Getting to G.O.L.D.

The VISIBLE LEARNING® Approach to Unleashing Education Improvement

Arran Hamilton
John Hattie

A JOINT PUBLICATION

CORWIN

Cognition Education Group

Getting to G.O.L.D.

The VISIBLE LEARNING® Approach to Unleashing Education Improvement

Arran Hamilton | John Hattie

A JOINT PUBLICATION

CORWIN | Cognition Education Group

CORWIN
A SAGE Publishing Company

FOR INFORMATION:

Corwin
A SAGE Company
2455 Teller Road
Thousand Oaks, California 91320
(800) 233-9936
www.corwin.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Acquisitions Editor: Ariel Curry
Editorial Assistant: Eliza Erickson
Production Editor: Laureen Gleason
Copy Editor: Integra
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Sarah J. Duffy
Graphic Designer: Lysa Becker
Marketing Manager: Kerry Garagliano

ISBN 978-1-0718-1998-2 (web PDF)

Contents

About VISIBLE LEARNING®

In 2008, Professor John Hattie published *Visible Learning*, a synthesis of more than 800 meta-studies covering more than 80 million students. The book revealed what education variables have the biggest impact on learning and created a new mindset that has swept educators around the world. Applying the Visible Learning® methodology means that students are taught to know what they need to learn, how to learn it, and how to evaluate their own progress. Using the Visible Learning approach, teachers become evaluators of their own impact on student learning. The combination causes students to drive their own learning. Since 2008, Professor Hattie has teamed with highly influential educators to expand the Visible Learning canon with books, including *Visible Learning into Action*, *Visible Learning for Teachers, Visible Learning for Mathematics*, and *Visible Learning for Literacy*.

Visible Learning+™ is the model of professional learning that takes the theory of Hattie's research and puts it into a practical inquiry model for teachers and school leaders to ask questions of themselves about the impact they are having on student achievement. Visible Learning+ is a result of the collaboration between Professor John Hattie and Corwin to help educators translate the Visible Learning research. Through a global network of partners, Visible Learning+ professional learning is implemented in over 20 countries in North America, Europe, and the Pacific.

Learn more at www.visiblelearningplus.com

About Corwin

Corwin, a SAGE Publishing company, was established in 1990, first as a professional book publisher, now as a full-service professional learning company, offering professional development solutions on the topics that matter most and the delivery methods that work best to achieve a school or district's objectives. Its many resources range from a library of over 4,000 books to on-site consulting to online courses and events. At the heart of every professional learning experience is the book content and author expertise that have made Corwin the most trusted name in professional development.

Learn more at www.corwin.com

About the Authors

Dr. Arran Hamilton is Group Director of Education at Cognition Education. His early career included teaching and research at Warwick University and a period in adult and community education. Dr. Hamilton transitioned into educational consultancy more than 15 years ago and has held senior positions at Cambridge Assessment, Nord Anglia Education, Education Development Trust (formerly CfBT), and the British Council. Much of this work was international and has focused on supporting Ministries of Education and corporate funders to improve learner outcomes. He currently oversees the design and delivery of Cognition Education's portfolio of education improvement programs, with live projects in the UK, Australia, New Zealand, and the Pacific Islands.

Professor John Hattie is Laureate Professor at the Melbourne Graduate School of Education at the University of Melbourne and Chair of the Australian Institute for Teaching and School Leadership. His areas of interest are measurement models and their applications to education's problems, and models of teaching and learning. He has published and presented over 1,000 papers, supervised 200 thesis students, and published 31 books, including 18 on understanding and applying the Visible Learning® research.

Acknowledgments

We would like to thank Dr. Brian Hincho, Dr. Shaun Hawthorne, Mary Sinclair, and Christophe Mullings from Cognition Education; Ariel Curry and Julie Smith from Corwin; and Distinguished Professor Viviane Robinson and the eight anonymous reviewers who commented on an early draft of this paper. Any remaining errors or omissions are entirely our own.

Introduction

Achieving an Olympic Gold is the pinnacle of success for world-class athletes. But running, say, 100 meters in under 11 seconds is fiendishly difficult. It requires elite performance and getting to gold requires both the *what* and the *how*. The *what* consists of the burgeoning research on the "running economy": the efficiency of different running gaits, training regimes, diets, and equipment/attire to enhance performance on the track.

The *how* consists of athletes' approaches to implementing these learnings, in order to give them critical edge. It involves the support of a coach who can advise on technique, and often the method of implementation is highly scientific. It starts with the analysis of athlete-specific data and leads to the development and testing of hypotheses (or hunches) about how to improve, based on review of global research. This involves large doses of trial and error because what works best for Usain Bolt does not necessarily work best for Tyson Gay or Florence Griffith-Joyner.

What + How = Impact.

In our world of education, there are many parallels. Like Olympic coaches, every educator and education system seeks gold for their learners. Although in this case, gold is at least a year's growth for a year's teaching input.

Educators are also becoming increasingly adept at drawing on the burgeoning research on what works best to enhance their performance. Through the Visible Learning research, we have harvested and synthesized the findings of more than 96,000 research studies involving more than 300 million students: distilling these into 270+ separate influences on student achievement. These data are now publicly available for all on the Visible Learning MetaX, available at http://www.visiblelearningmetax.com/.

Like the Olympic coaches and their athletes, this means educators now have access to a good (enough) compass to point (broadly) in the direction of education gold. We know *what* works best or rather what has worked well previously in a range of different contexts. Educators can use these data to make forward-looking probability estimates, or bets, about what is more (or less) likely to work well in their contexts in the future.

However, one of the key remaining challenges for educators is that tricky question of *how*. An unanticipated outcome of the Visible Learning project is that (in some contexts) it has encouraged educators to simply pick the interventions with the highest effect size—irrespective of whether these address education challenges they genuinely have in their local context. This is the equivalent of a sports coach blindly changing their runner's footwear—based on the global data, rather than by looking closely at the specific areas of improvement that their athlete needs.

A second conundrum that educators then face is how to implement the identified approaches with fidelity. Too often, this has proved fiendishly difficult. Either the original program designers provided no explicit guidelines on *how* or those guidelines just didn't seem to work quite as intended in the new context. In fact, one of the most pressing problems we face in education is the challenge of replication, particularly at scale.

We do not yet have a well-developed science of improvement that enables educators to systematically

- identify education challenges worth solving;
- develop (or select) interventions that are likely to have impact;
- implement with fidelity;
- evaluate to determine whether there has been impact; and
- iterate to enhance impact.

This paper is about that challenge of selection and implementation. Our key purpose is to sketch out some protocols to enhance the probability that the quest for gold will actually lead to gold.

