Foundation. The Foundation endeavors to improve the life options of all students, especially underserved urban youth of color, whom we refer to as "Our Kids," by fundamentally redesigning the education system.

The report was created collaboratively by researchers from McREL and officers of the Stupski Foundation. Its purpose is to provide members of the Design Collaborative team with a review of key findings from the existing literature regarding critical research questions related to the systems diagnostics component of the Learning System and to offer recommendations for the development of this component. Together, the reports cover these topics:

- Assessment
- Curriculum
- Pedagogy
- Student Supports
- Systems Diagnostics
- Leadership
- College Readiness
- Our Kids

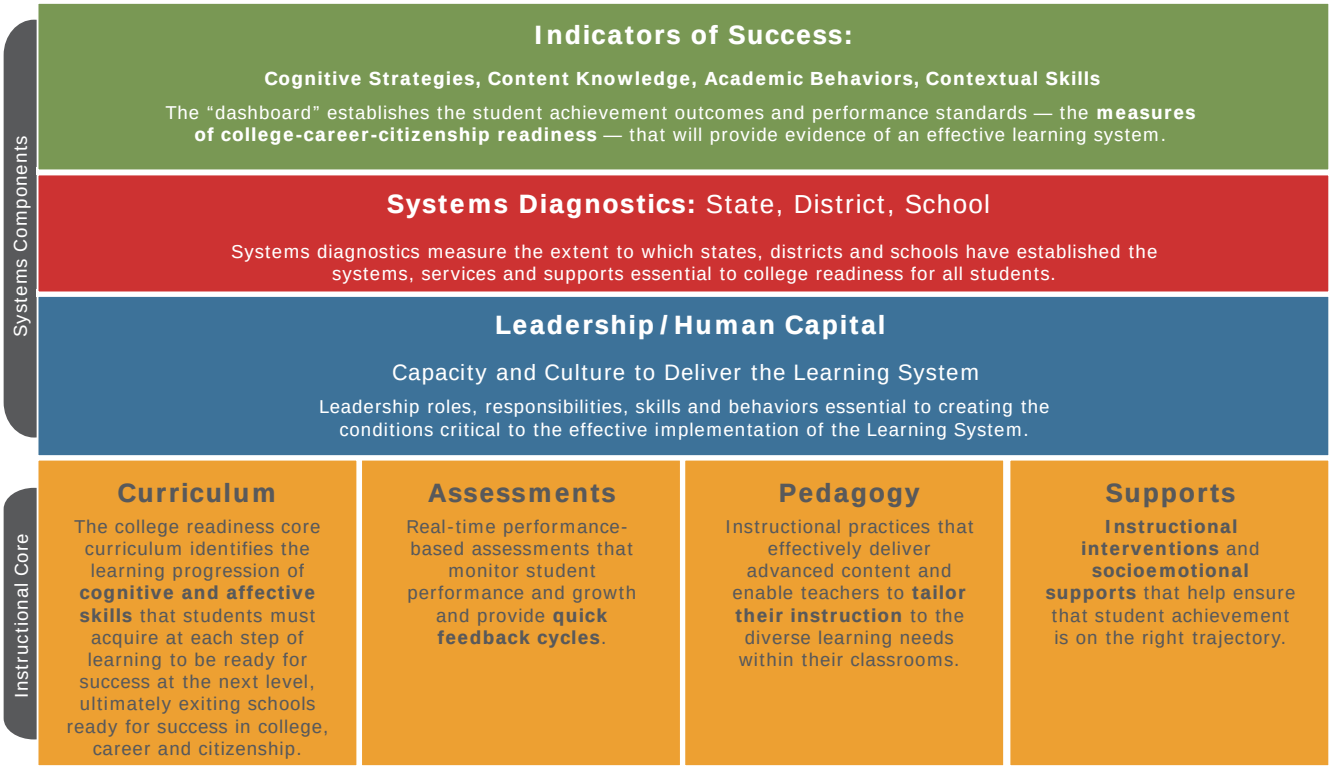
The first section of this report provides salient findings that emerged from the literature review. The second section offers a discussion of the findings along with several recommendations—framed as four key options—for how the Design Collaborative might proceed. A brief concluding discussion follows. Summaries of the studies and literature reviewed for this report are provided as separate documents.

About the Learning System

The Learning System is the product of the Stupski Foundation's extensive examination of research, best practices, and theories of action for improving education opportunities for all children. It is deeply rooted in the Foundation's mission to foster innovation in public school systems so that all students graduate ready for college, career, and success—as well as the notion that the United States' education system, in its current state, is incapable of accomplishing this goal. As stated on the Foundation's Web site, "The basic components of what public education systems need to teach all students to world-class standards, particularly those students for whom public schools are their only option, do not exist in any coherent, accessible or evidence-based way" (Stupski Foundation, n.d.).

Thus, the Foundation has focused its philanthropic efforts on supporting the "fundamental reinvention" of the American system of public education into one that prepares all children for the challenges of life, career, and citizenship in the 21st century. To accomplish this objective, the Foundation launched a multi-year, cross-sector collaboration among researchers and practitioners from inside and outside education to develop a new and comprehensive Learning System. In its June 2008 *Strategy and Program Overview*, the Foundation posited that this system includes seven components, shown in Figure 1 (see p. 6). The indicators of success are dependent on a definition of college readiness, which is addressed in the respective report. Although Our Kids is not an explicit component of the learning system, it is the foundation for the work the Stupski Foundation is committed to in the education sector. As such, the populations of students of color and students of poverty warranted a separate report.

Figure 1. The Learning System



About "Our Kids"

The Stupski Foundation is committed to addressing the academic needs of underserved populations, in particular, students who are of color *and* in poverty (which comprises 42% of African American students and 37% of Hispanic students) (Duncan & Magnuson, 2005). Despite a dramatic rise in minorities enrolling in college (a 50 % increase from 1995–2005), fewer minorities appear to be graduating. As shown in Figure 2 (see p. 7), in 2006, fewer minorities aged 25–29 reported having obtained an associate degree or higher than their older peers (aged 30 and over) (American Council on Education, 2008). This trend marks an important reversal in advances in educational opportunities for minorities and may mark the first time in history that a generation of students has demonstrated less educational attainment than its predecessors (American Council on Education, 2008).

Overview of methodology

McREL researchers followed a five-step process for translating findings into recommendations.

Step 1: Key assumption

After conducting an initial survey of relevant literature, Stupski Foundation staff members identified the following assumption to guide the literature review for the curriculum component:

U.S. public education needs more effective and real-time diagnostics that monitor and adjust the educational systems serving Our Kids.

Step 2: Identification of research questions

McREL researchers, in collaboration with Stupski Foundation staff members, generated these questions:

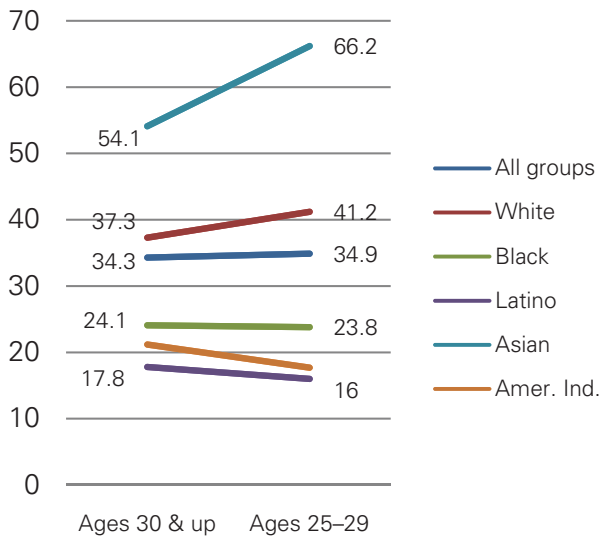
1. To what extent have states/districts established systems diagnostics?
2. What are the components and characteristics of the ideal systems diagnostics tool?
3. Which diagnostic models from other fields have relevance to education?

While McREL researchers concentrated on these three questions to guide the curriculum literature review, they did so always with an eye toward what worked for students of color and poverty. The discussion section addresses specifically how the findings inform decisions for this target population.

Step 3: Literature search

The three research questions guided a search for literature in several journal databases (e.g., Academic Search Premier, JSTOR, ERIC, Proquest, Academic Onefile, Educators Reference Complete), sites funded by the U.S. Department of Education (e.g., ERIC, What Works Clearinghouse, Doing What Works, National Laboratory Network, and those of national comprehensive centers and national education research centers), and other sources, including Google Scholar and Educational Policy Analysis

Figure 2: Percentage of U.S. adults with associates degree or higher, 2006



Source: American Council on Education, 2008

Archives. In addition, the Table of Contents of certain journals and the following Web sites were systematically reviewed because of their apparent relevance to the search topic: Education Resources Information Center (ERIC), Center for the Education and Study of Diverse Populations, Assessment and Accountability Comprehensive Center (AACC), Fordham Institute Center for Education Policy Research (CEPR), National Bureau of Economic Research (NBER), Education Commission of the States (ECS), National Center for Evaluation, Standards and Student Testing (CRESST), Center on Innovation and Improvement, EdSource, Council of Chief State School Officers (CCSSO), *EdWeek*, Southern Regional Educational Board (SREB), and *Sage Journals Online*. Sources were searched by the following keywords:

- Academic achievement
- Accountability
- Accountability and metrics
- African American students

- Assessment
- College entrance examinations
- College preparation
- College readiness
- Diagnostic indicators
- District improvement
- Effective schools research
- Ethnicity
- High reliability
- High reliability organizations
- High-stakes tests
- Minority groups
- Organizational development
- Organizational effectiveness
- Organizational models
- Our Kids
- Program evaluation in urban schools
- Race
- Restructuring
- School district accountability
- School district effectiveness
- School districts
- School effectiveness
- School improvement
- State accountability
- System diagnostics
- Turnarounds-district role
- Urban schools

The research team also examined reference lists of articles identified in the first scan to find other applicable studies for all the components of the Learning System. They identified 80 articles related to systems diagnostics. All identified articles were retrieved and reviewed with attention to research methods, outcomes, and recommendations for future study. Ultimately, the team summarized 46 articles related to systems diagnostics, which are in a separate annotated bibliography.

