Improvement Research Carried Out Through Networked Communities: Accelerating Learning about Practices that Support More Productive Student Mindsets

Anthony S. Bryk,
Carnegie Foundation for the Advancement of Teaching

David S. Yeager,
University of Texas at Austin
and Carnegie Foundation for the Advancement of Teaching

Hannah Hausman
Jane Muhich
Jonathan R. Dolle
Alicia Grunow
Paul LeMahieu
Carnegie Foundation for the Advancement of Teaching:

Louis Gomez,
University of California, Los Angeles
and Carnegie Foundation for the Advancement of Teaching

June 10, 2013
revision

A White Paper prepared for the White House meeting on
Excellence in Education: The Importance of Academic Mindsets

Please address correspondence regarding this manuscript to Anthony Bryk
(bryk@carnegiefoundation.org) or David Yeager (dyeager@utexas.edu).

Caveat

This paper is an early draft intended to share some of the improvement work ongoing at the Carnegie Foundation and the ideas that motivate it. The Carnegie Foundation for the Advancement of Teaching has embraced a “design-development” orientation. Improvement means rapidly testing ideas against both evidence and theoretical argument. This paper, like much of our daily work, enlivens a mantra of “definitely incomplete and possibly wrong” in some regards. The draft is shared in that spirit. The results illustrated in the paper are based on preliminary data analyses. They should not be quoted or cited without permission of the authors.

<u>Our Mission:</u> *The Carnegie Foundation for the Advancement of Teaching is committed to developing networks of ideas, individuals, and institutions to advance teaching and learning. We join together scholars, practitioners, and designers in new ways to solve problems of educational practice. Toward this end, we work to integrate the discipline of improvement science into education with the goal of accelerating the field’s capacity to learn to improve.*

Executive Summary

The research on academic mindsets shows significant promise for addressing important problems facing educators. However, the history of educational reform is replete with good ideas for improvement that fail to realize the promises that accompany their introduction. As a field, we are quick to implement new ideas but slow to learn how to execute well on them. If we continue to implement reform as we always have, we will continue to get what we have always gotten. Accelerating the field's capacity to learn *in and through* practice to improve is one key to transforming the good ideas discussed at the White House meeting into tools, interventions, and professional development initiatives that achieve effectiveness reliably at scale.

Toward this end, this paper discusses the function of networked communities engaged in improvement research and illustrates the application of these ideas in promoting greater student success in community colleges. Specifically, this white paper:

- Introduces *improvement research* and *networked communities* as ideas that we believe can enhance educators' capacities to advance positive change.
- Explains why improvement research requires a different kind of measures—what we call *practical measurement*—that are distinct from those commonly used by schools for accountability or by researchers for theory development.
- Illustrates through a case study *how systematic improvement work to promote student mindsets can be* carried out. The case is based on the Carnegie Foundation's effort to address the poor success rates for students in developmental math at community colleges. Specifically, this case details:
    - How a *practical theory* and set of *practical measures* were created to assess the causes of "productive persistence"—the set of "non-cognitive factors" thought to powerfully affect community college student success. In doing this work, a broad set of potential factors was distilled into a digestible framework that was useful to practitioners working with researchers, and a large set of potential measures was reduced to a practical (3-minute) set of assessments.
    - How these measures were used by researchers and practitioners for practical purposes—specifically, to *assess changes*, *predict* which students were at-risk for course failure, and *set priorities* for improvement work.
    - How we organized researchers to work with practitioners to accelerate field-based experimentation on everyday practices that promote academic mindsets (what we call *alpha labs*), and how we organized practitioners to work with researchers to test, revise, refine, and iteratively improve their everyday practices (using *plan-do-study-act* cycles).

While significant progress has already occurred, robust, practical, reliable efforts to improve students' mindsets remains at an early formative stage. We hope the ideas presented here are an instructive starting point for new efforts that might attempt to address other problems facing educators, most notably issues of inequality and underperformance in K-12 settings.

Improvement Research Carried Out Through Networked Communities: Accelerating Learning about Practices that Support More Productive Student Mindsets

Our nation's schools are, and have been for decades, in a constant state of reform.[1] By many accounts they are actually getting better. Unfortunately, our aspirations for schools are accelerating at a faster rate. Consequently, a growing chasm exists between noble aspirations and what schools can routinely accomplish.

Today we ask more of our public schools than ever before. No Child Left Behind compelled attention to the learning of all of our nation's students, not just some. The introduction of Common Core now substantially raises standards. So as a first priority, we want our schools to become more effective in advancing deeper learning for all students. Second, we live in a time where tremendous pressures exist on the public purse and will continue to do so. For several decades, education commanded increasing public resources. The conversation, however, has changed. Not only do our schools need to be much better, the whole enterprise must be more cost efficient. Put simply, we now expect schools and colleges to accomplish more with less. Third, far too many students remain disengaged, walking out the doors of our high schools and colleges and never completing their education. Turnover among teachers, principals, and superintendents is unacceptably high and morale is at an all-time low.[2]

A significant advance on any one of these three aims—greater academic effectiveness, cost efficiency, and human engagement—would be a major accomplishment. Simultaneously improving all three would be extraordinary. Yet this is precisely what our educational institutions must now do. How best to accomplish these "Triple Aims of Educational Improvement" is far from clear, but one observation does stand out. Success will continue to elude educational reformers unless we equip them and their institutions with better ways of understanding the practical problems they are facing and with more systematic approaches for addressing them.

Many are now seeing research on "student mindsets" as one helpful component for addressing disengagement and increasing learning at reduced cost. And, in fact, these ideas have demonstrated potential in carefully designed experiments. However, as we describe below, the history of school reform teaches us that even the best-tested ideas frequently fail to live up to their initial promise when used broadly. Why is this? And what can be done?

In this paper we argue that *networked communities* engaged in *improvement research* are useful for achieving such efficacy with reliability at scale. As we will show, when improvement research on academic mindsets is embedded within a broader instructional improvement effort, then important progress can be made on advancing student success. We illustrate this in the

---

[1] Cuban (1990)

[2] The 2011 MetLife survey of teachers shows morale at the lowest level in 20 years; the number of teachers who are very or fairly likely to leave the profession is 29%, up from 17% in 2009; those who are very satisfied are at 44%, down from 59% in 2009 (https://www.metlife.com/assets/cao/contributions/foundation/american-teacher/MetLife-Teacher-Survey-2011.pdf)

context of the extraordinarily high failure rates of community college students in developmental math. However, the themes outlined here are broadly relevant to other educational problems with other student populations.

Education's Modus Operandi

Over and over again, change efforts move rapidly across education, with little real knowledge as to how to effect the improvements envisioned by reform advocates (or even whether it is possible). When reformers took aim at the high drop-out rates and weak student engagement in high schools, massive effort sprung forth to create new small high schools. Little guidance existed, however, as to exactly how to transform large dysfunctional comprehensive high schools into effective small schools. Many of these efforts failed. When reformers focused attention on weaknesses in in-service professional development, a whole new organizational role—the instructional coach—was introduced into schools.[3] What coaches actually needed to know and be able to do, and the requisite organizational conditions necessary for them to carry out this work, was left largely unspecified. When reformers recognized the importance of principal leadership, significant investments were directed at intensive principal development programs.[4] Principals were urged to become instructional leaders even though demands on their time were already excessive and few or no modifications were offered to relieve the latter. The recent introduction of formal teacher evaluation protocols has greatly amplified this stress. When policymakers were unsatisfied with the rate of school improvement, high stakes accountability schemes were introduced. Unintended consequences abounded. The incidence of test-score cheating accelerated and select students were ignored, as accountability schemes directed attention to some students but not others.[5] The rapid introduction of value-added methods for assessing teachers began well before the statistical properties and limits of these methods were well understood.[6] Not surprisingly, a host of problems have emerged and political pushback is mounting. Reaching back a bit further, when corporate downsizing was the rage, school districts embraced site-based management. The actual domain for such local decision-making however was often left unclear and the necessary resources for carrying out local decisions was not provided.[7]

In each instance there was a real problem to solve, and in most cases there was at least a nugget of a good reform idea. Educators, however, typically did not know how to execute on these ideas; districts and states lacked the individual expertise and organizational capacity to support these changes at scale; and many policymakers ignored arguably the most important instrument

---

[3] Elmore and Burney (1997, 1998); Fink and Resnick (2001); Knight (2007).
[4] Fink and Resnick (2001)
[5] E.g., Jacob and Levitt (2003); Also see examples from Atlanta (http://gov.georgia.gov/press-releases/2011-07-05/deal-releases-findings-atlanta-school-probe as getting major attention, similar issues have arisen in other cities.) and Washington D.C. (http://takingnote.learningmatters.tv/?p=6232)
[6] See reports from the Gates Foundation on the MET study and critical consensus reviews at www.carnegieknowledgenetwork.org
[7] See Hess (1995); Bryk, Sebring, Kerbow, Rollow, and Easton (1998).

for any of this to work—developing will and agency for engaging these changes by our nation's teachers and principals.