In keeping with the theme of this collection of papers, we call our implementation approach the **G.O.L.D. Method**. As we will unpack, this involves four phases:

- **Goal Hunt**—agreeing on what needs to be improved
- **Opportunity Sift**—agreeing on the mechanisms that provide opportunity for improvement;
- **Lift Off**—implementing the improvement; and
- **Double Back**—measuring the impact and agreeing on where to go next, so that, ideally, we can also **Double Up** to scale the impact.

This has most similarity to the idea that "teachers are to DIIE for"; that is, educator impact comes from excellent Diagnoses, choosing high-probability Interventions, ensuring Implementation fidelity and appropriate dosage, and conducting Evaluations of the impact on the learning lives of students (see Hattie & Zierer, 2018).

On our journey, we will traverse the following waypoints:

Part 1 Overview of the Implementation Challenge
Why Is Education Reform Hard? This section celebrates the wonderful success that governments have made in scaling education for all but laments that progress is slowing and that improvement is getting ever harder.
Toward an Implementation Science for Education, which surveys a range of promising approaches that are currently being piloted in education, including Deliverology and Spiral of Inquiry. We argue that there are golden nuggets in each that could benefit from being extracted and re-blended.
Part 2 The Visible Learning® G.O.L.D. Model
Overview of G.O.L.D. This introduces the key phases and stages.
Goal Hunt. This phase outlines processes for the identification of education challenges worth fixing. In other words, this is about hunting for the right issues to address.
Opportunity Sift. This suggests protocols for sketching, building, and validating theories of improvement. In other words, this is about designing effective interventions.
Lift Off, which is where interventions are implemented.
Double Back. This is about evaluating the effectiveness of what you have done to identify how you can double up the impact.
Finally, we include our observations about moving from the Double Back phase to **Double Up**, that is, to achieve impact at scale.

PART 1

Overview of the Implementation Challenge

1.1 Why Is Education Reform Hard?

Let's start with some good news. The last 170 years has, educationally speaking, been a resounding success. When school systems first emerged, the trailblazers were like bush runners. They had no path, no world-class coaches, and no custom-made "running shoes." Despite this, those early pioneers ran like the wind and quickly massified education.

Since the 1850s, governments around the world have moved from an era of no publicly funded universal education to single-room primary schools, and then to the complex modern schools we see in many parts of the world today. Governments have literally pumped hundreds of trillions of dollars into this endeavor. Schools have been built. Teachers have been identified, trained, and hired. And according to World Bank data, currently 89% of children across the globe have had access to some form of schooling (UNESCO UIS, 2018).

In many countries, education is the single biggest area of public investment, with governments now collectively spending more than USD $3.46 trillion per annum on both basic and higher education (Hattie & Hamilton, 2020). The global spread of schooling has been one of the *greatest* successes of our era.

The global spread of schooling has been one of the *greatest* successes of our era.

Despite these successes, many stakeholders remain unhappy because of the following hurdles:

- **The Access Challenge.** There are still 263 million children (one in five adolescents and disproportionately girls) around the globe that are not *currently* in school (UIS, 2019).[1] This is more than the combined child-age population of the United States, Canada, the United Kingdom, the Netherlands, Scandinavia, Australia, and New Zealand.
- **The Equity Challenge.** In developed countries, where access to schooling is universal, education outcomes are still too much of a lottery—with an average of 14% of students in OECD countries failing to graduate high school (OECD, 2019, Secondary graduation rate indicator). This nongraduation rate includes 21% of boys in the United Kingdom, 18% of boys in the United States, and 17% of boys in Australia.
- **The Future Skills Challenge.** Employers and educators regularly lament that the education system seems to have been built for a bygone factory era and that it does not equip young people with the skills they need for the world of today, let alone tomorrow.

In many countries, policymakers appear very conscious of the fact that their collective investments in education are not bearing enough fruit. There is also growing concern about the potential for mass unemployment as algorithms become primed

[1]Many of these children have previously had some access to primary-level education.

to undertake the more "routine" tasks like stock picking, pathology, and legal review (Ford, 2017). The historian-cum-futurologist Yuval Noah Harari (2019) talks with great seriousness about the likely emergence of a *useless class* whose current employment activities could be entirely automated but who lack the creative and critical thinking skills to reinvent themselves. Indeed, the one thing our machine colleagues can't (yet) do is emulate the uniquely human skills of curiosity, creativity, collaboration, critical reflection, and communication that will likely be the key to success for future generations (Fullan & Scott, 2014; Jefferson & Anderson, 2017).

Recognizing the risks, education policymakers across both the developing and developed worlds keep themselves busy with the task of designing and implementing education reform. That playbook of reform varies from time to time and place to place, but as Pasi Sahlberg laments, it is increasingly looking like a Global Education Reform Movement (GERM) (Sahlberg, 2015), which is a standardized toolkit of policy interventions that centers on the initiatives tabulated in Table 1.

There can be many positive consequences of some of this, but GERM is premised on the belief that it is optimal to primarily focus on the levers external to the school and classroom. The opposite has been equally as fraught—the belief in a school leader or teacher's right to autonomy to choose their own teaching adventure, the focus on how to teach rather than on the impact of this teaching, the denial of expertise among educators, and the false belief that all educators are equal in their impact so leave them alone.

However, as nations increasingly use international rankings like the Programme for International Student Assessment to measure the comparative virility of their respective national education systems, we are witnessing greater convergence. The standardization of the measuring instrument seems also to be encouraging the standardization of policy prescription.

As one of us has argued elsewhere (Hattie, 2015), many of these system-level reforms are largely barking up the wrong tree or, rather, they don't bark high enough. National education systems are a composite of five interlocked layers, as shown in Table 2.

Table 1 The GERM Model of Education Improvement

Policy	Prescription
Standardized Teaching and Learning	• National teaching standards • National curriculum • Common data standards
Market-Based Reforms	• School inspectorate • School league tables • Teacher performance-related pay • Parental choice • School autonomies
Test-Based Accountability	• National standardized testing, linked to market-based reforms
Focus on Literacy and Numeracy	• Increasing teaching hours for math, reading, writing, and science

Source: Sahlberg, 2015.

Table 2 The Five Layers (or Levels) of Education Improvement

Layer	Description
Level 1: Big Picture	• Purpose of education, e.g., basic human right, economic productivity, and equal opportunities • National education challenges, e.g., ensuring all children get a quality education, increasing literacy rates, reducing truancy, zero tolerance to bullying, and unlocking creativity and unique human skills
Level 2: Education System Surface Structure	The policy instruments that are used to effect change at Levels 3 and 4: • Education funding, facilities, curriculum, school governance structures, assessment system, teacher recruitment and training processes, data monitoring systems, class size, length of school day, number of minutes devoted to specific subjects, budget for teacher professional development, and level of centralization versus decentralization
Level 3: Education System Deep Structure	What teachers actually do in classrooms and what they believe (espoused theories vs. theories in practice): • Teachers' theory of learning and their role in it, how they interact with learners, how they provide empathy, how they give and receive feedback, how they gather evidence and reflect on their strengths and areas for development, how they collaborate with each other, and the micro-pedagogies they employ
Level 4: The Learner	• What and how children learn; the facts, values, and skills that are transmitted and the fidelity of transmission (or the neurons that fire and wire together)
Level 5: The Legacy Effect	• The level of long-term contribution that schooling makes in equipping young people with skills for life versus the impact of other formal and informal mechanisms for learning, e.g., environmental context, family, peers, voluntary associations, and TV • The fadeout/degradation of learning over time, i.e., how much of what children learn in school is used and useful for life?