Step 4: Identification and cataloging of findings

The research team cataloged findings from the summarized articles using the following identifications:

Counterproductive *orthodoxies* (conventional ways of providing education which may be impeding student success)

Unmet needs (areas where students are not yet well served by the current system of education)

Next practices (a program or practice that needs to be developed, adapted, invented, and tested in response to an unmet need)

Promising practices (practices based on research but not supported by rigorous efficacy data)

Current *best practices* (practices demonstrated by research to be effective in improving outcomes for students)

Step 5: Generation of recommendations

In the final phase, research team members collectively reviewed key findings from the literature review in light of the following questions:

1. To what extent have states/districts established systems diagnostics?
2. What are the components and characteristics of the ideal systems diagnostics tool?
3. Which diagnostic models from other fields have relevance to education?

Responses to these questions were synthesized into recommendations, presented here as options for further action. These options include best or promising practices that should be *adopted* and scaled up or *adapted* to new settings or areas where there are gaps in practices that require new innovations to be *invented*.

Overview of the literature base examined

For this report, the research team began with the definition of diagnostics as the mechanisms that measure the effectiveness of a system at performing its intended task. Due to the nature of this topic, there is no definitive list of necessary components for educational diagnostic systems, nor is there an abundance of high-quality empirical research on the effectiveness of existing components. One can imagine the impossibility of randomly assigning various diagnostic components across a sample of districts to test their effects. Therefore, the majority of literature on this topic is descriptive or conceptual, often highlighting a particular tool used by a district or recommending a comprehensive set of tools that systems might adopt to assuage a given problem. In other words, the research team had few specific studies upon which to structure its recommendations. However, as the literature search was broadened into areas outside the field of education, patterns around the topic of systems diagnostics began to emerge.

Recommendations from the report are organized around a central hub of High Reliability Organizations (HRO). The HRO framework is prevalent in certain fields such as medicine; however, it is little discussed in education. One exception is the writings of G.T. Bellamy and a small number of colleagues. This small corpus provides the impetus for our recommendations by providing a structure for systems that cannot risk failure. From this hub, the research team was able to identify educational literature on critical diagnostic leverage points: high-quality staffing, data-driven decision making and defined site-level responsibility, longitudinal tracking systems, and timely and evidence-based interventions.

Findings

This report summarizes the literature around the hypothesis that academic success among Our Kids would improve if pre-K–12 education adopted the systems diagnostic approach utilized by High Reliability Organizations. There is little empirical work in the area of systems diagnostics, with the vast majority of the literature being conceptual or theoretical in nature. Consequently, randomized controlled trials of assessment systems or drop-out prevention programs are unlikely to exist, so formally testing this hypothesis is impossible at this time. Instead, the purpose here to describe the elements of HROs and offer a framework that applies those elements to educational systems diagnostics.

The High Reliability Organization model

Systems need diagnostic monitoring in order to sustain effectiveness. Jets negotiate airspace and land safely in crowded airports because air traffic controllers are able to make decisions based on real-time data. Nuclear power facilities within modernized countries are able to operate around populated areas without fear of accident due to vigilant monitoring systems and strict operating protocols that minimize risks of failure. Surgeons perform seemingly impossible life-saving operations on a routine basis without complications by using tested procedures imbued with overlapping layers of responsibility. In short, systems diagnostics are critical to monitoring and maintaining any effective system. This report examines the extent to which diagnostics are used throughout pre-K–12 education to monitor the systems that ensure the academic success and college, career, and citizenship readiness of students at-risk of failure—Our Kids.

There is little empirical work dealing directly with the efficacy of specific diagnostics systems. The literature primarily contains conceptual or theoretical work with descriptions of developing or existing programs. Rarely is the effectiveness of these programs directly measured. This report synthesizes the current thinking from academic leaders in the field of systems diagnostics as well as existing and emerging promising practices from the field. To frame our argument, the research team also looked outside of education for corollaries from other fields. Much has been made of the parallels between medical practice and education; this report builds on these parallels with examples of diagnostic medical models that could easily be transferred to education.

Framing this entire work is the High Reliability Organization framework. HROs operate in fields where systemic failure has catastrophic consequences. We argue that the HRO framework is appropriate and useful in initiating a discussion of systems diagnostics within the field of education. This work is based on the assumption that the U.S. public education system needs more effective and real-time diagnostics that monitor and adjust the educational systems serving all students, particularly Our Kids. Low-income, urban children face particularly steep challenges and often lack familial and community-based safety nets. For many of these children, the education system has become their last line of defense against academic failure. This report is based on the assumption that effective diagnostic tools could improve the systems that ensure Our Kids succeed in school and graduate ready for college or the workplace.

The proposed diagnostic system is scaled across multiple overlapping levels of operation. The three we highlight are school, district, and state; however, a comprehensive system should include classroom and federal levels of responsibility, as well. The authors of this report choose to focus on the three *core* levels as a choice of depth over breadth. Underpinning all levels are universal

operating principles of operational mindfulness, proactive operations as opposed to reactive operations, early identification of problems, and adaptive interventions that employ an all-hands-on-deck approach to problem solving.

Before continuing, the reader should be aware this report is designed to serve as a heuristic for conversations around effective systems diagnostics targeted toward the needs of Our Kids. The recommendations and options we describe are not the only areas where practice can be improved, nor are they comprehensive in scope. Rather, the purpose of this report is to provide readers with a seed from which critical conversations may begin.

Systems focused on preventing failure

Organizations operate within a certain context, and these contexts indelibly shape the organization itself. In *Better: A Surgeon's Notes on Performance* (2007), Dr. Atul Gawande describes how hospitals have shaped their operations based on the context in which they operate. A central focus, he points out, is a continual focus on the prevention of failure. Toward this end, hospitals have developed overlapping operational protocols to lessen the possibility of mistakes. This is evident in Gawande's description of a typical surgery preparation to remove a cancerous growth.

> The operation was not going to be difficult or especially hazardous, but the team had to be meticulous about every step. On the day of surgery, before bringing her to the operating room, the anesthesiologist double-checked that it was safe to proceed. She reviewed [the patient's] medical history and medications, looked at her labs in the computer and at her EKG. She made sure that the patient had not had anything to eat for at least six hours and had her open her mouth to note any loose teeth that could fall out or dentures that should be removed. A nurse checked the patient's name band to make sure we had the right person; verified her drug allergies with her, confirmed that the procedure listed on her consent form was the one she expected. The nurse also looked for contact lenses that shouldn't be left in and for jewelry that could constrict a finger or snag on something. I made a mark with a felt-tip ten over the precise spot where [the patient] felt the lump, so there would be no mistaking the correct location. Early that morning before her surgery, [the patient] had also had a small amount of radioactive tracer injected near her breast lump, in preparation for the sentinel lymph node biopsy. (p. 4)

This description of the pre-surgery ritual continues, but this excerpt makes the point. Systems, some as simple as verifying the patient's identity and others as complex as the injection of a radioactive tracer, were established to reduce the risk of failure. And, where human error would be catastrophic, hospitals have developed systems with overlapping levels of responsibility to ensure that any mistake that might slip through one level is caught by another. For example, Gawande physically marked the lump with a pen and

orally confirmed the nature of the operation with the patient even though the radioactive tracer was injected to indicate the location of the tumor. This does not imply that these systems are foolproof; Gawande writes extensively about the struggles to overcome both human and systemic failures. Rather, he illustrates through multiple examples how the organizational culture obsesses over failure, and in doing so, continuously seeks ways to avoid it. Although he does not identify them as such, Gawande is describing hospitals as High Reliability Organizations.

Operating norms such as these have relevance within the field of education. Heretofore, schools have not been widely viewed through the HRO lens. Academic failure is much less dramatic than a plane crash, nuclear meltdown, or death of a patient; and the negative consequences of a mistake or missed opportunity are not as immediate. However, the consequences of failure for a child are just as profound.

Elements of the High Reliability Organizations model

Often, diagnostics are viewed as passive data collection. A doctor orders a CT scan for a patient to assist in his diagnosis; a pilot checks instruments to determine the state of an aircraft before taking off. This type of monitoring is an important component of diagnostics, providing us with knowledge about the current health of the system. Unfortunately, when diagnostic systems are confined to this narrow definition, they are not as effective as they could be because they lack proactive and adaptive elements. It is important to track dropout rates and academic achievement; however, it is equally important to establish procedures that obviate system failures before they arise and improve operations based on lessons learned from failures. The HRO model provides a broader definition of systems diagnostics that contain these critical missing components (see Bellamy et al., 2005; La Porte, 1996).