In general, the press to push good ideas into large-scale use rarely delivers on the outcomes promised. In some locales a reform might work; in many places, however, it does not. At base is a common story of implementing fast and learning slow. As a field, we *undervalue learning to improve in a way that is systematic and organized*, and we lack a methodology to guide it. This should trouble all of us. If we continue to advance reform in the ways we have always done, we are likely to continue to get what we have always gotten. For a change to be successful, educators must learn how to adaptively integrate the new processes, roles, and/or materials brought forward by a reform into an already quite complex organizational system. Assuring efficacy as this adaptive integration occurs, however, is often largely ignored.

What is Improvement Research?

The central goal of improvement research is to accelerate the field's capacity to learn in and through practice to improve. We know from numerous sectors that this is key to transforming promising change ideas into initiatives that achieve effectiveness reliably at scale.[8]

The idea of improvement research taps a natural human inclination to learn by doing.[9] Informally, learning to improve already occurs in educational organizations. Individual teachers engage in it when they introduce a new practice in their classroom and then examine resulting student work for evidence of positive change. Likewise, school faculties may examine data together on the effectiveness of current practices and share possible improvement ideas. Improvement science seeks to bring analytic discipline to design-development efforts and rigorous protocols for testing improvement ideas. In this way, the "learning by doing" in individual clinical practice can cumulate in robust, practical field knowledge.[10]

Several tenets form this activity. The first is that within complex organizations *advancing quality must be integral in day-to-day work.*[11] While this principle may seem obvious on its face, it actually challenges prevailing educational practice where a select few conduct research, design interventions, and create policies, while vast others do the actual work. Second, improvement research is premised on a realization that education, like many other enterprises, actually has more knowledge, tools, and resources than its institutions routinely use well.[12] The failure of educational systems to *integrate research evidence productively into practice* impedes progress toward making schools and colleges more effective, efficient, and personally engaging. Third, improvement science embraces *a design-development ethic*. It places emphasis on learning quickly, at low cost, by systematically using evidence from practice to improve it. A central idea

[8] For an introduction to this field see Langley et al. (2010).
[9] This theme about learning in practice has a long tradition reaching back to contributions from both John Dewey (1916) and Kurt Lewin (1935).
[10] See Hiebert, Gallimore, and Stigler (2002).
[11] This is a central tenet in the Toyota Quality Management System. See Rother (2010).
[12] This problem is not peculiar to education. It is actually widespread across many different kinds of organizations. See for example Pfeffer and Sutton (2000)

is to make changes rapidly and incrementally, learning from experience while doing so. This is reflected in inquiry protocols such as the plan-do-study-act cycle.[13]

Fourth, and anchoring this learning to an improvement paradigm, is *explicit systems thinking*—a working theory as to how and why educational systems (and all of their interacting parts) produce the outcomes currently observed. These system understandings generate insights about possible levers for change. This working theory in turn gets tested against evidence from improvement cycles and consequently is revised over time. It also functions as a scaffold for social knowledge management —what a profession has learned together about advancing efficacy reliably at scale.

Fifth, improvement research is *problem-centered*. Inquiries are organized in order to achieve specific measurable targets. Data on progress toward these targets directs subsequent work. Disciplinary knowledge and methodologies are now used in the service of achieving this practical aim. In the case study we illustrate below, the "core problem" is the extraordinarily high failure rates in developmental mathematics, while the "target" involves tripling student success rates in half the time.

Finally, and arguably most importantly, improvement research maintains a laser-like focus on quality improvement. In this regard, *variability in performance is the core problem to solve.* This means attending to undesirable outcomes, examining the processes generating such outcomes, and targeting change efforts toward greater quality in outcomes for all. This pushes an empirical focus to look beyond just mean differences among groups.

The Power of Research Carried Out through Networked Communities

Improvement research can occur within individual organizations such as a school, a district, or a college. We can learn much more, and faster, when such activity is carried out through networks deliberately structured for learning through improvement research.[14] We know, for example, that networks can be a source of innovation and quickly solve problems that had once been thought difficult and even intractable.[15] We posit that translating this network structure into practical problem-solving in education would make it possible to exploit the innumerable improvement cycles already occurring. Teachers, principals, and educational leaders are experimenting almost daily, trying new approaches as they seek to "get it right" for their students. And given the size and scope of American education, this experimentation occurs on a grand scale. For just about any question one might pose, the odds are very good that someone else has already thought of it, and has even made significant headway toward solving it. Individual educators and institutions are learning every day, yet as a field, we lack capacity to organize, refine, and build on those lessons. Networked improvement communities (NICs) seek to transform these private individual efforts into tested, collective knowledge for improvement.

---

[13] Langley et al (2009)
[14] This proposition is anchored in general accounts of social learning theory in the context of formal organizations. See Engelbart (1992, 1995). In the current context see Bryk, Gomez and Grunow (2011) or Dolle et al. (2012)
[15] See Surowiecki (2005) also Nielsen (2012).

Key to actualizing this power is that each site shares common data and has a common working theory of the problem. A tool called a *driver diagram* is often used to represent this.[16] Practices that show promise in reducing a given problem in one or a small number of contexts are tested further under more diverse conditions. Those that yield inconsistent results might be refined and retried quickly at very small scale. Based on rapidly accumulating evidence, some change ideas may simply be abandoned as not workable.

The focal concern across all such inquiries is whether positive outcomes can be made to occur more reliably. *The ability to replicate quality outcomes under diverse conditions is the ultimate goal.* A Networked Improvement Community aims to unite the discipline of improvement science with the power of networks to innovate and learn. It portends a new form of a scientific community—a community engaged in disciplined inquiry about improving practical affairs.

You Cannot Improve at Scale What You Cannot Measure

The core principles of improvement science[17] have important implications for carrying out this work, and high among them is the centrality of measurement. At least three different types of measures are involved in conducting improvement work in education. First, global outcome data on problematic concerns—for example, student drop-out rates or pass rates on standardized tests—are needed to understand the scope of the problem and set explicit goals for improvement. These data sources are designed principally to be used as *measures for accountability*.

A second and different class of instruments is designed in the course of original academic research. These *measures for theory development* aim to generate data about key theoretical concepts and test hypotheses about the inter-relationship among these concepts. Such measures are also useful in the early stages of designing experimental interventions to demonstrate that in principle changing some individual or organizational condition can result in a desired outcome. Such research helps to identify change ideas that might be incorporated into a driver diagram.

Both of these types of measures, although very informative, are insufficient on their own for conducting improvement research. A third class of data, *practical measures for improvement*, is also needed to inform change efforts. For a summary of the goals and limitations of these three types of measures, see Table 1.

Measurement for Improvement (aka Practical Measurement)

The practical work of improvement introduces several new considerations that have important measurement implications. First, improvement efforts require *direct measurement of intermediary targets* (i.e., "mediators") in order to evaluate key change ideas and inform their

[16] For more on driver diagrams, see http://rd.carnegiefoundation.org/wp-content/uploads/2011/05/bryk-gomez_building-nics-education.pdf
[17] The kinds of practical inquiries illustrated are specific examples of "improvement research" i.e. practical disciplined inquiries aimed at educational improvement. The general methodology that guides these individual inquiries is referred to as "improvement science" (Berwick, 2008).

continued refinement. For example, is a student's mindset actually improving in places where a change has been introduced, and for whom and under what set of circumstances? Second, practical measurement often presses toward *greater specificity* than occurs with measurement for theory development. Educators need data closely linked to specific work processes and change ideas being introduced in a particular context. Third, increased sensitivity can be gained when measures are *framed in a language specific to the populations targeted for improvement* (e.g. adult community college students) and *contextualized around experiences common* to these individuals (e.g. classroom routines they are likely to experience). Fourth, and most significant from a practical perspective, they need to be *engineered to embed within the constraints of everyday school practice*. For example, a survey routinely given to students during regular classroom time would need to be brief—for instance, no more than 3 minutes. These are described in row 3 of Table 1 and in Table 2.