The implicit assumption behind many current and recent national education reform initiatives is that there is a direct causal chain or golden thread that is cast down from Level 1 and that sways vigorously at Level 5. In other words, the assumption is that policymakers set goals; these goals are converted into uniform instruments; these instruments, once implemented, significantly impact, and for the better, what teachers do in classrooms. It is further assumed that this change in teacher behaviors then directly results in children learning faster, harder, and better and that there is a causal relationship between what goes on in this whole schooling ecosystem and life outcomes. And the assumed direction of that relationship is school transmitting to life rather than vice versa.[2]

Our sense is that what happens at Level 3 has a very strong impact on Level 4 and that this, in turn with various out-of-school influences, has *some* impact in laying appropriate foundations for Level 5. However, we are not at all convinced that much of what happens at Levels 1 and 2 has any significant impact on Levels 3–5. Our contention is that both segments operate, too often, in parallel universes.

[2]Not all educational theorists hold with this view, however, from Dewey to Illich to Beane to Yong Zhao, to name a few.

There are many explanations for why the growing global body of research about what works translates into limited impact in schools and classrooms. First, education is often a highly politicized area of public service delivery (Nordstrum et al., 2017). This means that new programs can sometimes be implemented based on political or ideological considerations and without any clear linkage to the evidence on what works or any evaluation of actual impact. In the United States, for example, federal funding for education programs has rarely been tied to impact data (Slavin, 2002).

Second, even where appropriate and research-based interventions have been identified, achieving implementation with fidelity is extremely difficult (Berman & McLaughlin, 1975; Cuban, 1993; Darling-Hammond & Snyder, 1992; Elmore & McLaughlin, 1998; Rivlin & Timpane, 1975; Robinson, 2018). Interventions and approaches that researchers agree on paper *should* be effective often do not produce the intended results or significantly change what educators do in the classroom (Cohen & Moffitt, 2009; Correnti & Rowan, 2007; Rowan, Correnti, Miller & Camburn, 2009).

Ultimately, the success of any educational reform is directly proportional to the number of teachers who are willing to question their own approaches and to look seriously at ways of doing things differently, multiplied by the number of change agents that can help them on their journey. No wonder that Larry Cuban (1998) in his analysis of the impact of John Dewey's progressive education movement in the United States concluded that even at its peak no more than 25% of teachers had fully converted to using student-centered approaches. The majority did not change anything at all and many of those teachers who said they were progressivists employed hybrid approaches that retained many of the features of their preexisting teaching practice.

The success of any educational reform is directly proportional to the number of teachers who are willing to question their own approaches and to look seriously at ways of doing things differently, multiplied by the number of change agents that can help them on their journey.

To overcome these challenges, we need to widen our focus from *what works* to *how we can make it work*. We need to ask what works best for *whom* and under what conditions; or why it worked earlier and how we can make it work again. In short, we need a science of improvement (including effective *implementation* of those improvements) for education.

1.2 Toward an Implementation Science for Education

In the parallel field of health, implementation science has had a reasonably long history. Methods for effective rollout and scale-up of health reforms have been iterated and improved since the 1960s (Pressman & Wildawsky, 1984; Wandersman et al., 2008). These approaches emerged because empirically tested health intervention programs were not having the desired impact, once implemented in primary care settings.

The research in health care suggested

1. **The actual *process* of implementation strongly influenced whether there was impact** (Kelly, 2012; Meyers et al., 2012).
2. **There is a powerful relationship between the *beliefs and values* of health care practitioners and the outcome of implementation**; that is, when the new approach aligned with health care professionals' worldviews, it was more likely to be implemented with enthusiasm (Aarons, Green, & Miller, 2012). Ergo, implementation must tackle beliefs.
3. **Practitioners should not have sole responsibility for implementing research-based interventions**. Instead, accountability for success also lies with the researchers and program developers, who should work in tandem (Meyers et al., 2012).

In our business of education, implementation science has had a relatively short history. The *Handbook of Implementation Science for Psychology in Education*, edited by Kelly and Perkins, was published in 2012. This was the first weighty tome on the psychology of education implementation.

There are also several promising improvement methodologies for education that are being piloted in a range of education contexts (Table 3).

Table 3 Overview of Some Promising Education Improvement Methodologies

Methodology	Description	Reference
Deliverology	Developed by Sir Michael Barber, this approach focuses on scaled implementation of an agreed approach, with fidelity. The framework is intended to be generic (i.e., can be used within and beyond education) but has been employed in educational contexts in the United States. **Verdict:** Strong on governance, implementation, and scalability processes but limited focus on protocols for selecting appropriate interventions or for engaging with stakeholders' beliefs and values. More suitable for top-down change.	Barber, M., Kihn, P., & Moffit, A. (2011). *Deliverology 101: A field guide for educational leaders*. Thousand Oaks, CA: Corwin.

(Continued)

(Continued)

Methodology	Description	Reference
Positive Deviance	Centers on grassroots exploration and fact finding to identify positive outliers or "deviants" to whatever issue stakeholders seek to resolve. The idea is to catalogue positive deviant behaviors that can be replicated and scaled up. **Verdict:** Strong on innovating problem solving by scaling up pockets of effective practice, but limited focus on how stakeholders select their issue, how they should scale up, or how to address the fact that many instances of positive deviance are nonreplicable.	LeMahieu, P. G., Nordstrum, L., & Gale, G. (2017). Positive deviance: Learning from positive anomalies. *Quality Assurance in Education*, *25*(1). https://doi.org/10.1108/QAE-12-2016-0083 Pascale, R., Sternin, J., & Sternin, M. (2010). *The power of positive deviance: How unlikely innovators solve the world's toughest problems*. Boston, MA: Harvard Business Press.
Spiral of Inquiry	Developed by Helen Timperley, Linda Kaser, and Judy Halbert, the Spiral of Inquiry provides frontline educators with six subroutines to enhance student learning outcomes: scanning, focusing, developing a hunch, learning, taking action, and checking. **Verdict:** Strong on identification of areas for improvement and the development, implementation of local action research projects to implement and measure improvement and on privileging stakeholder beliefs. Less focus on specific implementation processes or on scaling up.	Timperley, H., Kaser L., & Halbert, J. (2014). *A framework for transforming learning in schools: Innovation and the spiral of inquiry*. Victoria, Australia: Centre for Strategic Education.
Agile for Education	Range of approaches, including Simon Breakspear's *Teaching Sprints*, which have been adapted from the generic Agile model and applied to educational settings. **Verdict:** These frameworks have similar focus and benefits to the Spiral of Inquiry model and are valuable process drivers for school or departmental professional learning communities to undertake fast improvement cycles. They may have less value in driving system-wide reform.	Breakspear, S. (2017). Embracing agile leadership for learning: How leaders can create impact despite growing complexity. *Australian Educational Leader*, *39*(3), 68–71.
Reduce Change to Increase Improvement	This is a body of school improvement research and processes developed by Viviane Robinson. It proposes a highly selective improvement focus and offers protocols for coaches to engage with and understand stakeholder beliefs around change. **Verdict:** Wonderfully strong on "less being more" and on mechanism for engaging with rather than bypassing educators' theories of action and their beliefs. Less coverage on the processes of implementing, evaluating, and scaling up improvement.	Robinson, V. M. J. (2018). *Reduce change to increase improvement*. Thousand Oaks, CA: Corwin.