High Reliability Organizations are characterized by what Hoy, Gage, and Tarter (2006) refer to as organizational mindfulness: "Mindfulness is ongoing scrutiny of existing experiences, appreciation of the subtleties of context, and identification of novel aspects of context that can improve foresight and functioning" (p. 308). Among high functioning organizations, mindfulness is the foundation of the culture itself. From the central node of mindfulness, radiates an array of related characteristics. Weick and Sutcliffe (2001) propose the most frequently cited characteristics of HROs:

- Preoccupation with the operational breakdowns that lead to systemic failure
- Reluctance to simplify operations and interpretations of lessons learned
- Sensitivity to the importance of ground-level decision making
- Commitment to organizational resilience in the face of systemic failures
- Deference to situational expertise over hierarchy or rank

Although these features typically are present in successful organizations, they are not necessarily a recipe for success. The vicissitudes of an effective organization are far too complex. Indeed, there are times in which one or more of these features might impede organizational effectiveness. What these represent are the successful features of HROs, under normal operations, that are relevant to an expanded conceptualization of systems diagnostics. From this position, the authors of this report hypothesize that the adoption of a comprehensive systems diagnostic approach utilized by HROs would improve academic achievement, retention rates, and college readiness among Our Kids.

Systems diagnostics within the HRO framework operate on three simultaneously occurring levels: proactive operations, responsive monitoring,

and adaptive intervention. Central to each is the identification of critical inputs and outputs into the system. Examples of inputs are the curriculum, classroom instruction, and teacher recruitment. Examples of outputs are achievement levels, retention rates, and operational efficiency. An important design element of any diagnostic system is identification of these elements and development of valid and reliable instruments for their assessment.

Proactive operations. Proactive operations refer to normal operating procedures established to eliminate risks of failure before they arise. Central to the development of proactive operations is an organizational culture focused on identification of failure points and continual self-improvement. The underlying premise is that the extensive use of proven (or promising) practices aimed at improving the academic and social supports for Our Kids will avert many problems before they arise. Establishing policies to ensure classroom teachers have the necessary skills to work with urban populations is an example of this concept.

Responsive monitoring. Responsive monitoring is considered the traditional purview of diagnostic systems. It is the more passive monitoring phase that involves data collection and analysis for decision making. However, this model expands the use of assessments for formative rather than summative purposes and incorporates data collection for the purpose of dropout prevention.

Adaptive intervention. Adaptive intervention involves how organizations respond to problems. High Reliability Organizations understand that systems do occasionally fail and problems slip through. Once detected, it is incumbent on the organization to react quickly and rectify the problem. A school, for example, that experiences a high truancy rate might receive stricter oversight from an intervention committee consisting of school personnel, district leadership, community leaders, and content experts—an all-hands-on-deck approach that immediately brings additional resources to bear on nascent problems before the ineffective practices that allowed the problem to arise become entrenched.

Moreover, interventions should serve as action research with the most effective treatments informing the system's core operations. In this way, the entire diagnostics operation functions as a learning system within a continual feedback loop that incorporates effective innovations into core operations. Bellamy and colleagues (2005) refer to this as the "Swiss cheese" model of HROs. As they explain, any single layer contains holes through which problems might slip. However, when additional layers are added, it becomes increasingly unlikely that the holes within all the layers will align. A failure at one level will likely be caught at another level. Rather than depending on a single diagnostic layer, the HRO model of systems diagnostics depends on multiple simultaneously occurring layers that avert, monitor, and rectify systemic failures.

Discussion & Recommendations

Contained within the three diagnostic layers of the HRO Framework are four areas for improvement for education systems diagnostics (see Table 1). Each area describes an option for improving educational opportunities for Our Kids by leveraging promising practices at a systems level. To maintain cohesiveness, each option is organized around a seminal work (or works) on the topic with supporting literature supplementing content where necessary.

Although this report focuses on systems diagnostics targeted at the performance of Our Kids, much of the literature from which the recommendations were developed is generic toward student demographics. The McREL research team used its best professional judgment

Table 1. High Reliability Organization model applied to educational systems diagnostics

High Reliability Organization Approach to Systems Diagnostics

Proactive Operations →	Responsive Monitoring →	Adaptive Intervention
Focus on Improvement of Core Operations	**Continual Monitoring of System Performance**	**Timely Interventions where Systems Have Failed**
1) Actively recruit good teachers & support their professional development. 2) Build a cohesive culture focused on data-driven decision making with defined site-level responsibility.	3) Develop & utilize longitudinal data collection systems for the critical areas of dropout prevention & academic performance.	4) Enact timely and evidenced-based interventions where systems have failed.

← Feedback Loop for Adaptation of Core Operations

to ensure that any recommendations applied to an at-risk population. Additionally, we remind the reader that these are areas within an expanded definition of systems diagnostics that includes proactive and adaptive components. Not all will be oriented toward measurement; however, all are recommended for inclusion in a comprehensive diagnostic system.

Option 1: Actively recruit good teachers & support their professional development

Effective teachers are among a district's most important assets. According to some educational researchers, staffing schools with the right personnel is the most important thing a school system can do to improve achievement. Studies that examine the impact various education inputs have on student achievement consistently find that teachers have the largest impact even over important variables, such as per pupil expenditure or class size (Nye, Konstantopoulos, & Hedges, 2004; Rowan, Correnti, & Miller, 2002; Sanders & Rivers, 1996; Snipes & Casserly, 2004). However, all teachers are not equally effective. Conventional estimates propose that one standard deviation in teacher quality accounts for approximately 0.1 standard deviations in student achievement (Rivkin, Hanusek, & Kain, 2005; Rockoff, 2004). To put this in perspective, class size for a typical 5th-grade classroom would have to be reduced by approximately 13 students to obtain the effect of moving from an ineffective to an effective teacher (Rivkin et al., 2005).

Similarly, a report by McKinsey & Company (2007) concluded that high-performing school systems consistently recruit high-quality individuals and develop these people into effective instructors. Their report contains analysis of the best performing school systems as identified by the OECD's Programme for International Student Assessment (PISA); a survey of current literature; and interviews with more than 100 experts, policymakers, and practitioners. School systems were located in regions as diverse as Alberta (Canada), Australia, Belgium, Finland, Hong Kong, Japan, the Netherlands, New Zealand, Ontario (Canada), Singapore, and South Korea. Seven additional systems that showed "strong improvement trajectories" were also included in the study. This sample represents myriad cultures, educational systems, policies, and geographic regions; yet, despite these wide cultural and geographic differences, teacher quality remained the paramount educational input.

Potential benefits and drawbacks of this option

Due to its comprehensive nature, the McKinsey report serves as the seminal work for our examination of this option. So how exactly did the top-performing school systems in this study ensure they have the highest quality teachers?

Recruiting talented candidates. Most school systems in the report broke with the conventional method of recruitment and instead used a more proactive approach. Traditional recruiting techniques involve selecting candidates from a graduate pool that have already completed (or nearly completed) training. However, McKinsey & Company (2007) argue that inefficiencies in this system lead to oversupplies of labor that dissuade high performers from even entering the field. This creates an iterative downward spiral that continually lowers the status of the profession.

Top-performing systems, on the other hand, select candidates into the training program itself, matching the number of slots with an anticipated number of job openings. For example, Singapore proactively screens candidates for a limited number of positions within its teacher preparation programs. Matriculates become employees of the Ministry of Education, receive a salary, and receive teaching positions upon graduation. This practice of screening candidates before they begin training also is practiced in fellows and residency programs in United States cities such as Chicago, Boston, and New York, albeit on a much smaller scale.

These alternative pathways into the profession provide an option to policymakers who have limited ability to reshape a system's recruitment process. Alternative entry routes are not simply a means for supplementing the teaching force; recruiting talented candidates is a first step in increasing the status of the profession as a whole. As the authors explain, "In all school systems there are powerful feedback loops associated with the status of the teaching profession. Once teaching became a high-status profession, more talented people became teachers, lifting the status of the profession even higher" (p. 22). Boutique preparatory routes allow for distinct branding opportunities that are useful for recruiting desirable candidates, particularly into districts with historically poor reputations for quality.

A case study of a high-performing Midwestern urban school offers another alternative for effective staffing. Bowers (2008) applies a combined theoretical framework from Jim Collins's (2001) research on effective organizations described in Collins's book, *Good to Great*, and studies of NYC's Community School District 2—a celebrated urban district success story. Among his findings is the importance of hiring the right teachers. "In *Good to Great*, Collins (2001) uses the metaphor 'getting the right people on the bus'" (p. 41). Interestingly, a major point of Collins's study is that the leaders of the *Good to Great* companies first got "the right people on the bus. . . and the wrong people off the bus" (p. 41) and then "leveraged the collective talent of those people as a team to figure out where the bus should go. . . " (p. 167).

As any administrator will acknowledge, this is difficult to accomplish with the added pressure from teachers unions whose task is to protect the jobs of it members, regardless of their effectiveness. To reduce the risk of poor staffing decisions, the school in this study used substitute teachers as a hiring pool. As a trial run, effective individuals were hired as long-term substitutes and received professional development opportunities. By following this method, schools can offer permanent positions to those who are a good fit while releasing those who are a poor fit.

Addressing the importance of quality professional development. Having the best teachers in place is critical, but they must also have effective and timely training. Boston Public Schools, for example, conducts accountability reviews every four years that includes measures of professional development (Barnes, 2004). There are two primary challenges to this goal. The first is defining effective instruction.[1] The second challenge involves disseminating this information to staff.

[1]To read more on this topic, readers may want to consult the accompanying report on Pedagogy.