Uses of Improvement Measures

These measures serve several functions. First, practical measures assist educators in *assessing changes*; that is, they can help practitioners learn whether a change that they have introduced is actually an improvement. For this purpose measures need to be sensitive to changes in the short term and quickly accessible to inform subsequent improvement efforts.

A second use for a practical measure is *predictive analytics*. This use answers questions regarding which individuals or groups of individuals are at higher risk for problematic outcomes within a given setting. They can guide educators better to target their attention, including supplemental learning supports, in some places rather than others.

A third use for practical measures is *priority setting*. When practitioners are engaged in improvement work, they have to make choices about where best to focus their efforts. Practical measures provide empirical guidance in making these choices. Educators seek high levels of outcomes for all students. The drive for more equitable outcomes directs attention toward weakening over time the predictive relationships mentioned above. For example, as we have found in the Pathways network, students who express concerns about belonging in college at the beginning of the term are at higher risk for not completing the first semester of instruction. To the extent that this indicator remains predictive semester after semester with each new group of students, reducing belonging uncertainty becomes a direct target for subsequent improvement efforts.

Building on these broad themes, below we illustrate them in the context of an effort to address a major educational issue. This effort, carried out by the Carnegie Foundation for the Advancement of Teaching, embeds improvement research on academic mindsets within a networked improvement community working more broadly on changes to curriculum and instruction. Therefore it is a helpful case study for imagining how improvement research and networked improvement communities may be helpful for addressing student mindsets with efficacy and reliability at scale.

Table 1. Types of Measurement.

Type	Typical use	Sample research question	Common Features	Implications for psychometrics	Limitations for improving practice
Accountability	Identifying exemplary or problematic individual teachers, schools, or districts.	"Which schools should we put on probation?"	Summative, global performance measures, typically collected once a year, often toward the end of the academic year.	Extremely high reliability at the level at which you are rewarding or punishing.	Data are typically reported after school year has concluded. Students providing data do not directly benefit as they are about last year's instruction / teacher / curriculum. Causes of differences are opaque and not tied to specific practices.
Theory development and testing	Test a theory regarding the relations among two or more conceptual variables.	"Does low self-efficacy predict lower grades?"	Goal is to detect stable individual differences among students, teachers, or schools on construct of interest. Administer long, somewhat redundant surveys assessing multiple small variations on the some concept. Typically used to maximize estimated relations between latent variables.	High internal consistency reliability and construct validity as typically assessed via factor analyses. Goal of minimizing error variance in construct measure is key to goal of maximizing estimated relations between latent variables.	Impractical to administer as a part of standard practice in classrooms. Often unable to detect the effect of changes in the short term, and so not informative for rapid improvements.
Improvement	Determining whether an educational change is an improvement.	"If I do X in my classroom, will it create a sense of belonging for my most marginalized students?"	Very brief and embedded in daily work. Measure only select aspects of constructs that are an intentional focus of improvement work, that are tied to a practical theory, and explicitly signal processes that are actionable by educators. Sensitive to changes.	Focus is on predictive validity within classrooms, between classrooms, and / or between schools. Improvement goal is to drive predictive relations with course outcomes to zero. Factor structure and internal consistency reliability are not the primary concerns.	In many cases these have not yet been developed. Requires building systems (web-based or otherwise) for easy collection and rapidly reporting on data. Measures that are relevant in one context may have a different meaning in another.

Table 2. Use Cases for Improvement Measures.

Practical need	Research question	Measurement specification
Assessing changes	Did the change that I implemented actually lead to an improvement?	Repeatable measures that are sensitive to targeted changes over the short term.
Predictive analytics	Which individuals are highly at risk for the problematic outcome?	Brief, highly predictive measures that are practical to administer and quickly analyzed and reported on. Strongly related to key outcomes.
Priority setting	Which drivers of the problematic outcome continue to be at problematic levels? (And should be a subsequent improvement priority?)	Brief, highly predictive measures that are practical to administer and quickly analyzed and reported on. Strongly related to key outcomes.

A Case Study: Improving Developmental Mathematics Outcomes in Community Colleges

The United States is unique in the world in providing a redemptive path to postsecondary education through community college. Over 14 million students are enrolled in community college, seeking opportunities for a productive career and better life. Community college students are more likely to be low income, the first in their family to attend college, an underrepresented minority and underprepared for college.[18] Between 60 to 70 percent of incoming community college students typically must take at least one developmental math course before they can enroll in college-credit courses.[19] However, 80 percent of the students who place into developmental mathematics do not complete any college-level course within three years.[20] Many students spend long periods of time repeating courses and most simply leave college without a credential. As a consequence, millions of people each year are not able to progress toward their career and life goals. Equally important, these students lack command of the math that is needed to live in an increasingly quantitative age and to be critically engaged citizens.

The reasons for the low success rates are complex. Developmental math instruction often does not use research-based learning materials and pedagogic practices than can foster

[18] Bailey, Jenkins and Leinbach (2005); also see Rutschow et al. (2011).
[19] U.S. Department of Education (2008); Bailey, Jeong, and Cho (2010)
[20] Bailey, Jeong, and Cho, (2010)

deeper student learning. Traditional math curricula do relatively little to engage students' interest and demonstrate the relevance of mathematical concepts to everyday life.[21] Many students have had negative prior math experiences leading to the belief that "I am not math a person." These beliefs often trigger anxiety and poor learning strategies when faced with difficult or confusing math problems.[22] This is compounded for some students (e.g., women, African Americans) who identify as part of a group that has been stereotyped as not good at math.[23] Research also tells us that students struggle to use the language of mathematics effectively, struggle to understand problem situations that require mathematical reasoning, and communicate their learning to others orally and in writing.[24]

A Pathways Strategy

To address these long-standing challenges, the Carnegie Foundation for the Advancement of Teaching formed a network of community colleges, professional associations, and educational researchers to develop and implement the Community College Pathways (CCP) program. The program is organized around two structured pathways, known as Statway™ and Quantway.™ Rather than a seemingly random walk through a maze of possible course options,[25] students and faculty are now joined in a common, intensive year-long experience toward *ambitious learning goals* that culminate in the awarding of college math credit. Statistics and quantitative reasoning, respectively, provide the conceptual organizers for the Pathways. Both Pathways emphasize the core mathematics skills needed for work, personal life, and citizenship. They stress conceptual understanding and the ability to apply it in a variety of contexts and problems. Developmental mathematics objectives are integrated throughout.

Three research-based principles vitalize the Pathways instructional design:

1. **Productive struggle**. As shown by Hiebert and Grouws,[26] students are more likely to retain what they learn when they expend effort "solving problems that are within reach and grappling with key mathematical ideas that are comprehendible but not yet well formed."[27] Consequently each new subject matter topic begins with a rich problem that engages students' thinking and stimulates this struggle to understand.
2. **Explicit connections to concepts**. Sometimes math is taught with a focus on procedural competence at the price of advancing real conceptual understanding.[28]

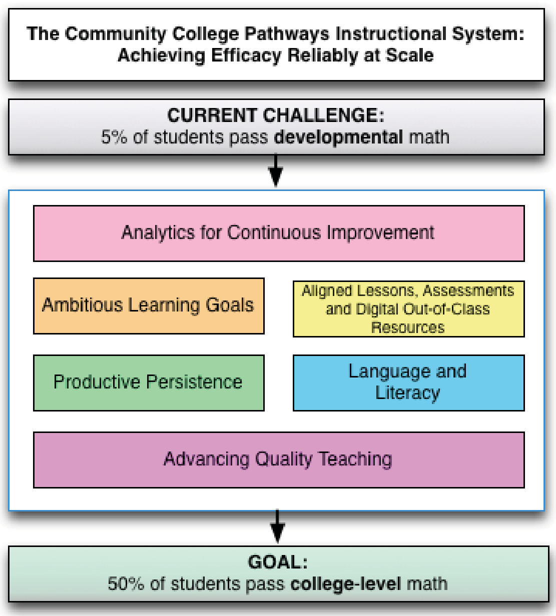
[21] Carnevale and Desrochers (2003); also see Hulleman & Harackeiwicz (2009).
[22] Blackwell, Trzesniewski, and Dweck (2007); Haynes, Perry, Stupinsky and Daniels (2009); Beilock, Gunderson, Ramirez, and Levine (2010)
[23] Cohen, Garcia, Purdie-Vaughns, Apfel and Brzustoski (2009); Walton and Spencer (2009).
[24] Gomez, Lozano, Rodela, and Mancervice (2012); Schoenfeld (1988)
[25] Zeidenberg and Scott (2011)
[26] Hiebert and Grouws (2007)
[27] Schmidt and Bjork (1992)
[28] Boaler (1998)

Research suggests making explicit connections between mathematical or statistical facts, ideas, and procedures can improve both conceptual and procedural understanding.[29]

3. **Deliberate practice**. Classroom and homework tasks are designed to overcome gaps in understanding, apply what is learned, and deepen facility with key concepts.[30] Deliberate practice eschews rote repetition for carefully sequenced problems developed to guide students to deeper understanding of core concepts.[31]

These three design principles are actualized in the *Lessons, Assessments, and Out-of-Class Resources* that form the curriculum for each pathway.