Methodology	Description	Reference
Hexagon Tool	A six-part planning tool for schools to identify local needs and then evaluate prebuilt programs and their suitability for the local context. Focus areas are need, fit, resources, evidence, readiness, and capacity. **Verdict:** A useful thinking tool to support individual schools to select evidence-based programs, but it does not specifically address or support implementation or evaluation, postselection.	Blase, K., Kiser, L., & Van Dyke, M. (2013). *The hexagon tool: Exploring context*. Chapel Hill, NC: National Implementation Research Network, FPG Child Development Institute, University of North Carolina at Chapel Hill.
Learning to Improve	Developed by researchers at the Carnegie Foundation for the Advancement of Teaching, this methodology provides a framework for the identification of areas for improvement, improvement hypothesis development, and Agile-like improvement cycles. **Verdict:** A powerful set of tools with a strong focus on starting local and then scaling. Also advocates involvement of external researchers/coaches to work alongside those implementing improvement.	Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Other promising approaches include the following:

- *Collective Impact* (Kania & Kramer, 2011), which focuses on establishing a centralized infrastructure, dedicated staff, and structured processes.
- *Scaling Up Education Reform* (Bishop, O'Sullivan, & Berryman, 2010), which centers on establishing goals, developing new institutions, spreading reform, gathering evidence of impact, and creating opportunities for all stakeholders to take ownership of the reform.
- Many publications by Michael Fullan and Andy Hargreaves on scaling up (including Hargreaves & Fullan, 2012).

While there are encouraging case studies supporting the effectiveness of these approaches, there are (as yet) no meta-analyses that review or compare the impact of these different methodologies at scale. There is a significant gap in both the literature and the Visible Learning MetaX database.

At present, the best that we can do is harvest the golden nuggets from these approaches to improvement and synthesize them into a practical set of (hunch-based) principles and tools that educators can use to support more effective education implementation runs.

PART 2

The VISIBLE LEARNING® G.O.L.D. Model

Overview

Each of the approaches to effective improvement outlined in the previous section has its own in-built "twist" or areas of emphasis. There are, however, some commonalities between all these approaches. To some degree, they each emphasize the importance of *carefully* carrying out the following steps:

1. selecting the area for improvement;
2. developing (or identifying and localizing) one or more interventions to generate impact;
3. implementing the agreed interventions, using robust project management tools and approaches; and
4. leveraging the scientific method to gather data during implementation, to verify and enhance impact.

What we set out to do in the remainder of the paper is to fashion these nuggets into an explicit improvement methodology that we call the Visible Learning® G.O.L.D. Model (see Table 4).

We make no claim that G.O.L.D. is revolutionary. It merely takes the best tools and processes that we have (shamelessly) scavenged from the implementation methodologies outlined in the previous section and blends these with trial-and-error insights that our teams have gleaned from direct involvement in large-scale improvement programs in the United Kingdom, United Arab Emirates, Malaysia, Australia, New Zealand, Europe, and the Pacific Islands.

Our hunch is that these processes and tools will be most valuable to educators that wish to embark on large-scale improvement or what can be called

Table 4 The Visible Learning® G.O.L.D. Model

Phase 1 Goal Hunt	Phase 2 Opportunity Sift	Phase 3 Lift Off	Phase 4 Double Back
Finding/diagnosing an education challenge worthy of everyone's time and effort	Systematically investigating mechanisms for improvement to agree on the best-fit approach	Implementing the agreed improvement strategies	Explicitly and scientifically measuring the impact and agreeing where to go next

Education Challenges. As Table 5 outlines, these are highly complex problems with severe impact on future quality of life (e.g., employability, health, dignity, adaptability, resilience, happiness, and life expectancy) and where stakeholders agree that they want to make resolution their number one crusade. Approaches to resolution often involve the establishment of a taskforce or guiding coalition that works with and across multiple schools—sometimes even a whole system. These tools will also be of use to individual schools, where they are united in identifying and addressing an agreed whole-school education improvement initiative.

We now unpack the four phases of G.O.L.D. and present some practical tools that can be used to support successful resolution of substantial education challenges.

Table 5 Defining Education Challenges

	Education Enhancement: addressed by educators as part of what they do in their classrooms	**Education Improvement:** more complex, addressed by local groups of educators via, e.g., professional learning community	**Education Challenges:** highly complex and high risk, requires highly structured intervention via, e.g., G.O.L.D. thinking approach and additional resources
Perceived Impact	Some potential risks to quality of life if not tackled	Moderate risk to quality of life if not tackled	Significant risk to quality of life if not tackled
Confidence in the Data	Data often white noise or statistical anomalies	Often long-term underlying pattern in the data	Usually incontrovertible evidence
Stakeholder Consensus	No agreement that a problem exists	Agreement that the problem exists but there are differences of opinion about how to solve it	Strong agreement that the problem exists, but there are differences of opinion about how to solve it AND often limited time
Resourcing and Reversal	Does not require external intervention Problem usually self-correcting	Benefits from external intervention, i.e., help from outside the system Problem rarely self-correcting	Obvious to everyone that it requires external intervention Problem almost never self-correcting
Complexity	Relatively easy to solve	Difficult to solve	Fiendishly difficult to solve—solutions often generate unanticipated consequences
Scale	Classroom	School	Multiple schools

Phase 1 Goal Hunt

Education challenges are often complex to define and fiendishly difficult to resolve. Change initiatives are more likely to be successful when

- the benefits of change have been clearly defined,
- there is strong consensus that the education challenge needs resolving,
- there is deep understanding of the causes, and
- stakeholders explicitly map out what success looks like.