The second part of the challenge in instruction is, at least from a system management perspective, much more complex: giving thousands of teachers (in some cases hundreds of thousands of teachers) the capacity and knowledge to deliver great instruction reliably, every day, across thousands of schools, in circumstances that vary enormously from one classroom to the next—and with very little oversight. (p. 26)

While this task is difficult, it is not impossible. Partially through targeted staff development, Boston boosted its number of students meeting state mathematics standards from 25 percent to 74 percent and increased those meeting English standards from 43 percent to 77 percent. Similar results were seen from targeted professional development in England. In a literature review of best practices in professional development, the Wisconsin Department of Public Instruction (2000) found that effective development programs must be linked with the overall school vision and teachers' personal interests/needs. Successful programs are typically longitudinal and grounded in applicable research, and compliment the school's strategic plan.

Monitoring staffing quality. Among the top-performing school systems reviewed in the McKinsey report, two primary tools were used to monitor staffing quality. The first was regular student examinations, such as those mandated under NCLB. The assumption is that the ultimate responsibility for student achievement rests with the teacher; therefore, a poorly performing student is indicative of a poorly performing teacher. The second tool was school reviews or inspections that evaluate schools on a set of performance benchmarks.

Many school systems highlighted in the report have established external groups whose sole purpose is to conduct formative school reviews. School systems in New Zealand, Hong Kong, and England have independent inspectors. The Office for Standards in Education (Ofsted), the independent inspector in England, reports directly to Parliament, while Hong Kong has the Hong Kong Examinations and Assessment Authority (HKEAA), which remains part of the Ministry of Education. Regardless of the reporting structure, "In many of the top-performing systems, responsibility for monitoring outcomes has been separated out from the responsibility for improving those outcomes" (p. 37).

Diagnostic questions about recruiting & supporting teachers

McKinsey & Company (p. 41) provide the following key diagnostic questions to use when evaluating the health of a school system's staffing policies.

For getting the right people to become teachers. . .

1. What is the average academic caliber of people who become teachers?
2. How is the teaching profession viewed by university students and recent graduates?

3. How rigorous are selection processes in teacher training?
4. What is the ratio of places on initial teacher education courses to applications?
5. Do recruitment policies take advantage of natural screening opportunities such as long-term substitutes?

For developing effective teachers. . .

6. What is the total amount of coaching new teachers receive in schools?
7. What proportion of teacher time is spent in professional development?
8. Do teachers reflect on and discuss practice?
9. How much focused, systematic research is conducted into effective instruction and then fed back into policy and classroom practice?

For ensuring every student performs well . . .

10. What system-wide checks exist on the quality of school performance?
11. What action is taken to tackle underperformance? (Adapted from McKinsey & Company, 2007)

Option 2: Build a cohesive culture focused on data-driven decision making with defined site-level responsibility

By definition, organizations require some degree of cohesiveness. It is the degree and nature of this cohesiveness that predicates an organization's effectiveness. Most urban systems are relatively large bureaucracies and, as such, require a hierarchical structure. Effective organizations purposefully develop flexible hierarchies that allow for individual innovations within a scaffolded framework (Hoy, 2003). A report by MDRC for the Council of the Great City Schools (Snipes, Doolittle, & Herlihy, 2002) addresses the way large school systems create cohesiveness cultures that focus on data-driven decision making. The report draws from the retrospective case studies of three urban school districts—Houston Independent School District, Charlotte-Mecklenburg Schools, and Sacramento City Unified School District—selected for demonstrated academic improvements and narrowing the achievement gap over a (minimum) three-year period. Features of these three school systems were compared to those of similar districts that did not demonstrate marked academic improvement. The MDRC Report serves as the seminal work for this section.

Potential benefits and drawbacks of this option

All districts in the study faced similar challenges: endemically poor academic achievement, political conflict between members of the school board and the superintendent, high turnover rates among teachers, low expectations and a less than demanding curriculum, lack of instructional coherence, high student mobility, and overall unsatisfactory business practices. With similar challenges, why did certain districts thrive while others did not?

Developing a cohesive and focused organizational structure. Among other things, the case study districts developed a cohesive organizational structure focused on improving student achievement. The authors describe the steps the district took:

- They adopted or developed districtwide curricula and instructional approaches rather than allowing each school to devise its own strategies.
- They supported these districtwide strategies at the central office through professional development and support for consistent implementation throughout the district.
- They drove reforms into the classroom by defining a role for the central office that entailed guiding, supporting, and improving instruction at the building level.
- They committed themselves to data-driven decision making and instruction.

The Houston Independent School District, for example, initiated several programs with the specific goal of improving student achievement. The Balanced Approach to Reading program was developed to align the district's reading curricula with the state standards on which students would be tested. While it took curricular decision making away from the purview of the individual school, it created cohesiveness within the district regarding the material to teach. Charlotte-Mecklenburg Schools also implemented a "top down" curricular decision-making process and began professional development workshops and seminars for teachers to solicit feedback and build support for the new curriculum. By contrast, control districts put forth little effort to develop a cohesive districtwide culture or centralize curricular development. Indeed, the authors point out the disconnect between the central office and individual schools affected all parts of the system:

- The policies and practices of the central office were not strongly connected to intended changes in teaching and learning in the classroom.
- The districts gave schools multiple and conflicting curricula and instructional expectations, which they were left to decipher on their own. (p. 6)

Comparison District 1 suffered from high teacher and superintendent turnover. There was little cohesive leadership at the district level, and data management systems were outdated. District 2 suffered from a bloated and ineffective bureaucratic structure that drained funds away from its schools. Furthermore, it was embroiled in a legal battle over desegregation that drew resources away from academic reforms.

Building capacity for data-driven decision making. An important element not covered in the MDRC Report is that of building capacity for data-driven decision making and dissemination of results. Having data alone is not enough; personnel throughout all levels of the organization must understand how to use data to positively affect achievement. As explained by Wohlstetter, Datnow, and Park (2008), "At the most fundamental level, data-driven decision making plans involve an implicit contract between the central office and the schools whereby the district assists in the collection and analysis of data, but the responsibility for interpreting results and developing solutions is decentralized to schools and individual teachers" (p. 240). This principal-agent framework of shared responsibility requires an implicit (or sometimes explicit) agreement between parties.

In order to facilitate the effective use of data, Wohlstetter and colleagues recommend obviating teacher-level resistance by including school-level personnel in planning sessions on data use. This provides needed information to teachers about the proper use of data, but also allows district-level personnel to understand teachers' needs. And, although macro-level decisions are made at the district level, schools need the autonomy to make

instructional decisions on their own. Data usage should be frequent, and findings should be shared with all stakeholders, including community leaders and parents (Camblin, Gullat, & Klopott, 2003).

Diagnostic questions about developing a cohesive, data-driven culture

The following questions have been developed from the literature. They are designed to probe the amount of cohesion across decision-making levels and adherence to a data-driven culture.

For curricular cohesiveness and alignment. . .

1. Are curricular decisions being made at the district level?
2. Do teachers receive the necessary professional development to implement the curriculum?
3. Is there an alignment between curriculum and state standards?

For data-driven decision making. . .

4. Is the infrastructure in place to support data collection and dissemination?
5. Are teachers involved in planning sessions over the use of data?
6. Is data shared with all stakeholders?

Option 3: Develop & utilize longitudinal data collection systems for critical areas

Educational leaders need real-time, reliable data on school performance. While there are any number of important areas that require monitoring—two particularly critical areas for at-risk populations are dropout prevention and student academic performance. Under NCLB, tracking this data is no longer optional. However, there is considerable variation in the way it is done; and some methods are more effective than others. Moreover, data itself does nothing to improve the educational opportunities of students. Systems must be in place to use data in timely and effective ways (Snipes & Casserly, 2004). This section reviews some promising practices in data collection and use for dropout prevention and performance tracking.

If students are not in school, neither the most effective teaching nor most engaging curricula will do any good. Dropout rates among Our Kids have reached epidemic levels with approximately one-half of minority students exiting school before graduation. The figures are even more startling within the largest districts, with dropout rates ranging between 60–70 percent for minority populations.

Compounding this problem is the fact that, for a variety of reasons, dropout figures likely underestimate actual numbers (Orfield, Losen, Wald, & Swanson, 2004). The two most frequently reported statistics are from the National Center for Education Statistics (NCES) and the Current Population Survey (CPS), which the Census Bureau conducts. NCES estimates rely primarily on state-reported data, which many experts consider inaccurate. CPS estimates rely on self-report data administered through surveys of students the bureau can track down. In short, more accurate data is needed.

To keep students in school, systems should focus on three key areas: accurately tracking dropout rates, understanding and using the predictors of dropping out to better monitor student progress, and adopting or developing dropout preventions programs that provide Our Kids the tools and motivation they need to succeed in postsecondary education.

Dropout prevention and tracking systems

Accurately tracking dropout rates is difficult. High mobility among Our Kids means that schools often have to guess whether a student has dropped out or simply has moved to another school. Compounding this difficulty, many students exit schools during transition periods

when they are hardest to track. For example, a study by the Southern Regional Education Board (SREB) reports that 38 percent of Florida students who drop out during high school do so during their freshman year (Bottoms & Timberlake, 2007). There is considerable debate surrounding the most accurate source for dropout data, but a number of experts recommend using the Cumulative Promotion Index (CPI), which is calculated from enrollment figures provided annually by each district to the Common Core of Data (Orfield et al., 2004).[2] CPI estimates are typically at odds with state reported figures.

Florida, for example, reported 2004 minority dropout rates at approximately four percent, while the CPI estimated closer to 50 percent. In order to develop more accurate measures, Orfield and colleagues (2004) recommend that states begin longitudinally tracking students through a "unique common identifier" (p. 15). Of course, any tracking number is useful only if the student remains within the boundaries of the system. Ultimately, a national tracking number might serve a more useful function. Until that time, and because there is currently no authoritative measure, districts are cautioned to use and report estimates from multiple sources.