Figure 1. A Conceptual Model for the Carnegie Foundation for the Advancement of Teaching's Pathways Initiatives.



Four additional supports complement this instructional core of ambitious goals and aligned instructional materials. First, integrated throughout the Pathways is an evidence-based package of student activities and faculty actions, called *Productive Persistence* that aims to increase student motivation, tenacity, and learning skills for success. Strategies focus on reducing student anxiety,[32] increasing their sense of belonging[33], and strengthening their belief that they can learn math (i.e. countering the fixed mindset beliefs).[34] Specific activities focus on developing the skills needed to be effective students and the flexible mindsets necessary to utilize those skills. These include targeted

[29] Hiebert and Grouws (2007)
[30] Ericsson (2008); Ericsson, Krampe, and Tescher-Römer (1993)
[31] Pashler, Rohrer, Cepeda, and Carpenter (2007)
[32] Jamieson, Mendes, Blackstock, and Schmaeder (2010); Ramirez and Beilock (2011)
[33] Walton and Cohen (2011)
[34] Dweck (2006)

student interventions, guidance to help faculty create more supportive classroom environments, and a lesson structure that encourages active student engagement.

Second, given students' diverse backgrounds, Pathways also attend to the *Language and Literacy* demands in materials and classroom activities, and supports are interwoven so that learning is accessible to all. Third, informing continuous evidence-based improvement is a *Rapid Analytics* capacity designed to focus attention on what is (and is not) working, where and for whom, and under what set of circumstances. While on-average improvements are important, the NIC goal is efficacy in every college, in every classroom, and for all of the diverse sub-groups of students who enroll. This component provides empirical feedback informing ongoing efforts toward these targets. Fourth is an *Advancing Quality Teaching* component. The Pathways programs aim to provide instructors with the knowledge, skills, and supports necessary to experience efficacy in initial use, to develop increasing expertise over time, and to engage the larger networked community in research on improving their collective practice.

Focusing in on Productive Persistence

Productive Persistence refers to the "non-cognitive factors" that allow a student to successfully complete their academic coursework—the *tenacity* to persist and the *learning strategies* to do so productively. As we outline below, one key component of Productive Persistence is a set of explicit academic mindsets. In the remainder of the paper, we illustrate some early uses of these data in guiding improvement efforts.

As noted above, much important research has focused on identifying non-cognitive factors that *predict* student outcomes.[35] The critical practical question, however, is how these factors can be affected *reliably, at scale, and by diverse practitioners working in diverse settings?* There is a limited set of precise psychological interventions that have demonstrated effects in randomized experiments,[36] but they have almost never been tested with community college students, and, at least in the published literature, have only been tested at a small scale.[37] There are also many related interventions that have been the subject of high-quality evaluations—such as "learning communities,"[38] intensive mentoring,[39] or comprehensive student success courses.[40] On average, however, these have had little or no effects on student performance or credit attainment beyond the treatment period. There are also many clinical rules of thumb about effective student engagement tactics but little evidentiary basis for these practices too.

More generally, there is no coherent framework that unites researchers and practitioners. Instead, there is a "Tower of Babel" about student success, where academics and

[35] For reviews, see, e.g., Farrington et al. (2012); Dweck, Walton, and Cohen (2011)
[36] See Yeager and Walton (2011)
[37] Though for unpublished data see Yeager, Paunesku, Walton, and Dweck (2013)
[38] Weiss, Visher, Wathington, Teres, and Schneider (2010).
[39] Visher, Butcher, and Cerna (2010)
[40] Rutschow, Cullinan, and Welbeck (2012)

practitioners use different names for the same concepts or the same name for very different interventions. How then can one frame the "non-cognitive" field in a way that leads to productive improvement—that brings some coherence and organization to an effort to try to dramatically and systematically improve student outcomes in this domain?

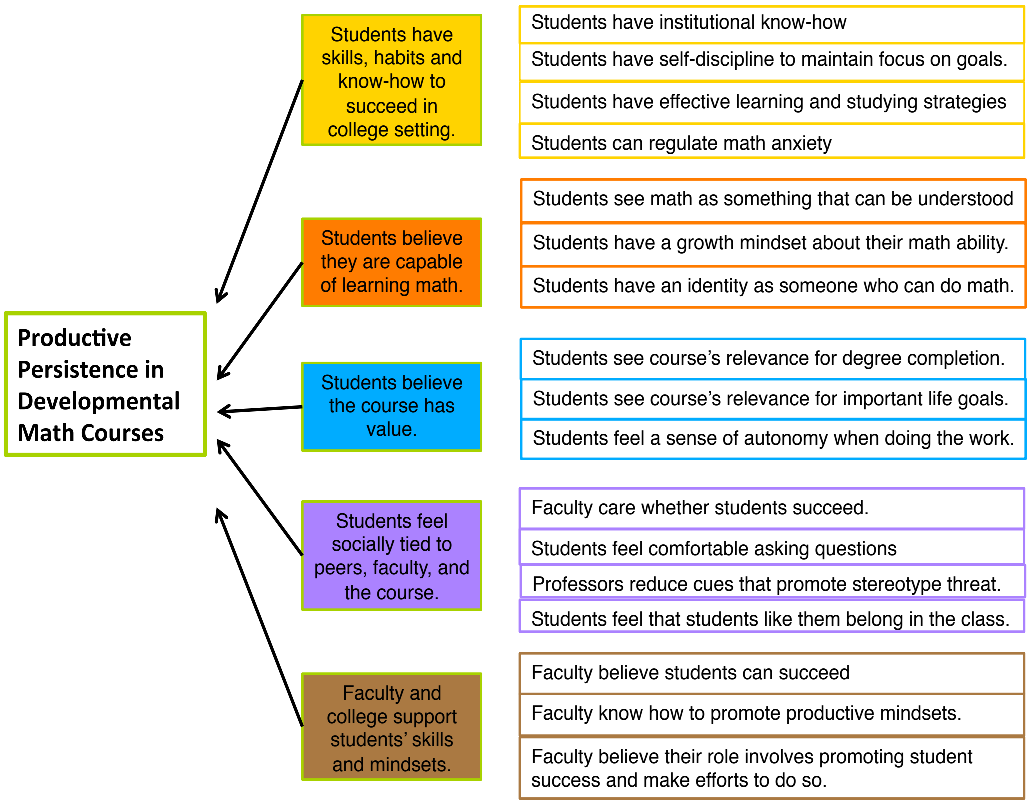
In this context, the Carnegie Foundation's "Productive Persistence" work sought to create a practical theory that can provide a common language to guide improvement. It then sought to develop and test practical measures of the components of that practical theory, so that researchers and practitioners can build an evidentiary basis for practices that reliably improve student motivation and engagement.

Creating a Practical Theory to Guide Improvement

A first step to beginning the improvement work was to create a "practical theory" that researchers and practitioners could agree on and then represent it in a shared concept map. A practical theory is not a *disciplinary* theory, in that it does not seek to document novel features of human psychology or social structures that shape the ways people in general think or behave. Instead, it draws on both the wisdom of practice as well as insights from academic theories to guide practice improvement. While disciplinary theories emphasize novelty, counter-intuitiveness or fine distinctions—and as a result have a highly important role in science—a practical theory uses only those distinctions or novel ideas that can reliably motivate action to solve practical problems in diverse contexts among diverse practitioners. A practical theory is also not a *general educational* theory. It is not designed to be an account of all relevant problems (for instance, motivation among students of all levels of ability or of all ages). Rather, it is co-created with community college practitioners and tailored for the challenges faced specifically by developmental math students. Nevertheless, what might be learned from efforts to create practical theories may be informative both for disciplinary theories and for practical theories of other educational problems. The current productive persistence map consists of five key sub-concepts detailed in Figure 2.

To re-iterate, the virtue of a practical theory is not that it is new or non-obvious. To the contrary, the virtue of a practical theory is that each element is immediately recognizable to both practical experts and theoretical experts, each of whom deeply understands the problem of practice through their own lens. This is important, we believe, for whether the theory is a useful guide for practice improvement and whether it will be adopted by both communities.