Therefore, we need to be clear before we embark on change that the goals are worthwhile to warrant the time and investment in attempting systematic change. **This means we need to undertake a Goal Hunt.**

A **Goal Hunt** is a comprehensive inquiry:

1.1 The inquiry starts with the clear purpose of identifying education challenges that are worth everyone's time and effort in solving; that is, **what is it are we trying to make better?**

1.2 During a Goal Hunt, education challenges are systematically detected, catalogued, reviewed against local and global evidence, and ranked; that is, **are we sure we have identified the most appropriate area to make better? And are we sure that we haven't bitten off more than we can chew?**

1.3 Ideally, the fullest range of stakeholders participate in the Goal Hunt process and they are given the opportunity to state their preferences and share their beliefs; that is, **do we ALL agree that solving this education challenge is the best use of our time and resources? Are we all committed to making this better?**

1.4 Stakeholders work collaboratively to build and validate causal models to diagnose and explain WHY their selected education challenge exists; that is, **do we share the same beliefs about the cause of this education challenge? What are the different potential causal explanations we need to keep in mind for when we move to the Opportunity Sift phase?**

1.5 Stakeholders collaboratively set the (provisional) improvement goal, that is, **this is where we are now, and this is where we are going to get to, and this is the measure we are going to use.**

And if you can't find an education challenge that most stakeholders agree is worth resolving, put down your hunting implements and get back to the day job.

In the following subsections, we outline how you might bring a Goal Hunt to life in your context.

1.1–1.3 Hunting for an Education Challenge Worth Solving

To bring about change you must have a clear sense of what you want to change and why. At the system, district, or school level, the first step is to establish a G.O.L.D. Team that meets, explores, analyzes, and then sifts and sorts all the potential education challenges. This is akin to the way that Coach Glen Mills identified stride frequency and balance as the key challenges or improvement areas for Usain Bolt.

The initial role of this hunting party is to find as many challenges as there are worthy of resolution, as possible. The question that we ask ourselves when looking at challenges in the educational landscape is "What's the worst that could happen if we did nothing?" Table 6 illustrates some of the education challenges uncovered by goal hunters in different contexts.

The next step is to identify which (if any) education challenges are worthy of systematic resolution/remediation. One way of addressing this is to employ the collective wisdom of the crowd and ask all relevant stakeholders (including educators, students, parents, and community members) to literally vote on which education challenge they think is the most important to resolve.

There are two very good reasons for doing this. The first is that, as Sir Francis Galton discovered when he asked the crowd at a country fair to guess the weight of an ox (while individual guesses or

Table 6 Identifying Education Challenges Worth Solving

GOAL HUNT	
Education Challenge	**Potential Consequences** ***What's the worst that could happen if we did nothing?***
Teacher recruitment and retention • Difficulty in recruiting qualified math and science teachers • 50% of teachers leaving profession after 5 years	• Reduced quality of math and science lessons • Fewer students opting for science/technical careers • Higher lifetime cost of teacher training
17% of graduating students not achieving the minimum standard of literacy	• Students are unable to progress to higher education • When students enter the workforce they are limited to routine roles that are susceptible to automation
Information and communication technology (ICT) equipment in the school district is reaching end of life and needs to be replaced	• There will be insufficient functioning ICT equipment across our schools • Our learners may not develop ICT skills
9% of students are not regularly attending school	• Student learning is hampered • Students do not graduate from school • Restricted employment opportunities in adulthood
Teacher professional development is at an all-time low in offerings and quality	• Quality of teaching is significantly affected • Student learning outcomes are significantly hindered

opinions were generally wide of the mark), the mean average of all guesses or opinions was extremely close to the ox's true weight (Galton, 1907). What Galton uncovered was the *collective wisdom of the crowd*, that is, when we combine the mean average of everyone's intuition, often it is not far from the empirical truth. The second is that, in our educational context, by engaging with all stakeholder groups, people will be more likely to buy in to the idea that the highest-ranked education challenges are crusades worthy of significant investment of their own time and energy. We are also more likely to be able to determine stakeholders' belief and value processes, which will be useful later, as we design and then implement our intervention.

One way of undertaking the voting exercise, if stakeholders have identified a very long list of potential challenges, is by comparison of matched pairs; for example, the following box shows the comparison of a pair of education challenges.

9% of students are not regularly attending school	versus	ICT equipment in the school district is reaching end of life and needs to be replaced

The idea is that every education challenge is compared to every other education challenge, and the total number of votes for each comparison is then tallied and ranked. Stakeholders are asked to choose which item from each pair that they rate has the greatest priority for resolution. This process is repeated until every option has been pitted against every other option. The ratings are then aggregated into an overall ranking of perceived importance.

An important aspect of this process is that it involves strong involvement from local stakeholders, thereby ensuring that whatever priority interventions emerge clearly come from the community that will go on to lead resolution.

However, vote counting alone might lead to the identification of inappropriate education challenges. As you pit your list of identified education challenges against one another, it is also important that you look at the quality of evidence you have at your disposal to make your collective rankings. Do you have any of the following conditions?

- **High Confidence**, for example, a significant amount of quantitative and qualitative data that have been collected from various sources, which are mutually corroborative.
- **Medium Confidence**, for example, some robust data exist, but there are gaps that require leaps of faith, or there are strong data, but they pull in different directions.
- **Low Confidence**, for example, while there is strong intuition and consensus that the problem might exist, there is no hard supporting evidence. Are judgments entirely based on gut reaction?

You will also want to consider whether the education challenges you have uncovered are related to **outcomes** or **inputs**. An outcomes challenge is one where the existing education system is resulting in reduced life chances for learners (e.g., lower employability, health, dignity, adaptability, resilience, happiness, and life expectancy). Whereas an inputs challenge is when there is some change in the education operation system that indirectly threatens the possibility of an outcomes crisis down the line (e.g., challenges in recruitment and retention of teachers or school leaders).

Going from Deductive to Inductive

The method for education challenge identification that we have outlined in the text is deductive. It starts with stakeholders laying down ideas and then looking at the evidence to see whether or not this conforms to their perception of reality. One potential downside to the deductive method is that it might encourage stakeholders to identify superficial or unimportant education challenges.

Another approach is to undertake your education challenge hunt inductively. This means that you don't prejudice your search for a worthwhile education challenge with preconceptions about what needs to be fixed and then take a public vote. Instead you go on a data hunt. You start by looking at metric data, undertake learning walks, and collect teacher and student voice data to identify/diagnose a range of potential education challenges. You can then consider which of the challenges have the most severity of impact.

Depending on the resources at your disposal, you could divide your G.O.L.D. Team into two separate hunting parties (i.e., deductive and inductive); compare and synthesize the results from both at the end.

If you can't find an education challenge big enough to be worth solving, we would suggest that you disband your G.O.L.D. Team.

But if you have identified an education challenge for which there is near-unanimous agreement, we suggest that you then attempt to define the challenge as explicitly as possible. For example, if your selected challenge is that 9% of students are not attending school regularly, you will want to identify whether there are any patterns in the demographic of that 9% and/or the duration of absenteeism, as shown in Table 7.