The ability to predict who will drop out gives districts a jump start on averting crises before they arise. Paradoxically, some educational policies targeted at improving the achievement of at-risk populations may actually exacerbate the dropout problem. A longitudinal mixed-methods analysis of Texas data reveals that, "the state's high-stakes accountability system has a direct impact on the severity of the dropout crisis" (McNeil, Coppola, Radigan, & Heilig, 2008). One reason for this is the use of retention policies for high school freshmen. Accountability policies make student retention more appealing for a school because the failing test score does not count against Adequate Yearly Progress (AYP). However, according to estimates by the authors, retaining a student can increase his risk of dropping out by as much as 250 percent.

Of course, there are many student-level factors that contribute to the risk of dropping out, including lack of participation in extracurricular activities, frequent absenteeism, unstable home environment, low family income, and having non-graduate parents (McNeil et al., 2008; Nichols, 2003). As McNeil and colleagues report, these predictors are additive. While any one factor alone may not be predictive, as they accumulate, so does an individual's likelihood of dropping out of school.

Dropout prevention programs have a long history. Some, such as *Talent Search* (Constantine, Seftor, Martin, Silva, & Meyers, 2006), have existed since

[2]The National Governors Association developed its own plan for collecting, analyzing, and reporting dropout data called the Graduation Counts Compact. The agreement was signed by the governors of all 50 states in 2005. For a description of the agreement, see the National Governors Association's report, *Implementing Graduation Counts: State Progress to Date* (2008).

implementation of the Higher Education Act of 1965 and have been tested across the nation; others are site-developed and have a much smaller profile. Regardless, all of the dropout prevention programs with proven effectiveness reviewed for this report have similar features to draw upon. Four programs included in this report have been empirically studied, and those studies were reviewed by the Institute of Education Sciences What Works Clearinghouse (WWC):

- Achievement for Latinos through Academic Success (ALAS)
- Talent Search
- Check & Connect
- Accelerated Middle Schools

All met WWC standards completely or with reservations (the top two acceptable levels of evidence).[3] A common element of each program is the early identification of at-risk students in middle school or freshman year of high school. Additionally, each program was founded on a strong interpersonal component. Students identified as being at risk of dropping out were assigned an adult mentor who built a genuine relationship with the student. All mentors focused on behavioral modification, study skills, problem solving skills, and academic advisement. Mentors within the Talent Search program went a step further and actually coordinated site visits to postsecondary institutions and helped students complete financial aid applications. The structure of each program varied. The Accelerated Middle Schools program, for example, can operate as a stand-alone school or as a school within a school. Regardless of the structure, each of these programs has significantly reduced dropout rates among participants by tailoring its components to the unique needs of the students and school.

Tracking academic performance

Any discussion of systems diagnostics should include the topic of monitoring academic performance. Whether or not to monitor academic performance is no longer a decision districts have to make; the No Child Left Behind Act of 2001 requires annual testing of students in grades three through eight. There is clear evidence that accountability systems have led to performance gains (Hanushek & Raymond, 2002).

The relevant discussion now centers on how to measure academic performance. At present, NCLB requires states to measure achievement against criterion-referenced state standards. In other words, student performance (and, by association, school performance) is measured against a set of fixed standards established by the state. In assessment terminology, this is a cross-cohort model because it compares the performance of different student cohorts as they move through the system. For example, the academic performance of a 3rd-grade cohort from the previous year is compared against the 3rd-grade cohort of the current year—different groups of students within the same grade. The rationale behind this model is that all students should obtain mastery over defined grade specific curricula. However, there are problems with this model.

The primary complaint with cross-cohort comparisons is that student performance in a given year is the culmination of all prior educational experiences in addition to non-education factors. The 3rd-grade student's achievement scores are really the culmination of all he or she has learned K–3. Due in part to this central criticism, the U.S. Department of Education is currently allowing ten states to pilot an alternative measure of academic performance that accounts for prior educational experiences—the gain model.[4] For over a decade, researchers and policymakers have debated how

[3]U.S. Department of Education (July, 2008); U.S. Department of Education (September, 2006; October, 2006; December, 2006).

[4]The current terminology used by the U.S. Department of Education is "growth model." For purposes of consistency with the reviewed literature, this report uses the term "gain" rather than growth.

students and schools should be assessed (Meyer, 1997). While there are currently no definitive answers, there are some promising practices from which to draw lessons.

Potential benefits and drawbacks of this option

Districts need data systems that longitudinally track student performance (Huggins, 2004). Prior to the passage of NCLB, Minneapolis Public Schools instituted a comprehensive longitudinal assessment system that incorporated cross-cohort comparisons, gain scores, and value added models. While the system has changed over the years, it serves as a strong example of assessment done right. Heistad and Spicuzza (2000) presented an overview of the system at the 2000 meeting of the American Education Research Association. Their paper serves as the seminal work for the discussion of this option.

Developing multifaceted assessment models. The Minneapolis model was ahead of its time. Before accountability mandates, the city took the lead developing a multifaceted assessment model whereby the performance of both students and school could be evaluated. In total, the city used the following academic indicators:

- Change in achievement level compared to performance standards
- Student achievement level compared to expected national norm growth
- Student achievement compared to predictors of levels of performance based on pretest score and student demographics (Heistad & Spicuzza, 2000)

As the authors note, "cross-cohort indicators examine change over time, albeit from different groups of students" (p. 6). However, the city went beyond this assessment and instituted a student gain model. Gain models adjust for a student's prior experience by substituting or statistically adjusting for entering achievement levels (Ladd & Walsh, 2002). This allowed Minneapolis to estimate the learning that took place over the assessed year. While this model did not factor out non-educational factors, it had the benefit of accounting for the cumulative educational experience of the student, thus providing a more accurate gauge of progress made during that year. Additionally, the city instituted a value-added model to evaluate overall school effects. Considered "exotic" by some, value-added models are a mainstay of economists and educational researchers because they provide the estimated contribution of a specified variable over and above the effects of other inputs.

The Minneapolis model controlled for the following aspects: free/reduced lunch status, English-language learner status, special education status, gender, race, guardians living at home with the child, and poverty level of the child's

neighborhood (p. 20). What remained after statistically controlling for these factors was an estimation of the contribution the school made to student performance. As the authors explain, the value-added model was used to identify schools that "beat the odds based on pre-test scores and student characteristics" (p. 8).

Minneapolis is not alone is it use of value-added models. Memphis, Tennessee, has recently piloted its own version to estimate the overall school effect across grades and subjects (Booker & Isenberg, 2008). While helpful, value-added models have their problems. An investigation of those used by North Carolina and South Carolina suggest that they likely underestimate the effects of low-performing schools and overestimate the effects of high-performing schools (Ladd & Walsh, 2002). The important lesson is that there is no panacea; however, gain models and value-added models provide important information not available through those traditionally in use by most school systems.

Diagnostic questions about collecting & utilizing longitudinal data

The following questions have been developed from the literature to probe the extent which a school system has established effective longitudinal dropout prevention and academic performance tracking systems.

For dropout prevention. . .

1. Do school and district personnel recognize the risk factors of dropping out?
2. Are students who exhibit multiple risk factors flagged for additional services?
3. Has the district established accurate methods for tracking dropout rates?
4. Are students required to take college entrance exams?

For academic performance. . .

5. Does the school and district utilize multiple models to gauge student/school performance?
6. Do models account for prior levels of achievement?

Option 4: Enact timely & evidence-based interventions where systems have failed

Data collection is useful insofar as it leads to better-informed decision making. Occasionally, that data may lead to the conclusion that the systems established to support Our Kids have failed. When this happens, timely interventions are needed that quickly and effectively deal with the problem before it becomes endemic. These interventions can occur at any level. A teacher may realize that, despite participation in a dropout prevention program, a student has missed over a week of school with no explanation of his or her whereabouts. A principal might notice that a grade-level team is not tailoring instruction based on assessment data. However, this report focuses on the systems level of diagnostic interventions—that which takes place when a school or district fails.

Unfortunately, there is a dearth of empirical literature around systems-level interventions. However, several case studies of state takeovers and school reconstitutions have been conducted. From these, the McREL research team compiled promising practices for interventions when systems fail.

Potential benefits and drawbacks of this option

By definition, a failure means that a system is not functioning as expected. A policy brief by the Education Commission of the States (ECS) examines case studies of situations when system-

level failures have required district or state intervention (Ziebarth, 2002). This ECS brief serves as the seminal work for this discussion.

Ziebarth notes that there are two general options available to educational leaders when a school or district is failing—reconstitution or full state takeover. Both are considered the "nuclear option" and their use is rare. As of April 2002, 19 states had policies in place to allow for school reconstitution and 24 had policies for full state takeover. Historically, interventions such as these have taken place for reasons such as high dropout rates, unsafe conditions, poor/declining academic performance, and chronic complaints from parents and community members about overall quality.

Taking over failing schools. Reconstitution is the smaller of the two options and takes place when the district, or occasionally the state, takes over a failing school. School reconstitution involves "creating a new philosophy, developing a new curriculum, and hiring new staff at a low-performing school" (p. 6). Among the case studies reviewed for the ECS report, the authors found "anecdotal evidence" that reconstitutions may help stabilize schools and bring an increase of community support.