Figure 2. A Concept Map for Productive Persistence in Developmental Math



Creating a Practical Measure

After identifying and refining the five conceptual areas represented in Figure 2, a next step was to create a set of practical measures to assess each. Because many of the ideas in the driver diagram had come from the academic literature, there were measures available for each. A comprehensive scan of the field located roughly 900 different potential survey measures of the constructs included in our productive persistence driver diagram.

By and large, however, these measures failed the test of practicality. For instance, no faculty member is likely to administer a 900-item battery as a part of their regular classroom routines. More concretely, many items were redundant, theoretically-diffuse, double-barreled questions with vocabulary that was confusing for respondents learning English or with low cognitive ability. Predictive validity, a primary criterion for a practical measure, was rare. For instance, an excellent recent review of existing non-cognitive measures[41] located 196 survey instruments coming from 48 empirical articles. We led a team of coders who reviewed each of these. We could locate objective validity evidence (i.e., correlations with test scores or official grades) for only 12, or roughly 6 percent of measures. Administration in community college populations was even more rare; we could find only a tiny set of papers that measured the drivers identified in our practical theory and showed relations to objective course performance metrics among developmental mathematics students. Of course, many of these past measures were not designed for improvement; they were designed to test theory and as such were often validated by administering them to large samples of captive undergraduates at selective universities. Practical measurement, by contrast, has different purposes and therefore requires new measures and different methods for validating them. Therefore, our team took the list of 900 individual survey items and reduced them to roughly 26 items that, in field tests with community college students, took an average of three minutes to answer.

Guided by theory. We were able to create practical measures in part by looking to the experimental literature to learn what effectively changes one or more of the core elements in the concept map. We then re-wrote items so that they tapped more precisely into the causal theory. For instance, while an enormous amount of correlational research has focused on the impact of "social connections" for motivation,[42] some experimental literature focuses more precisely on a concept called "belonging uncertainty" as a cause of academic outcomes in college.[43] The theory is that if a person questions whether they belong in a class or a college, it is difficult to fully commit to the behaviors that may be necessary to succeed, such as joining study groups or asking professors for help. Of significance to practical measurement, it has been demonstrated that an experimental intervention can mitigate the negative effects associated with this mindset.[44] Such experimental findings provide a basis for item reduction. Instead of asking students a large number of overlapping items about liking the school, enjoying the school, or fitting in at the school, our practical measure asked a single question: "When thinking of your math class, how often, if ever, do you wonder: Maybe I don't belong here?" A similar process was repeated for each of the sub-concepts shown in Figure 2. Possible items for inclusion were then further scrutinized and refined through a comprehensive process that applied principles of optimal survey item design illustrated below.[45]

Contextualizing and pre-testing. After an initial period of item writing, the survey items next went through a process of customization to the perspectives of community college practitioners and students. Because practical measures must be brief, we did not have the luxury of asking items that large portions of students would misconstrue, or that large numbers of faculty would find irrelevant. Just as a political opinion poll needs to pass the

---

[41] Atkins-Burnett, S, Fernandez, C, Jacobson, J, & Smither-Wulsin, C. (2012); for a similar review see U.S. Department of Education (2011).
[42] e.g., Wentzel and Wigfield (1998)
[43] Walton and Cohen (2007)
[44] Walton and Cohen (2011)
[45] Krosnick (1999); Schuman and Presser (1981); see, Yeager, Bryk, and Hausman (2013).

test of face validity with the public who may or may not want to agree with the findings, our practical measures, in order to help faculty critically evaluate their practice, needed to be interpretable, face valid, and actionable, from the perspective of practitioners. Moreover, following best practices, we also conducted cognitive pre-tests with current developmental math students to surface ambiguities or equivocations in the language.[46] We paid special attention to how the items may have confused the lowest performing students or students with poor English skills—both groups that would be especially likely to under-perform in developmental math, and therefore groups that ideally the practical measures would help us learn the most about how to help.

Embedding in the instructional system. After this process and some initial piloting, the brief set of measures was embedded in the Pathways online instructional system (a website with the textbook and homework). Students were directed by faculty to complete the items as part of the first day either in class or at home. In this way, drivers of students' motivation and engagement can be assessed efficiently and practically on a regular basis.

Illustrative Examples Using Practical Measurement to Improve

Assessing change. A first use for practical measures is to assess whether changes implemented in the classroom were, in fact, linked to improvements in students' reports about Productive Persistence. An assumption in improvement research is that there will be variability in local practice and there will be variability in results. The challenge for improvement researchers is to measure both of these so as to learn how to change practice in ways that reduce variability in performance and create quality outcomes for all.

Our example centers on assessing whether a package of activities for "Starting Strong" successfully led to changes in Productive Persistence in the first few weeks of a developmental math instruction. As noted, both practitioners and researchers contend that the first few weeks of a term are a crucial "sensitive period" for student engagement. When students draw early conclusions that they cannot do the work or that they do not belong then they may withhold the effort that is required to have success in the long term, starting a negative recursive cycle that ends in either course withdrawal or failure.[47] Similarly, in the first few class periods, students join or do not join study groups that can be an important resource for course success, after which time membership tends to close. This informal network participation is also predictive of student learning over time.[48] Consequently, we theorized that that if faculty successfully created a classroom climate that helped students see their academic futures as more hopeful and that created strong social ties to peers and to the course, students might gradually put forth more effort and, seeing themselves do better, might show an upward trajectory of learning and

[46] For instructions, see, e.g., Presser, et al., (2004).
[47] Cook, Purdie-Vaughns, Garcia, and Cohen (2012)
[48] Vaquero and Cebrian (2013)

engagement.[49] One key "Starting Strong" activity was a direct-to-student *growth mindset intervention* delivered either in class as a worksheet or via the Internet during the first week of the course. The growth mindset intervention is a precise, brief (30 minutes) theory-based persuasive reading and writing exercise that is designed to powerfully shift students' beliefs away from the view that being a "math person" or not is something that is fixed and toward the view that math ability can be grown and developed. The customization of this activity for developmental math students in community college and its evaluation in randomized experiments has been described elsewhere.[50] A second bundle of activities consisted of classroom routines in the form of a "Starting Strong" Package. This consisted of a set of initial classroom routines targeted toward reducing anxiety, increasing interest in the course and forming supportive students' social networks.

Table 3. Productive Persistence Use Cases for Practical Measures

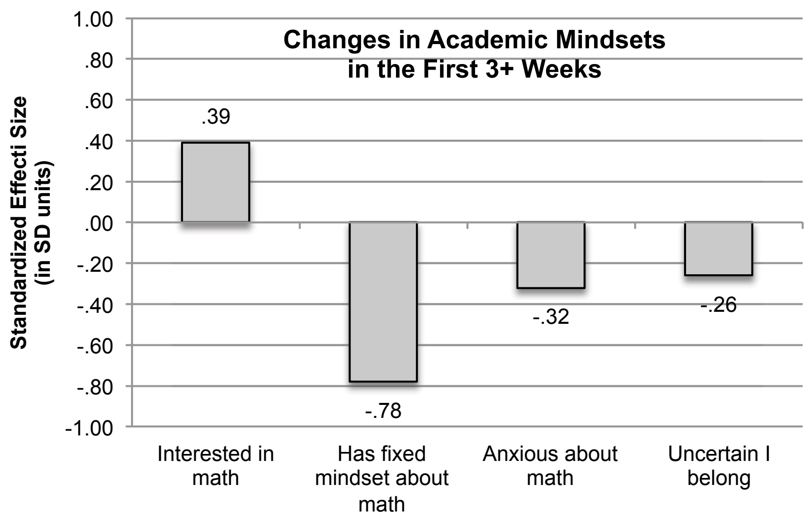
Practical need	Research question	Example
Assessing changes	Did the change that I implemented actually lead to an improvement?	Analysis of changes in targeted productive persistence objectives (Figure 3)
Predictive analytics	Which of my students have characteristics that put them at risk for failing?	Develop and deploy an "at-riskness" indicator (Figure 4)
Priority setting	Which sub-groups of students may still fail to benefit from the program? (And what should I target as a subsequent improvement priority?)	Students who are still uncertain about their belonging one month into the course are likely to drop out and fail (even after controlling for other factors) (Figure 5)

As a first look at efficacy, we examined the Productive Persistence survey on the first day of class and after three weeks of instruction. Evidence on the efficacy of the Productive Persistence Starter Package, presented in Figure 3, was encouraging. The results, presented in standardized effect sizes, show moderate to large changes in four measured student attitudes after the first three weeks of exposure to Statway. As instruction began, students' interest in math increased, their beliefs about their inability to learn math (i.e. a fixed mindset) decreased, math anxiety decreased, as did their uncertainty about

[49] For a psychological analysis, see Garcia and Cohen (2012)
[50] Yeager, Paunesku, Walton, and Dweck (2013)

belonging. However, these effects did not occur in every college and for every sub-group of students; these results, in conjunction with predictive validity findings presented below, informed subsequent improvement priority setting.