1.4 Map the Causal Drivers

In order to resolve an education challenge, we need to develop and test hypotheses (or hunches) about the potential causes. This requires the G.O.L.D. Team to consider all the possible causal factors that could be contributing to the education challenge.

One way of identifying the causal factors to an identified challenge is called *The Five Whys* approach, which was originally developed by Sakichi Toyoda and was used within the Toyota Motor Corporation (Ohno, 1988). It involves asking and collecting data about each hypothesized link in the causal chain. The box on the following page illustrates an example using the Five Whys approach.

The idea is that you ask as many "why" questions as you can think of and then use these to

Table 7 Education Challenge Breakdown Structure

Education Challenge	Breakdown Structure	What Does This Mean?
9% of students are not regularly attending school	• Not regularly attending is defined as missing two days in any single week; and/or more than four days in any single term • Of nonattending cohort—36% only just meet the threshold definition; 64% are persistently not attending • 82% of nonattenders are boys between ages 13 and 17 • 63% of nonattenders are low SES group • Student voice collection was inconclusive—students unwilling/unable to articulate the reason for their absence	• Our challenge group is 64% of 9% or 5.76% of the student body. • Most of our nonattenders are adolescent boys from lower SES (socioeconomic status) group. • We don't know the causes/whether there are patterns in those causal mechanisms. • But given that majority of nonattenders share similar demographic features, it is reasonable to speculate that there may be an overlap in the underlying causal mechanism.

The Five Whys Approach

Education Challenge: 9% of children are not coming to school and it's interfering with their learning

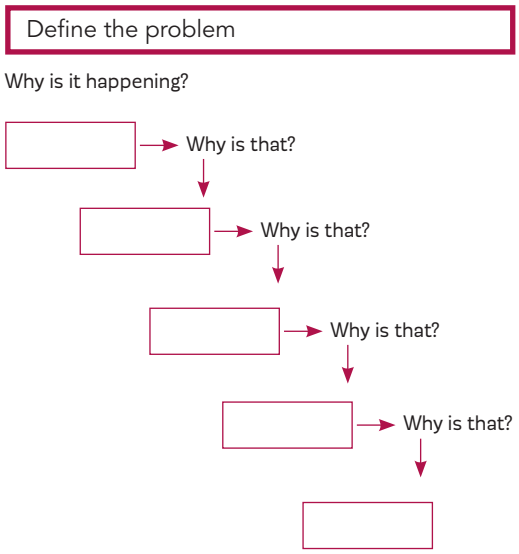
1. Why? They don't want to come.

2. Why? They can't relate the learning to their lives.

3. Why? It's boring and not contextualized.

4. Why? Our teachers are not differentiating to their needs.

5. Why? Our teachers have not been trained and supported to do this.



build a causal model. It is extremely likely that, depending on their different worldviews/beliefs, stakeholders will come up with multiple five-ways. This process provides a window into stakeholder beliefs, and it is important to interrogate each model and come to a shared understanding.

In Figure 1, we illustrate how the collective or majority view can be presented in the form of a path analysis. However, you will notice that in the illustration we have only peeled back one layer of causation on each of these potential "whys" (system-level discovery teams may find it beneficial to add at least one more layer of "whys" that directly map to the big six factors listed in Figure 1).

You could also present your analysis as a fishbone diagram or an issue tree. However, what you have mapped out is just a theory. You now need to collect data to see whether your hunches about the causes of the education challenge are plausible. Table 8 shows how each potential causal variable can be investigated and recorded. However, the verification methods you undertake might vary considerably depending on the nature of your education challenge. Sources of data might include global research including Visible Learning MetaX, student voice, learning walks, parent interviews, teacher observation, and locally administered surveys.

At the end of the review process, the G.O.L.D. Team redraws their path analysis to reflect what they have uncovered from their research. This finalized causal driver map will be central to the crafting of interventions or solutions during the Opportunity Sift stage.

1.5 Set the Quantifiable Improvement Goal

The final stage of the Goal Hunt involves setting, agreeing, and locking specific and measurable success criteria for the selected education challenge (Table 9). As Peter Drucker once said, "What gets measured gets improved."