With a state takeover, the legislature, state board of education, or court hands managerial responsibility of a district to a third party. The most celebrated example of state takeover happened with Chicago Public Schools, which saw strong academic improvement in addition to decreases in corruption within the central office. However, not all state takeovers have been as successful in improving student achievement. Typically, greater improvements are realized within central office operations and infrastructure. "The bottom line is that state takeovers, for the most part, have yet to produce dramatic and consistent increases in student performance, as is necessary in many of the school districts that are taken over" (p. 5).

Another report that examines a similar set of case studies finds some limited benefits to state takeover. Wong and Shen (2001) conclude that academic gains are present, but only within the poorest performing elementary schools. This is not surprising, as these are the target of most takeovers.

Before taking action against a failing school or district, policies should clearly specify the triggers for intervention, and these standards should be enforced consistently across sites. Of course, states and districts are hesitant to take over failing systems. School board and legislatures typically do not have the necessary resources to run complex systems. However, it may not be necessary to take action against a failing system. The threat of takeover itself may be enough to shake up leadership and spur reforms.

Focusing interventions on a limited number of high-priority goals. When these reforms do not take place, or are ineffective, there are certain things leaders can do to increase their chances of a successful intervention.

A review of successful turnarounds across different organizational types reveals common elements (Hassel & Hassel, 2009). First, successful turnaround campaigns focused on a "few early wins." Rather than attempting to reform the entire system, interventions focused on a limited number of high-priority goals with equally high visibility. According to the authors, an example of this might be increasing the reading scores within a struggling elementary school. Early wins allow a turnaround campaign to garner early support and build momentum. Second, effective turnarounds purge operating norms that stand in the way of high performance. During the turnaround campaign of Continental Airlines, executives ceremonially burned a copy of the voluminous regulations manual that had become emblematic of over-bureaucratization within the company. Finally, get the right people into key leadership positions quickly. All turnaround campaigns face resistance—some from individuals who hold differing opinions about new policies but genuinely work for best interest of the organization, and others who are simply protecting turf. Having leadership in a position to help win over the former and eliminate the latter is crucial to the success of an intervention.

Diagnostic questions about systems-level interventions

The following questions have been developed from the literature to probe a system's capacity for school restructure or takeover in addition to the scope of the intervention.

1. Should a state takeover a low-performing district/school? What are the trigger points for takeover?
2. If officials of low-performing districts/schools are given the same authority as takeover officials, can they enact the same reforms?
3. Does the state/district have the necessary resources to enact the required intervention?
4. What is the next step if an intervention fails to yield sufficient improvement?
5. Will new leadership be needed or will existing leaders be retained?
6. Which specific policies or systems allowed failure to occur? How can these be changed? (Adapted from Ziebarth, 2002)

Final Thoughts

Providing an educational experience that ensures college readiness (let alone, success) to underprivileged students is a daunting yet critical goal.[5] This report began with the assumption that re-forming the educational systems that support Our Kids under the HRO framework will greatly improve their effectiveness. The HRO framework is not new; however, it is yet to be comprehensively applied to educational organizations.

To some, the idea of a continual focus on failure might seem anathema to education. Shouldn't the education of at-risk students focus on positive achievements? The response to that question is that Our Kids are at-risk because they are surrounded by conditions that threaten academic success. Gawande's (2007) meticulous surgery preparation, described in the introduction of this report, illustrates what happens when organizations systematically focus on threats of failure. Protocols and habits of the mind are designed and incorporated into core operations that obviate those failures before they occur. Only by focusing on threats while celebrating successes can an organization develop a culture of continuous self-improvement.

HROs are learning organizations, and a key element of HROs is their extensive use of systems diagnostics throughout the operational continuum. Diagnostics are not only viewed as a tool for data collection, but they are also embedded with core operations and interventions. In crafting a learning system for Our Kids, developers should consider ways in which the HRO framework of systems diagnostics conceptualized in this report would help produce a more effective and reliable educational experience.

The reader should keep in mind, however, that there is little research literature around the application of the HRO framework to education. As such, the four articulated options are not exhaustive. Improvements such as aligning high school exit requirements with postsecondary coursework (Achieve, 2008; Conley, 2007; Dounay, 2006; Huggins, 2004); improving school leadership (Shannon & Bylsma, 2004); conducting regular school audits (Bauer & Mitchell, 1997); developing students' affective skills (Williams & Noble, 2005; Ward & O'Shaughnessy, n.d.) and the use of timely student remediation (Nunley, Shartle-Galotto, & Smith, 2000) may serve as equally important options. The authors of this report encourage their further exploration and inclusion in the proposed model.

In addition, it is useful to draw connections between the various components of the framework. Core operations are the preventative medicine of an effective diagnostic system. An important feature of core operations that did not receive much attention in earlier sections is that of a culture of high expectations that focuses on academic achievement (Bellamy, 2005). Our Kids face many challenges; the sympathetic reflex to those facing numerous challenges is to provide accommodations. These often take the form of lowered expectations. Though lowering rigor may help students' grades in the short term, it leaves them unprepared for the challenges of postsecondary education. The fact is, most high

[5]See the description of Project GRAD (Graduation Really Achieves Dreams) for an example of the difficulties facing even the most well-designed programs (Snipes, Holton, & Doolittle, 2006).

school exit exams and state achievement tests are not challenging enough to serve as screening mechanisms for college performance (Achieve, 2004; Adelman, 2006). Providing additional accommodations beyond this, though well intentioned, is a myopic practice that damages Our Kids' chances for long-term academic success. Instead, high expectations should be matched with the necessary academic and social supports.[6]

Finally, monitoring system performance allows for informed decision making and course corrections. When weaknesses in the system are detected or effective interventions are identified, they should serve to update core operations. If a school has implemented an innovative dropout prevention program that has proven effective over several years, the district might look at implementing the program across multiple schools.

Whether a particular core operation or intervention is effective should be determined by its impact on academic achievement. Having an accurate picture of achievement is contingent upon timely and reliable assessments. Unfortunately, assessments are often summative—given at the end of the academic year, when their ability to inform instruction is limited. For effective systems-level diagnostics, assessments should occur more often and take a more formative role (Quint, Sepanik, & Smith, 2008). As the Assessment report in this series of reports explains, frequent assessment for formative purposes allows for more informed instructional decision making at the school level. It also provides district executives more data points on which to judge the health of the system. Rather than waiting until the end of school year, when it is too late for students, immediate interventions can be put in place for struggling schools or districts.

This report examined the use of systems diagnostics in public education. While diagnostics are used effectively in certain domains, there is no comprehensive and systematic use throughout education. In short, education's use of diagnostics is haphazard and lacks an overarching framework. The way High Reliability Organizations, such as hospitals, think about diagnostics provides an excellent example for education. The high-failure risk environment of a hospital is analogous to that of urban schools. Fatigue, miscommunication, and infection conspire against a patient's health in a hospital just as poverty, crime, and low expectations conspire against student success in an urban school. The difference is that the former recognizes these threats and has established comprehensive diagnostics to handle them, while the latter has not. This report is intended to begin a dialogue on this important topic and provide a framework, based on the HRO model, on which to build a comprehensive diagnostic system.

---

[6]See the accompanying reports on Our Kids and Student Supports.

References

Achieve, Inc. (2004). *Do graduation tests measure up? A closer look at state high school exit exams.* Washington, DC: Author.

Achieve, Inc. (2008). Closing the expectations gap, 2008. *An annual 50-state progress report on the alignment of high school policies with the demands of college and careers.* Washington, DC: Author.

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college.* Washington, DC: U.S. Department of Education.

Barnes, F. (2004). *Part III: Rubrics and standards of practice for use with the school-improvement guide.* Providence, RI: Annenberg Institute for School Reform.

Bauer, S. C., & Mitchell, S. M. (1997, March). *The quality audit: A framework for internal analysis of the capacity for change.* Paper presented at the Annual Meeting of the American Education Research Association, Chicago, IL.

Bellamy, G. T., Crawford, L., Marshall, L. H., & Coulter, G. A. (2005). The fail-safe schools challenge: Leadership possibilities from high reliability organizations. *Educational Administration Quarterly, 41*(3), 383–412.

Booker, K., & Isenberg, E. (2008). *Measuring school effectiveness in Memphis.* Princeton, NJ: Mathematica Policy Research, Inc.

Bottoms, G., & Timberlake, A. (2007). *Giving students a chance to achieve: Getting off to a fast and successful start in grade nine.* Atlanta, GA: Southern Regional Education Board.

Bowers, A. J. (2008). Promoting excellence: "Good to Great," NYC's District 2, and the case of a high-performing school district. *Leadership and Policy in Schools, 7*(2), 154–177.

Camblin, S. J., Gullatt, Y., & Klopott, S. (2003). *Strategies for success: Six stories of increasing college access.* Boston, MA: Pathways to College Network Clearinghouse.

Collins, J. (2001). *Good to Great.* New York: Harper Collins Publishers, Inc.

Conley, D. T. (2007, November). *How do you know that your students are ready for college?* Paper presented at the Early College High School Network, Pomona, CA.

Constantine, J. M., Seftor, N. S., Martin, E. S., Silva, T., & Meyers, D. (2006). *A study of the effect of the Talent Search program on secondary and postsecondary outcomes in Florida, Indiana, and Texas: Final report from phase II of the national evaluation.* Washington, DC: Mathematica Policy Research for the U.S. Department of Education.

Dounay, J. (2006). *Embedding college readiness indicators in high school curriculum and assessments: Policy Brief.* Denver, CO: Education Commission of the States.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children, 15*(1), 35–54.

Gawande, A. (2007). *Better: A surgeon's notes on performance.* New York: Picador.