Figure 3. Initial Evidence on the Efficacy of the Productive Persistence Starter Package.

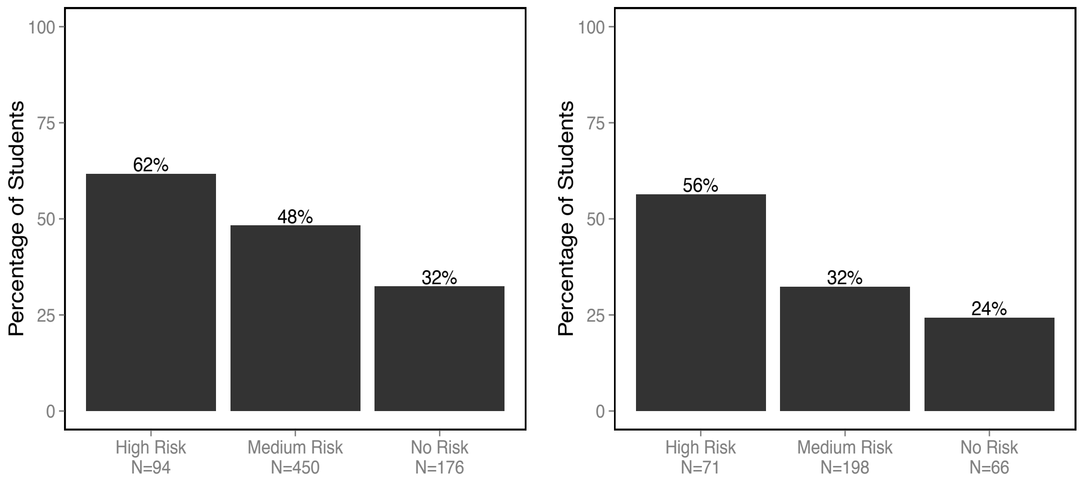


Also noteworthy, the Productive Persistence survey items explain 40 percent of the variance in the baseline measure of students' mathematics skills collected on the first day of class. These questions also predicted student performance on online homework tasks completed during the first two months and students' persistence in use of this platform, even after controlling for baseline math test results.

Predictive analytics. Predicting a student's probability of successfully completing the course is another key use for our practical measure of Productive Persistence. We posited that if faculty could identify early on those students at-risk for failure, tailored interventions to increase the likelihood of success might be possible. For this purpose, we developed an "at-riskness" indicator based on student responses to the Productive Persistence questions asked on the first day of the course.

Data from three of the sub-concepts shown in Figure 2 were used to form the at-riskness indicator: (1) Skills and habits for succeeding in college; (2) Fixed mindset about math ability; and (3) Mindsets about social belonging. Data on students' interest in the course were not included in the at-riskness indicator. Although the Pathways were deliberately re-designed to provide meaningful and interesting material, students would not be expected to provide meaningful information about this on the first day of the course. The measures about faculty's mindsets and skills were also not included in the at-riskness indicator since our objective was to understand variance in *student* risk factors *within* classrooms, not risk factors at the teacher level. We empirically derived cut points that signaled problematic versus non-problematic responses on five different risk factors embedded within the three domains. We then summed the number of at-risk factors to form an overall at-riskness score ranging from 0 to 5.[51]

Figure 4. Productive Persistence At-Riskness Indicator Predicts the Percentage of Students Who Fail the End-of-Term Common Assessment With a Score of ≤ 60%.



(a) Statway ($\chi^2(2)\, p < 0.001$) (b) Quantway ($\chi^2(2)\, p < 0.001$)

High risk = 3, 4, or 5 risk factors; *Medium risk* = 2 or 1 risk factors; *No risk* = 0 risk factors.

As illustrated in Figure 4, this analysis produced striking differences in student success rates as a function of the productive Persistence indicator. Students with high risk on day 1 were twice as likely to fail as compared to low-risk classmates. Testifying to the robustness of these findings, these results replicated in both the Statway and the Quantway colleges, and the day 1 at-riskness indicator predicted end-of-course exam performance even when controlling for background mathematical knowledge and student demographic characteristics such as race/ethnicity, home language and number of dependents.[52]

Priority setting. As noted earlier, a third use of practical measures is to identify priorities for subsequent improvement activity. To do this, we examined how the separate components, comprising the Productive Persistence survey, predicted the outcomes of interest. In varying degrees, each of these components was a target for change in the initial Starting Strong Package. Any component that continued to strongly predict outcomes, despite initial design efforts to ameliorate its effects, is a possible target for further improvement activity. We found that one survey item, assessing belonging uncertainty, was the single best predictor of whether students dropped the course before the end of the semester, even after controlling for background math knowledge and

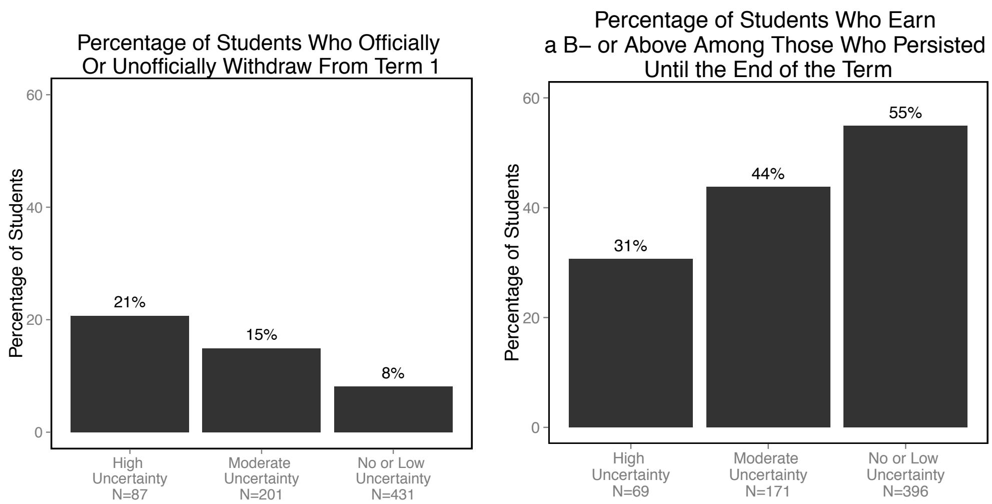
---

51 For detailed appendices describing how this was done, see Yeager, Bryk, and Hausman (2013)

52 See Yeager, Bryk, and Hausman (2013) for statistical models.

demographic-personal characteristics such as race/ethnicity, income, number of dependents in the home, and number of hours worked (See Figure 5). Furthermore, among students who did not withdraw from the course, this item was also an excellent predictor of whether students met the minimum threshold for being prepared for subsequent math coursework (a B- or better).[53]

Figure 5. Relation of Single-Item Measure of Belonging Uncertainty to Course Outcomes (Results combined for the Statway and Quantway NICs).



Belonging uncertainty survey item: "When thinking about the Statway [Quantway], how often, if ever, do you wonder 'Maybe I don't belong here?'" Response coding: *No or low uncertainty* = "Never" or "Hardly ever"; *Moderate* = "Sometimes"; *High uncertainty* = "Frequently" or "Always." χ^2 (2) tests p < .0001.

Pulling it All Together: Using Practical Measures in Networked Communities to Accelerate Learning to Improve

Hub Analytics: Looking Across the Network

Networked improvement communities require a coordinating hub.[54] A core capacity of the hub is to collect, manage, analyze, and share data across the network. This plays several important improvement functions. First, common measures are critical for understanding whether or not local changes are, in fact, improvements. This data helps explain performance differences across classes, instructors, colleges, regions, and the network at large. The Pathways analytic infrastructure is designed to answer these questions. A second, closely related function of the hub analytic capacity is analyzing variation in performance to predict how a change to curriculum, instructional practice, or

[53] See Yeager, Bryk, and Hausman (2013) for statistical models.
[54] Bryk, Gomez, and Grunow (2011)

the online instructional platform is likely to affect student outcomes when tried in a new classroom or college. In doing so, the hub is regularly examining three major sources of variation in outcomes: characteristics of students, classroom instruction, and institutional context. In essence, this infrastructure is the backbone of a decentralized, network-wide learning system with the capacity to leverage an immense amount of system data to guide improvements. Third and more generally, the hub takes the lead for the network in developing common frameworks, tools, and routines that are critical to coordinating continuous improvement efforts across a dispersed professional community. These artifacts also facilitate network learning and engagement that is essential to scaling improvement within an education system.