Figure 1 Example Path Analysis

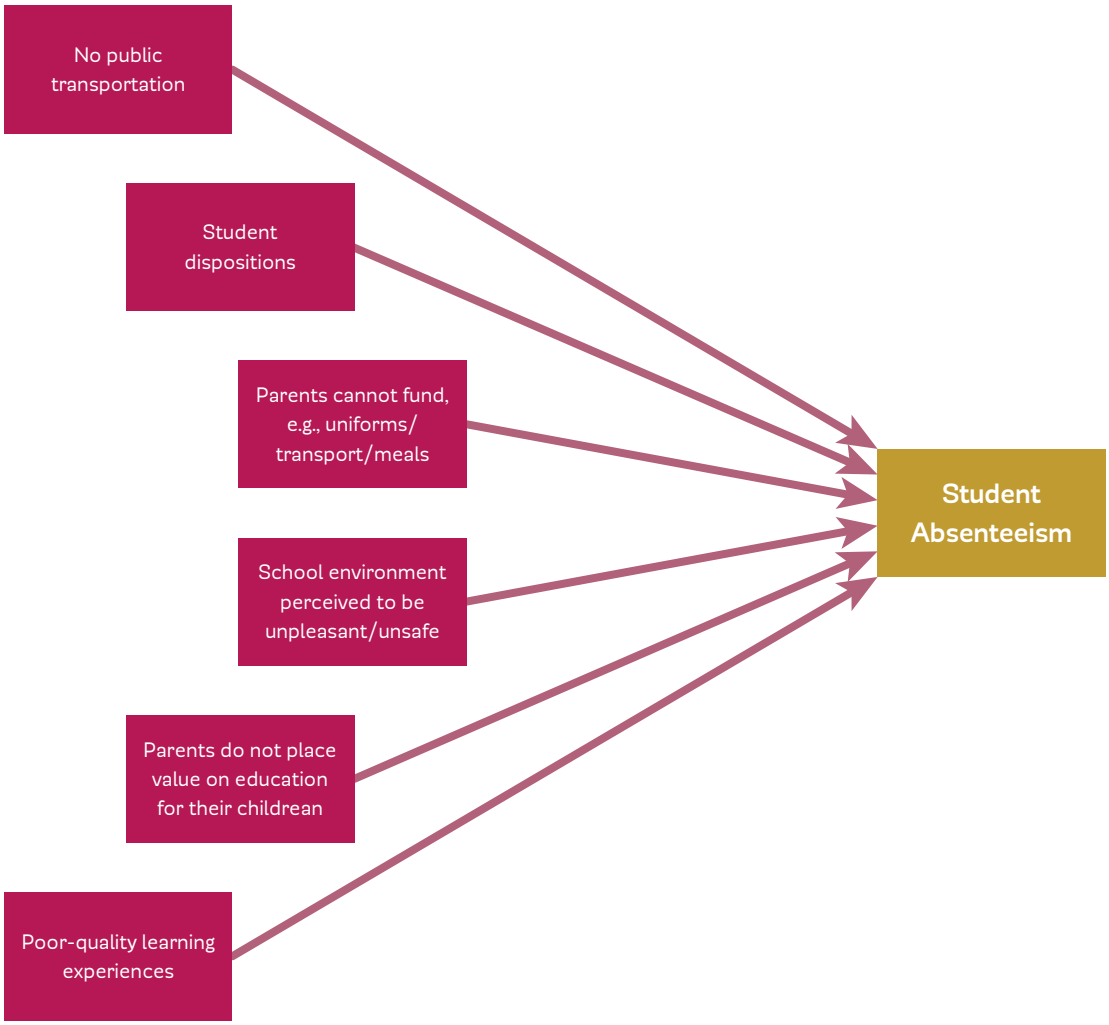


Table 8 Causal Driver Verification

Causal Variable	Causal Hypothesis	Sources of Verification Data	Outcome of Verification	Variable Remains?
No public transportation	Students are not attending school because the transportation infrastructure is inadequate	Student geolocations Bus routes	**Not verified.** All students live within 30 minutes' walk from schools, within the district.	**Removed**
Poor-quality learning experiences	Students are not attending school because they do not enjoy their classes and do not believe what they study is relevant	Student voice Lesson observations Curriculum review	**Verified.** Students consistently reported that they found lessons unengaging. Lesson observations also suggested disengagement.	**Remains**

Table 9 Setting the Success Criteria

Current Situation	"To Be" Situation	By When	Measured How
9% of students are regularly not attending school	97% of students are regularly attending school	December 2022	• School attendance register data • Regular attendance defined as 7 or fewer absences during school year
Justification of Selection of "To Be" Values			
Review of regional comparator data suggests that on average 3% of students regularly do not attend school. Our goal is to achieve the same rate of attendance as regional comparators.			

Phase 2 Opportunity Sift

Attempts to resolve large-scale education challenges are more likely to be successful when:

- interventions are explicitly designed;
- the designs draw on global and local research, rather than intuition alone;
- a range of alternative interventions are proposed, reviewed, ranked, and refined by local stakeholders;
- the review process includes analysis of "human factors" including cognitive biases and the degree to which the proposed intervention aligns with or contradicts stakeholders' existing worldviews;
- the selected intervention has an evidence base that suggests a high probability of success and that the approach is adaptable to the local context; and
- stakeholders build an explicit and phased success map, with measurable improvement targets.

Therefore, before we blindly rush ahead and implement any approach or intervention, we need to review and sift all the potential opportunities or options to select and design the approach that has the highest probability of being effective within the local ecosystem.

This means we need to undertake an Opportunity Sift.

An **Opportunity Sift** is a systematic solutions-design exercise:

2.1 It starts by going back to the causal drivers that were hypothesized and validated during step 1.4 in our Goal Hunt. The idea is to sketch out all the different opportunities and options for blocking, weakening, or reversing the causal drivers of an education challenge; that is, ***what are ALL the different ways and tools we can use to fix this? And what evidence is there that these ways will work for us?***

2.2 Next comes the development of a theory of improvement. This takes the best opportunity sketches from step 2.1 and maps them out into a full delivery model that details the resources, activities, and outputs and how each of these contributes to solving the education challenge; that is, ***how are we going to make this happen? What resources do we need? Who is going to do what?***

2.3 We need to recognize that it's unlikely we will have developed the perfect theory of improvement the first time. Therefore, we need to explore all the ways it can be iterated and consider the human factors (aka local beliefs and values) and whether our improvement model engages with these; that is, ***what are all the different ways we could wiggle (iterate) our design and which wiggles do we think will result in better impact? And does what we propose to do align with the way our stakeholders think and feel?***

2.4 Finally, we need to develop a success map. This builds on the baseline and quantifiable improvement goal targets you already set in step 1.5. It builds these out into a full results framework that sets short-, medium-, and longer-term targets for each activity, output, and outcome; that is, ***how and when are we going to measure whether we are on track?***

At the end of the Opportunity Sift process, you will have identified the best options for improvement and built these into a cohesive improvement initiative, ready to **Lift Off.**

In section 2.1, we outline how you might undertake an Opportunity Sift process in your context.

2.1 Reverse Engineer from Causal Drivers to Opportunity Sketches

The opportunity sketches activity builds on the work you undertook in step 1.4, where you mapped and validated the causal drivers that generate your agreed-on education challenge. The idea is to sketch out as many potential options as possible to resolving/remediating/reversing each of the agreed causal drivers of your education challenge. And, again, you might wish to include a range of stakeholders in the process of making your long list.

In Table 10, we illustrate how you could create an opportunity sketch. From your work in steps 1.1–1.4 you can already complete columns 1 and 2 of the table. Completion of column 3 can be undertaken as a group activity with interested local stakeholders, as a primary research activity via interviews and discussions with a range of stakeholders, or as a piece of secondary desk research with review and validation by local stakeholders.

However, you are unlikely to have the time and resources to implement all the initiatives that you identify in your opportunity sketches. So, you need a mechanism to whittle down to the opportunities that have the highest probability of being effective in your local context.

One place where you can find global data on the types of interventions that have been successful in a range of contexts is the Visible Learning MetaX. Other very good sources of research data are shown in Table 11.

Table 10 Creating an Opportunity Sketch

3. Opportunity Sketches ← (i.e., actions you can take to reduce/block/ reverse the influence of column 2 variables on the column 1 education challenge)	2. Causal Drivers/ Challenge Components (1.4) ←	1. Education Challenge Statement (1.1–1.3) ←
• Vocational education option for adolescent boys, e.g., group project building a car	School is not enjoyable for absentee learners	Student Absenteeism
• Project-based learning curriculum with range of student-selected options that students can relate to		
• Introduction of an equity/cultural competency program for teachers, e.g., Culture Counts Plus		
• Introduction of equity program, e.g., Culture Counts Plus parent engagement strand	Parents do not see value of school for their children	
• Home–school partnership program		
• AI-driven early warning system to pre-identify at-risk learners	Natural dispositions in some students make them more inclined to disengage	
• Tangible cash/goods awards to encourage/ reward attendance		
• Training for targeted students in grit/ self-regulation		
• School visits from successful alumni role models		

Using the data available in the sources given in Table 11 and what you know about your local context, you can then rank each of your proposed opportunity sketches, as shown in Table 12.

And in Table 13, we outline some criteria that you can use to undertake this ranking exercise.

This is akin to the kind of opportunity selection process that Coach Mills might deploy as he explores the potential impact of different kinds of footwear: interval training regimes, dietary changes, and breathing techniques to improve Usain Bolt's chances of getting to gold.