Hanushek, E. A., & Raymond, M. E. (2002, June). *Lessons about the design of state accountability systems.* Paper presented at Taking Account of Accountability: Assessing Policy and Politics, Harvard University.

Hassel, E., & Hassel, B. (2009). The big U-turn: How to bring schools from the brink of doom to stellar success. *Education Next, 9*(1), 21–27.

Heistad, D., & Spicuzza, R. (2000). *Measuring school performance to improve student achievement and to reward effective programs.* Paper presented at the American Education Research Association conference, New Orleans, LA.

Hoy, W. K. (2003). An analysis of enabling and mindful school structures: Some theoretical, research, and practical considerations. *Journal of Educational Administration, 41*, 87–108.

Hoy, W. K., Gage, C. Q., & Tarter, C. J. (2006). School mindfulness and faculty trust: Necessary conditions for each other? *Educational Administration Quarterly, 42*(2), 236–255.

Huggins, E. M. (2004). *A seamless P–16 system of education: Best practices high school reform.* Prepared for the Best Practices Panel. Retrieved from http://www.ode.state.or.us/teachlearn/specialty/pre-post/p16reformfinal.pdf

LaPorte, T. R. (1996). High reliability organizations: Unlikely, demanding and at risk. *Journal of Contingencies and Crisis Management, 4*(2), 60–71.

Ladd, H. F., & Walsh, R. P. (2002). Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review, 21*(1), 1–17.

McKinsey & Company. (2007). *How the world's best-performing school systems come out on top.* New York: Author.

McNeil, L. M., Coppola, E., Radigan, J., & Heilig, J. V. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives, 16*(3), 1–48.

Meyer, R. H. (1997). Value-added indicators of school performance. In E. Hanushek & D. Jorgensen (Eds.), *Improving the performance of America's schools: The role of incentives* (pp. 171–196). Washington DC: National Academy Press.

National Governors Association. (2008). *Implementing graduation counts: State progress to date, 2008.* Washington, DC: National Governors Association.

Nichols, J. D. (2003). Prediction indicators for students failing the state of Indiana high school graduation exam. *Preventing School Failure, 47*(3), 112–120.

Nunley, C. R., Shartle-Galotto, M. K., & Smith, M. H. (2000). Working with schools to prepare students for college: A case study. *New Directions for Community Colleges,* 111.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257.

Orfield, G., Losen, D., Wald, J., & Swanson, C. B. (2004). *Losing our future: How minority youth are being left behind by the graduation rate crisis.* Cambridge, MA: Harvard Education Publishing Group.

Quint, J. C., Sepanik, S., & Smith, J. K. (2008). *Using student data to improve teaching and learning.* New York: MDRC.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 9*4(2), 247–252.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the "Prospects" study of elementary schools. *Teachers College Record, 104*(8), 1525–1567.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Shannon, G. S., & Bylsma, P. (2004). *Characteristics of improved school districts: Themes from research*. Olympia, WA: Office of Superintendent of Public Instruction.

Snipes, J. C., & Casserly, M. D. (2004). Urban school systems and education reform: Key lessons from a case study of large urban school systems. *Journal of Education for Students Placed at Risk, 9*(2), 127–141.

Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: Council of the Great City Schools.

Snipes, J. C., Holton, G. I., & Doolittle, F. (2006). *Charting a path to graduation. The effect of project GRAD on elementary school student outcomes in four urban school districts.* New York: MDRC.

U.S. Dept. of Education, Institute of Education Sciences, What Works Clearinghouse. (2008, July). *WWC Intervention Report: Accelerated Middle Schools.* Washington, DC: Author.

U.S. Dept. of Education, Institute of Education Sciences, What Works Clearinghouse. (2006, September 21). *WWC Intervention Report: Check & Connect.* Washington, DC: Author.

U.S. Dept. of Education, Institute of Education Sciences, What Works Clearinghouse. (2006, October 5). *WWC Intervention Report: Achievement for Latinos through Academic Success*. Washington, DC: Author.

U.S. Dept. of Education, Institute of Education Sciences, What Works Clearinghouse. (2006, December 28). *WWC Intervention Report: Talent Search.* Washington, DC: Author.

Ward, T., & O'Shaughnessy, T. (n.d.) *A three-step approach to ensuring that your students are college-ready*. Eugene, OR: Center for Educational Policy Research.

Weick, K., & Sutcliffe, K. (2001). *Managing the unexpected: Assuring high performance in an age of complexity.* San Francisco: Jossey-Bass.

Williams, N. J., & Noble, J. P. (2005). *School-level benefits of using PLAN over time: Research report series 2005-01.*

Wisconsin Department of Public Instruction. (2000). *Characteristics of successful schools.* (2nd ed.) . Madison, WI: Author.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement, 19*(3), 239–259.

Wong, K., & Shen, F. X. (2001, Aug.-Sept.). *Does school district takeover work? Assessing the effectiveness of city and state takeover as a school reform strategy.* Paper presented at the Annual Meeting of the American Political Science Association. San Francisco, CA.

Ziebarth, T. (2002). *State takeovers and reconstitutions.* Denver, CO: Education Commission of the States.

References Reviewed but Not Cited in This Report

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of "No Child Left Behind." *Race, Ethnicity and Education, 10*(3), 245–260.

Mintrop, H., & Trujillo, T. (2007). *The practical relevance of accountability systems for school improvement: A descriptive analysis of California schools.* Los Angeles, CA: Center for the Study of Evaluation.

Southern Regional Education Board. (2000). *Getting results with accountability: Rating schools, assisting schools, improving schools.* Atlanta, GA: Southern Regional Education Board.

Venezia, A., & Bueschel, A. C. (2001). *The Bridge Project: Strengthening K–16 transitions policies.* Stanford, CA: Stanford University.

Appendix

Literature review method

In June 2008, the Stupski Foundation created a conceptual framework for the reinvention of American education. The framework identified seven essential components and focused on delivering 21st century college readiness for all students, but especially for "Our Kids," children of color and poverty. The Foundation explained that "graduating all students from high school with the knowledge and skills that qualify them as 'college ready' is the most meaningful and measurable way to increase life choices and options for all children, but most especially children of color and poverty" (About the Foundation, para. 3).

The Learning System includes four core teaching and learning components: Curriculum, Assessments, Pedagogy, and Supports. Surrounding these components, are three organizational components necessary to support the core: Leadership/Human Capital, Systems Diagnostics, and a Dashboard of College Readiness Indicators (College Readiness Learning System, n.d.).

The Foundation envisions convening a Design Collaborative, a cross-sector group of researchers, practitioners, and designers from inside and outside education, to "define, develop and continually improve" (Design Collaborative, n.d.) all of the components. To orient Design Collaborative members to the accumulated and maturing knowledge base related to each of the components and to children of color and poverty, the Foundation contracted with Mid-continent Research for Education and Learning (McREL). McREL conducted eight literature reviews—one on each of the components plus one on Our Kids—to identify and integrate theories and philosophical perspectives, issues, scientifically based research practices, unmet needs, and innovations relevant to designing one or more of the system components to accelerate learning for Our Kids.

This Appendix contains a description of the review method, including a general explanation of McREL's approach and descriptions of the particular procedures used for each phase of the review: identification of key hypotheses and research questions, literature search, identification and cataloguing of finds, and generating and communicating recommendations.

McREL's overall approach

Since the primary users of the reviews are the members of the Design Collaborative, the qualitative, iterative approach taken for the literature reviews sought to achieve the multiple goals of identifying emerging ideas, counterproductive orthodoxies, and promising practices relevant to the reinvention of the Learning System. Thus, eight research teams were assembled, each with one or more researchers familiar with the respective topic areas.

Qualitative approach. A *qualitative approach* shares several practices with those of *systematic* reviews, including comprehensive searches and transparency to reduce bias, but it differs with respect to inclusion/exclusion criteria. Systematic reviews emphasize explicit and a priori inclusion/exclusion criteria and criteria for evaluating the methodological quality of individual studies, carefully limiting the sources of evidence to support inferences about cause and effect relationships (Cooper, Hedges, & Valentine, 2009). The qualitative approach emphasizes diverse sources and types of evidence and knowledge to support a broader base of inferences (Pope, Mays, & Popay, 2007; Suri & Clarke, 2009).

The qualitative approach is particularly well-suited to the review's purpose and audience because the Design Collaborative needs both empirical studies and other literature to identify possible innovations for the current education system. An assumption underlying the Foundation's work to fundamentally reinvent American education is that the current system fails to deliver college readiness for all students, especially Our Kids. This assumption is supported by research indicating that students of color and in poverty have low high school and college graduation rates, and research from the last two years shows that college graduation rates for minority and poor students have further declined (American Council on Education, 2008). Therefore, a priority for the Foundation's work is to identify innovations that have not yet been studied, with the intent to evaluate their effectiveness. Literature specific to innovations is found outside the traditional scientific or academic journals.

Inclusive approach. McREL researchers adopted an inclusive approach, searching for and including phenomenological reports describing the experiences of Our Kids in and out of school and documenting the challenges and successes of their teachers and educational leaders. The researchers included literature on innovative, emerging models and untested ideas, as well as reports on mature, well-specified models with experimental evidence of effectiveness. Relevant quantitative research literature included correlational and experimental studies and meta-analytic reviews. Narrative reviews of research were included, as were policy briefs and position papers produced by opinion leaders and professional organizations. Literature sources included the World Wide Web, peer-reviewed journals, and practitioner magazines. Each document was identified by type of literature and evaluated in terms of the quality of the supporting evidence. Care was taken to draw only those inferences appropriate to the quality of the evidence.