"Alpha Labs:" Conducting Disciplinary Research on Problems of Practice

The primary focus of this white paper has been on how improvement research builds on disciplinary knowledge and turns it toward problems of practice improvement. In a broad sense it has been about extracting the best of disciplinary knowledge and driving this toward practical ends.

The development and integration of a student growth mindset intervention was the first foray of this sort in the community college NIC. This work continues under what we now call an "Alpha Lab." The goal of the Carnegie Alpha Lab Research Network, led by James Stigler of UCLA, is to engage academic researchers from diverse backgrounds who may be interested in addressing high leverage problems identified by the NIC. Within the context of Productive Persistence, three such priorities now exist: 1) activities that might reinforce and extend effects of initial mindset interventions throughout the academic year (i.e., "boosters" of the mindset intervention); 2) interventions that might reduce/moderate student test anxiety; and 3) efforts to reduce belonging uncertainty. By providing a structure through which researchers can work on problems and priorities set by the Pathways, the Alpha Labs draw on cutting edge research to deepen understanding of problems and test theory-based solutions. The network hub at Carnegie facilitates relationships with Pathways colleges, provides access to Pathways data, helps identify funding sources, assists in research grant writing, and provides forums for sharing research finding.

Improving Clinical Routines in Support of Productive Persistence

Complementary improvements are also emerging out of practice. Whether the positive effects of direct interventions to address student mindsets persists over time is likely to depend upon students subsequent instructional experiences. Logically, if a designed intervention can change a mindset, subsequent experiences can as well. So in response, the community college NIC has organized to develop and improve key classroom routines thought critical to supporting Productive Persistence. The lead for change now is clinical rather than academic. But even here the NICs are drawing on theory-based insights to guide this design-development. Based on year 1 results, the current focus is on improving classroom routines that strengthen students' experiences of belonging and thereby attenuate belonging uncertainty (i.e. the improvement priority noted above).

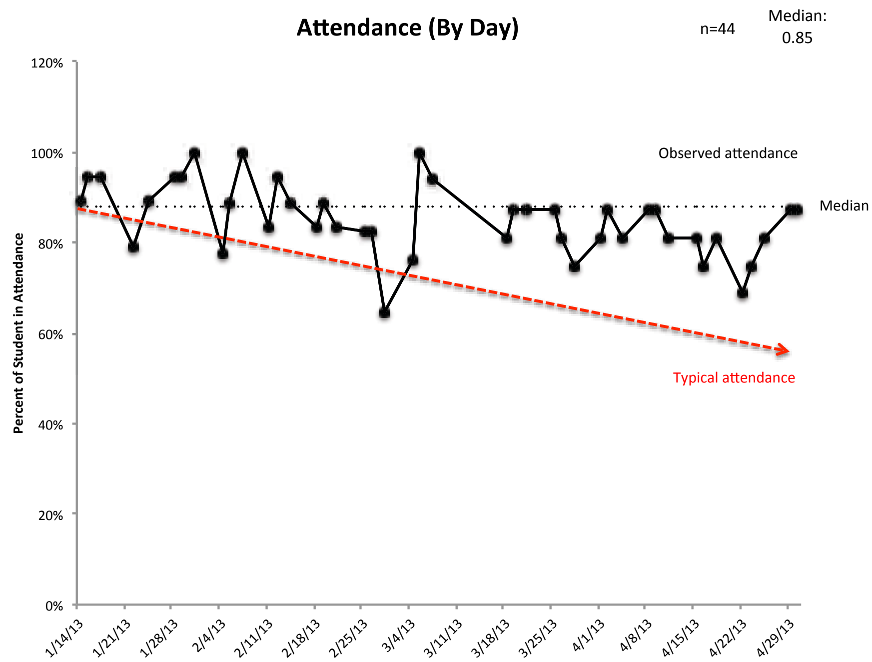
This past year network faculty have initiated improvement research on three change ideas to address mindsets about belonging / belonging uncertainty:

1. Faculty typically do not have a systematic way of reaching out to students who might miss a class. This hampers their ability to understand why students have been absent and limits their ability to encourage students to attend future classes. In response to this, a network faculty member developed and tested **routines and scripts for emailing absent students**. The script changes over the course of the semester as the relationship between the faculty member and the student evolves. The micro-dynamics of these routines were tested through plan-do-study-act (PDSA) cycles and revisions were made over time. The faculty member found that attendance was strong and has recommended that this script be further tested in multiple classrooms.

2. Another faculty member sought to build a sense of belonging by making students responsible for each other's presence in what was called a **"group noticing routine."** The routine consists of three distinct stages tested throughout the semester. In the first stage, the faculty member groups students and gets them to know each other outside of the immediate math content. In the next stage, groups are responsible for reporting to the faculty who is absent each day. In the final stage, groups take responsibility for contacting students who are missing in order to encourage them to attend future classes and give them any materials or information that they missed from class. Attendance remained strong across the semester (85 percent median attendance rate) and quite different from past experiences with similar student groups.

3. In the initial Starting Strong Package, faculty members were simply advised to give roles (e.g. monitor, reporter, facilitator) to students in the group. Through PDSA cycles, NIC faculty developed and tested a routine for **effective group role functioning.** During group work, students are given laminated cards that describe the expectations for their assigned role. The roles rotate throughout the course. A protocol was added that the student taking on the group facilitator role would also report on the performance of each student relative to the role they were assigned to play that day. The scores are then given to the faculty and incorporated into classroom participation grades. This process underwent initial testing by one faculty member before being tested in a different classroom. The two faculty members found that students worked together more effectively in the classroom and that attendance was strong (92 percent median attendance rate).

Figure 6. A Sample "Run Chart" for a Field-based PDSA Cycle Testing a "Group Noticing Routine"



All three of these routines appear promising in their initial tryout as demonstrated in run charts such as above. (Figure 6 is from the tryout of the noticing routine.) Given these initial encouraging results, all three routines will now be shared with other faculty this summer, tested by additional faculty under more diverse conditions, and likely refined. If comparable results continue to occur as the routines are picked up by more faculty and used across more varied settings, these practices would eventually become part of a common kernel of routines supported network-wide.

Conclusions

Education reformers are rightfully enthusiastic about the potential for research on mindsets to contribute to the improvement of student outcomes. Our emphasis on creating networks engaged in improvement research related to mindsets is not based on shortcomings in the academic literature but rather on the observation that there have been *many* promising ideas in the history of education, and, shockingly, very *few* examples of ones that have been successfully implemented with reliability at scale.

We have proposed that improvement research can be a helpful way forward. We have shown that it is possible to develop an understandable framework for the broad and disconnected field of "student success"–what we have titled "Productive Persistence." Next, we have shown that it is possible to adapt measures that were originally developed for research and use them in the context of improvement work on a new developmental math course. Importantly, even brief but fine-tuned measures could be highly predictive of course outcomes and help the networked improvement community assess changes, identify students at-risk, and set improvement priorities.

Finally, we note that this improvement work is only beginning. Through Alpha Lab partnerships, we expect that more disciplinary experimental work will further refine our theories and contribute additional insights and interventions to address more fully the psychological barriers contributing to underperformance. And through the work of faculty engaged in improvement research, this too will generate new, related practices that have potential to promote Productive Persistence. To do so, improvement measures—for instance ones that supplement optimized self-reports with direct assessments of behaviors that signify Productive Persistence—will need to continually be developed. While this work is nascent, we hope it provides some guidance for how the field might learn more quickly about academic mindsets in and through practice.

We end with a reminder that the goal of *any* educational change is to promote greater learning among many more students. Our hope in providing an integrated view of improvement research and networked communities is to illustrate how research on mindsets can be problem-centered—focused first and foremost on how to effectively address student underperformance at scale, as a part of more systematic efforts to improve the educational system. The Carnegie Foundation's NICs have attempted to *both* give students the educational experience they need—in terms of improved curriculum and instruction—*and* simultaneously address mindsets that might prevent students from fully benefitting from those changes—such as their fixed mindsets about ability, their feelings about not belonging, or their anxiety about math and statistics. We hope that this case study inspires others to consider analogous work in other settings—most obviously in K-12 schools—so that as a community of researchers and practitioners we can more effectively educate all of our students and expand our nation's human potential.

References

Atkins-Burnett, S, Fernandez, C, Jacobson, J, & Smither-Wulsin, C. (2012). *Landscape analysis of non-cognitive measures*. Princeton, NJ: Mathematica Policy Research.

Bailey, T., Jenkins, D., & Leinbach, T. (2005). *What we know about community college low-income and minority student outcomes: Descriptive statistics from national surveys*. New York, NY: Teachers College Community College Research Center, Columbia University. Retrieved May 13, 2013 from http://www.eric.ed.gov/PDFS/ED484354.pdf

Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review, 29*, 255-270.

Berwick, D. M. (2008). The science of improvement. *The Journal of the American Medical Association 299*, 1182-1184.

Blackwell, L. S., Trzesniewski, K. H. & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*, 246-263.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, USA*, *107*, 1060-1063.

Boaler, J. (1998). Opened and closed mathematics: Student experiences and understanding. *Journal for Research in Mathematics Education, 29*, 41-62.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In Hallinan, M. T. (Ed.), *Frontiers in Sociology of Education* (pp. 127-162). The Netherlands: Springer.

Bryk, A., Sebring, P. B., Kerbow, D., Rollow, S., & Easton, J. (1998). *Charting Chicago school reform: Democratic localism as a lever for change.* Boulder, CO: Westview Books.

Carnavale, A. and Desrochers, D. (2003). *Standards for what?: The economic roots of K-16 reform.* Princeton, NJ: Educational Testing Service.

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science, 324*, 400-403.

Cook, J. E., Purdie-Vaughns, V., Garcia, J., & Cohen, G. L. (2012). Chronic threat and contingent belonging: Protective benefits of values affirmation on identity development. *Journal of Personality and Social Psychology, 102*, 479-496.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19*, 3-13.

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York, NY: Macmillan.

Dolle, J., Wardrip, P., Russell, J. L., Gomez, L. & Bryk, A. (2012, June). *The power of improvement networks to transform educational inquiry: A preliminary exploration*, Poster session presented at the International Conference of the Learning Sciences, Sydney, Australia.

Dweck, C.S. (2006). *Mindset*. New York, NY: Random House.

Dweck, C.S., Walton, G.M., & Cohen, G. (2011). *Academic tenacity*. White paper prepared for the Gates Foundation. Seattle, WA.

Elmore, R. F., & Burney, D. (1997). *Investing in teacher learning: staff development and instructional improvement in Community School District #2,* New York City. Washington, DC: National Commission on Teaching and America's Future.

Elmore, R., & Burney, D. (1998). *Continuous improvement in Community District# 2, New York City.* Pittsburgh, PA: High Performance Learning Communities Project, Learning Research and Development Center, University of Pittsburgh.

Engelbart, D. C. (1992). *Toward high-performance organizations: A strategic role for Groupware.* Groupware '92. San Jose, CA: Morgan Kaufman Publishers.

Engelbart, D. C. (1995). Toward augmenting the human intellect and boosting our collective IQ. *Communications of the ACM, 38*, 30-32.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine. 15*, 988-994.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363-406.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago, IL: University of Chicago Consortium on Chicago School Research.

Fink, E., & Resnick, L. B. (2001). Developing principals as instructional leaders. *Phi Delta Kappan, 82*, 598–606.

Garcia, J. & Cohen, G. L. (2012). A social-psychological approach to educational intervention. In E. Shafir (Ed.), *Behavioral foundations of policy*, (pp. 329-350). Princeton, NJ: Princeton University Press.

Gomez, K., Lozano, M., Rodela, K., & Mancervice, N. (2012, November 8-11). Increasing access to mathematics through a literacy language lens. *Paper presented at the American Mathematical Association of Two-Year Colleges (AMATYC), Jacksonville, Florida*.

Haynes, T. L., Perry, R. P., Stupnisky, R. H., & Daniels, L. M. (2009). A review of attributional retraining treatments: Fostering engagement and persistence in vulnerable college students. In Smart, J. C. (Ed.), *Higher education: Handbook of theory and research* (pp. 227-272). New York, NY: Springer.

Hess, G. A. (1995). *Restructuring urban schools: A Chicago perspective.* New York, NY: Teachers College Press.

Hiebert, J., Gallimore, R., & Stigler, J. (2002). A knowledge base for the teaching profession: What would it look like, and how can we get one? *Educational Researcher, 31*, 3-15.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In Lester, F. K. (Ed.), *Second Handbook of research on mathematics teaching and learning,* (pp. 371-404). Charlotte, NC: National Council of Teachers of Mathematics.

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326,* 1410-1412.

Jacob, B., & Levitt, S. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, *118*, 843–877.

Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology, 46*, 208-212.

Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction*. Thousand Oaks, CA: Corwin Press.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567.

Langley, G. J., Moen, R., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance* (2nd ed.). San Francisco, CA: Jossey-Bass.

Lewin, K. (1935). *A dynamic theory of personality: Selected papers*. New York, NY: McGraw-Hill.

Nielsen, M. A. (2012). *Reinventing discovery: The new era of networked science*. Princeton, NJ: Princeton University Press.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14*, 187-193.

Pfeffer, J., & Sutton, R. (2000). *The knowing-doing gap: How smart companies turn knowledge into action*. Cambridge, MI: Harvard Business School Press.

Presser, S., Rothgeb, J., Couper, M., Lessler, J., Martin, E., Martin, J., & Singer, E. (2004). *Methods for testing and evaluating survey questionnaires*. New York, NY: John Wiley.

Ramirez, G., & Beilock, S. (2011). Writing about testing worries boosts exam performance in the classroom. *Science, 331*, 211–213.

Rother, M. (2010). *Toyota kata: Managing people for improvement, adaptiveness, and superior results.* New York, NY: McGraw Hill.

Rutschow, E. Z., Cullinan, D., Welbeck, R. (2012). *Keeping students on course: An impact study of a student success course at Guilford Technical Community College.* New York, NY: MDRC. Retrieved May 13, 2013 from http://www.mdrc.org/sites/default/files/Keeping%20Students%20on%20Course%20Full%20Report.pdf

Rutschow, E. Z., Richburg-Hayes, L., Brock, T., Orr, G., Cerna, O., Cullinan, D., ... & Martin, K. (2011). *Turning the tide: Five years of achieving the dream in community colleges.* MDRC. Retrieved May 13, 2013 from http://www.mdrc.org/sites/default/files/full_593.pdf

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well taught" mathematics courses. *Educational Psychologist, 23*, 145-166

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. New York, NY: Academic Press.

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Random House.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. (2011). *Measuring student engagement in upper elementary through high school: A description of 21 instruments.* (Issues & Answers Report, REL 2011–No. 098). Retrieved from http://ies.ed.gov/ncee/edlabs.

U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. (2008). *Community colleges: special supplement to the condition of education 2008.* (NCES 2008-033). Retrieved from http://nces.ed.gov/pubs2008/2008033.pdf

Vaquero, L. M., & Cebrian, M. (2013). The rich club phenomenon in the classroom. *Scientific Reports, 3*, 1174.

Visher, M. G., Butcher, K. F., & Cerna, O. S. (2010). *Guiding developmental math students to campus services: An impact evaluation of the Beacon Program at South Texas College.* New York, NY: MDRC. Retrieved May 13, 2013 from http://www.mdrc.org/sites/default/files/full_382.pdf

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*, 82-96.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes among minority students. *Science, 331*, 1447–1451.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*, 1132-1139.

Weiss, M. J., Visher, M. G., Wathington, H., Teres, J., & Schneider, E. (2010*). Learning communities for students in developmental reading: An impact study at Hillsborough Community College.* New York: MDRC. Retrieved May 13, 2013 from http://www.mdrc.org/sites/default/files/full_424.pdf

Wentzel, K. R., & Wigfield, A. (1998). Academic and social motivational influences on students' academic performance. *Educational Psychology Review, 10*, 155-175.

Yeager, D.S., Bryk, A., & Hausman, H. (2013). *The case for practical measurement.* Unpublished manuscript, University of Texas at Austin, Austin, TX.

Yeager, D.S., Paunesku, D., Walton, G., & Dweck, C.S. (2013). *How can we instill productive mindsets at scale? A review of the evidence and an initial R&D agenda*. A White Paper prepared for the White House meeting on "Excellence in Education: The Importance of Academic Mindsets."

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81*, 267-301.

Zeidenberg, M., & Scott, M. (2011). *The content of their coursework: Understanding course-taking patterns at community colleges by clustering student transcripts*. New York, NY: Community College Research Center, Teachers College, Columbia University.