McREL researchers judged the quality of the evidence in the context of the type of literature or study design and in relation to its relevance to answering particular questions. Guidance from Pope, Mays, and Popay (2007) on conducting reviews in the field of health research supports this approach:

> The inclusion of diverse sources of evidence in a review does not mean abandoning the rigor of a systematic review, but it does mean judging the quality of evidence in context and defining the relevance of evidence to answering specific questions, rather than defining some forms of evidence as intrinsically, and universally, of lower quality than others. (p. 1)

Each research team followed the five or six phases of any review process relevant to a quality knowledge synthesis (Cooper, Hedges & Valentine, 2009; Suri & Clarke, 2009). Table 1 (see p. 39) provides a side-by-side comparison of the phases of a systematic review of research (Cooper, Hedges & Valentine, 2009), a qualitative review (Suri & Clarke, 2009), and McREL's approach to this review.

Each team began by drawing from pertinent philosophical and theoretical literature and preliminary discussions with the Foundation to formulate hypotheses and research questions. Each team conducted extensive searches to find as much relevant literature as possible in order to include literature from the scientific and academic journals as well as literature from harder-to-find, cutting edge innovators. Additionally, teams revisited databases and alternative sources to purposefully search for additional literature written by authors identified by one or more stakeholders or to fill conceptual gaps that became apparent during the identification and cataloguing of findings and generating and communicating recommendations phases.

The phased process was iterative (Cooper, 2009) reflecting new understanding and insights as the search, analysis, interpretation, and discussions between component teams and between the Foundation and McREL progressed toward conceptual clarity and the exhaustion of new search hits. The number of documents included in each team's review was extensive, and the types of literature varied representing the experiential knowledge of a diverse group of stakeholders, including researchers, teachers, administrators, program developers, and leaders and scholars at the local and national levels.

Table 1: Phases of a literature review

Phase	Cooper, Hedges & Valentine (2009, p. 8)	Suri & Clarke (2009, p. 414)	McREL's approach
1	Problem formulation	Drawing from pertinent philosophical and theoretical discussions	Identification of key hypotheses
2		Identifying an appropriate purpose	Identification of research questions
3	Data collection	Searching for relevant evidence	Literature search
4	Data evaluation	Evaluating, interpreting, and distilling evidence	Identification and cataloguing of findings
5	Analysis and interpretation	Constructing connected understanding	Generating and communicating recommendations
6	Public presentation	Communicating with an audience	

Team approach. Teams were composed of researchers and practitioners with different areas of expertise. Teams met weekly, and team leaders from across teams met biweekly. Meetings were used to update other individuals and teams and share resources, pose and address questions, challenge assumptions, provide guidance on interpretation of evidence, open up new areas of consideration, clarify boundaries and overlap between system components, consider alternative perspectives, and develop connected understanding.

Identification of key hypotheses and research questions

McREL teams began by clarifying terms, relationships, and the conceptual scope of each review. Teams read and discussed a document produced during the Foundation's strategy definition process, *Research Guide for CRLS: Outline of Research Questions for Each Component of the CRLS* (n.d.). Included in this Guide were preliminary questions for each literature review. Teams previewed relevant literature, confirmed that the questions could be answered by the extant knowledge base, and posed additional questions when important issues related to accelerating learning for students of color and poverty were identified in the literature but missing in the Guide. The revised set of questions for each system component and Our Kids was reviewed and refined during ongoing dialogue between the Foundation and McREL.

Literature search

Multiple searches were conducted in a phased approach to identify as much literature as possible related to each system component and Our Kids. Teams conducted searches using multiple bibliographic databases: Academic Onefile, Academic Search Premier, Educators Reference Complete, ERIC, JSTOR, Proquest, and PsychInfo. Teams also conducted manual searches of journal and book tables of contents and reference lists of articles. Additional searches were conducted specifically to identify recent experimental and other research and reviews on the efficacy of interventions for accelerating learning of students of color and poverty. These searches were conducted by visiting the U.S. Department of Education What Works Clearinghouse Web site (http://ies.ed.gov/ncee/wwc/reports/) and the Campbell Collaboration Library of Systematic Reviews Web site (http://www.campbellcollaboration.org/library.php). Relevant documents were identified on state education agency (SEA) Web sites, and SEA officials were interviewed or named as seminal authors or sources of models that had been developed and implemented to monitor and accelerate learning of Our Kids.

Each team identified and used key terms and synonyms relevant to the topic for searching. Searches were conducted for literature published in the most recent 10 years (1998–2008); however, works by seminal authors and other recommended literature were included from outside these years. The search landscape varied for each team based on the topic and relevant sources; for example, while What Works Clearinghouse was a relevant source for the Pedagogy team, it was not a relevant source for the Leadership/Human Capital team. Internal review of search records and results led to additional leads on sources. Searching continued until all recommendations had been implemented and/or few new hits were identified.

Identification and cataloguing of findings

A coding protocol was developed and implemented to categorize the literature. Each team used the same protocol, adding categories and decision rules, as needed to organize the particular literature relevant to their topic. Each team leader and one or more members of each team were trained on the decision rules in the coding protocol and provided follow-up support to resolve uncertainties in its application. Team leaders periodically conducted quality assurance reviews of completed coding sheets and updated the protocol as needed during weekly team leader meetings or discussions with the Foundation. The coding protocol included identifying the following information:

- Full APA reference citation
- Category of literature (i.e., primary and secondary relevance)
- Type of literature (e.g., quantitative study, policy brief, program description)

- Locale
- Outcome
- Grade level
- Program or innovation name and description
- Main findings or points
- A recommendation for or against summarizing and including the selection in an annotated bibliography.

In addition, component teams added to the protocol by categorizing relevance to particular parts of their conceptual model or concept map.

Guidelines were developed and used by teams to identify counterproductive orthodoxies, unmet needs, next practices, promising practices, and best practices based on type of literature and quality of evidence. These were defined in the following ways:

- *Counterproductive orthodoxies:* Conventional ways of providing education which may be impeding success of Our Kids
- *Unmet needs:* Areas where Our Kids are not yet well served by the current system of education
- *Next practices:* A program or practice that needs to developed, adapted, invented, and tested in response to an unmet need related to accelerating learning for Our Kids
- *Promising practices:* Practices based on research but not supported by rigorous efficacy data from randomized controlled trials
- *Best practices:* Practices demonstrated by one or more randomized controlled trials to be effective in improving outcomes for Our Kids

The research team reviewing the college readiness component of the Learning System employed a slightly different process. Rather than using the categories above, this team reviewed literature on college readiness and categorized findings into four essential areas as defined by the Foundation and Conley (2007): cognitive strategies, content knowledge, academic behaviors, and contextual skills.

Component teams met weekly to discuss and categorize findings and to develop a conceptual map of the insights gained from the literature summaries and review. Teams used different conceptual mapping tools (e.g., SmartArt) to organize the insights (findings) and presented and discussed their respective maps at cross-team meetings. Features common across teams' concept maps were identified and a standard framework developed. Teams arranged findings onto the concept maps, identifying conceptual gaps and conflicting or discrepant findings, and returned to searching and reviewing to fill in the gaps and resolve or explain discrepant findings. The conceptual maps served as an organizing framework for report construction.

Generating and communicating recommendations

Working collaboratively, component teams drew conclusions from the insights (findings) derived from the review and identified potential options and recommendations for each component of the system. Teams used an iterative process of identification, reviewing for validity against the knowledge base, and further refinement until they determined they had identified the most promising options and that each was informed by the existing knowledge base.

Team leaders used the outcomes of team discussions and cross-team discussions, literature summaries, and the researcher's own review and integration of the literature to write a draft report of the findings. Draft reports were reviewed by knowledgeable internal experts and revisions in search strategies, interpretations of findings, and/or conclusions were made. Revised reports were reviewed by the Foundation and other outside reviewers prior to final revisions and production.

Although the wide-ranging literature searches produced reports on extensive baseline information related to Our Kids and each system component, the reports are living documents. As living documents, they bridge the creative and scientific enterprises of the past and present, and we envision the need to return to some of them for updating, extending, and drilling-down in the future.

References

About the Foundation. (n.d.). Retrieved from http://www.stupski.org/about_the_foundation.htm

American Council on Education. (2008). *Minorities in higher education 2008 twenty-third status report.* Washington, DC: Author.

College Readiness Learning System. (n.d.). Retrieved from http://www.stupski.org/college_readiness.htm

Conley, D. T. (2007). *Toward a more comprehensive conception of college readiness.* Eugene, OR: Educational Policy Improvement Center.

Cooper, H. (2009). Hypotheses and problems in research syntheses. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 19–35). New York: Russell Sage Foundation.

Cooper, H., Hedges, L., & Valentine, J. (2009). The handbook of research synthesis and meta-analysis. New York: Russell Sage Foundation.

Design collaborative. (n.d.). Retrieved from http://www.stupski.org/design_collaborative.htm

Pope, C., Mays, & Popay, J. (2007). *Synthesizing qualitative and quantitative health research: A guide to methods*. New York: McGraw Hill Education.

Research guide for CRLS: Outline of research questions for each component of the CRLS. (n.d.). San Francisco: Stupski Foundation.

Suri, H. & Clarke, D. (2009). Advancements in research synthesis methods: From a methodologically inclusive perspective. *Review of Educational Research, 79*(1), 395–430.

Stupski
Foundation

101 2nd Street, Suite 1100
San Francisco, CA 94105
Tel: 415.644.4800
Fax: 415.644.4801