By the end of this process you will have ranked all your opportunity sketches, and it is extremely likely that a small number of them are head and shoulders above the rest. These are the opportunities that you will carry forward to the next activity.

Table 11 Global Sources of Education Effectiveness Research Data

Source	Description of Contents
Visible Learning MetaX™ (Global)	Open-access education database of 1,700+ meta-analyses of more than 96,000 individual research studies, involving more than 300 million study participants
Education Endowment Foundation (UK)	Summary data on 34 types of education intervention and summary literature review reports
Iterative Best Evidence Synthesis Programme (NZ)	Literature reviews on 8+ common education challenges
What Works Clearinghouse (US)	Summary of programs/interventions with high-quality randomized control trial data (RCT) available
Best Evidence Encyclopedia (US)	Narrative literature review of 50+ common education challenges
Campbell Collaboration (Norway/Global)	Systematic reviews in a range of social policy areas including education, international development, crime and justice, and disability
Cochrane Collaboration (UK/Global)	Plain language summaries of health-related challenges and interventions
Health Evidence (Canada)	Free text-searchable database on health-related challenges and interventions, e.g., obesity, substance abuse

Table 12 Opportunity Sketch Ranking

Opportunity Sketches	Evidence of Impact 1–5 (5 = strong evidence)	Ease of Replicability 1–5 (5 = high ease)	Local Capacity to Implement 1–5 (5 = high capacity)	Cost of Implementation 1–5 (5 = low cost)	Total
School visits from successful alumni role models	3	2	1	4	10/20
AI-driven early warning system to pre-identify at-risk learners	4	3	3	3	13/20

Table 13 Potential Criteria for the Ranking of Opportunity Sketches

Factor	Criteria
Evidence of Impact	• Outcomes achieved, e.g., effect size • Number of studies and population of studies (e.g., in the Visible Learning MetaX, we include a confidence ranking for each influence) • Similarities between the context of the studies and your local environment
Ease of Replicability	• Is the intervention "productized" or do you need to build it yourself? • Are the steps easy to follow or open to wildly different interpretations?
Local Capacity to Implement	• Do you have access to high-quality internal or third-party technical assistance to support implementation? • Is there buy-in from stakeholders? Does the intervention model conform to local stakeholder beliefs/theory of action? • Do stakeholders have sufficient time to engage/participate at the levels required for success?
Cost of Implementation	• Total cost ÷ Total number of *Direct Beneficiaries*

2.2 Develop a Theory of Improvement

Now that you have identified your opportunities with higher probability of impact, you need to plan how you will deliver them to achieve impact. One framework for doing this is called a theory of improvement (see Table 14 for a worked example), and it involves providing convincing and cohesive answers to the following questions:

1. What resources do we need to deploy to bring about improvement (e.g., people, time, budget)?
2. What activities are we going to undertake with that resource to generate improvement in the education challenge area? Which stakeholders will we be engaging with during those activities? This will be drawn from the opportunity sketches from section 2.1 that you have most confidence in.
3. What are our assumptions about the way that these activities will lead to improvement?
4. What will the outputs of the activity be (i.e., the products created, the audience engaged with, and so on)?
5. What measurable short- and longer-term outcomes do we expect to see as a result of implementing the activity and by when?

2.3 Iterating Your Theory of Improvement

In larger-scale initiatives, rather than quickly developing a single theory of improvement and then launching the program, there are strong benefits to developing multiple theories of improvement. This approach allows you to consider all the different ways that you could construct and implement your improvement agenda.

Dr. Arran Hamilton and the team at Cognition Education have developed the Wiggle Tool to help stakeholders identify the various ways implementation can be varied or "wiggled." Table 15 provides a worked example. The columns detail a range of activity features that can be wiggled, such as who participates and who delivers training. And the rows are used to outline all of the potential wiggles and then to agree and select the ones with the highest probability of impact.

Table 14 Theory of Improvement Worked Example

Area for Change		Actions				Outcomes		
Education Challenge	Challenge Component	Resources	Activities	Assumptions	Outputs	Short-Term Outcomes	Medium-Term Outcomes	Long-Term Outcomes
Student absenteeism: 9% of students not regularly attending school Not regularly attending is defined as missing two days in any single week and/or more than four days in any single term • Of nonattending cohort 36% only just meet the threshold definition; 64% are persistently not attending • 82% of nonattenders are boys between ages 13 and 17	**A. School is not enjoyable for absentee learners**	• $XXX budget for district professional development (PD) funding • Release time for 50 identified teachers	Introduction of an equity program for teachers—e.g., Culture Counts Plus (CC+)	• Teachers will understand, agree with, and implement what they have been trained in • The approaches are effective in changing teacher behaviors toward an agentic thinking model • Those changes enhance student engagement	• 50 identified teachers successfully complete the Culture Counts Plus PD program	• 90% of teachers report that they are satisfied with program	• Lesson observation data show that 75% of teachers are implementing the CC+ approach consistently in their classrooms • Pre-/post-student voice interviews show 21% improvement in student perceptions	97% of students regularly attending school
		• Partnership with PD provider to activate the model, with estimated cost of $XXX • Release time of 20 hours for five teachers to be trained	Special whole-school project aimed at engaging adolescent boys—F1 competition	• The targeted students will find F1 engaging • Implementation of this 1-time/1-hit project will be sufficient to generate engagement • Other students will not become disengaged, i.e., unintended side effects	• 1x school-wide project implemented			

Area for Change		Actions				Outcomes		
Education Challenge	Challenge Component	Resources	Activities	Assumptions	Outputs	Short-Term Outcomes	Medium-Term Outcomes	Long-Term Outcomes
• 63% of nonattenders are low SES • Student voice collection was inconclusive—students unwilling/unable to articulate the reason for their absence		• 1x PLC network lead spending 15 hours per month planning meetings and supporting peer collaboration • 1 hour per month for teachers to attend meetings and 5 hours for each to trial new approaches outside meetings	Establish 12-month professional learning community model for teachers	• Teachers have the time and enthusiasm to engage • Teachers will undertake inquiry in appropriate areas and embed new locally appropriate approaches • New approaches will make learning more enjoyable for absentee learners			• Student voice surveys show XX% increase in student satisfaction	97% of students regularly attending school
	A. **Parents cannot fund ancillary costs associated with their children's participation in schooling**							

Source: Copyright © Cognition Education (2020). All rights reserved.

Table 15 Wiggle Tool Worked Example

Activity: Cultural Competency Training Program for Teachers					
	Feature 1 ***Participant Selection***	**Feature 2** ***Who Delivers Training***	**Feature 3** ***Incentives for Participation***	**Feature 4** ***Post-Training Support Mechanism***	**Feature 5** ***Duration of Training***
Variant 1	Mandatory for all	Delivered by in-house team	None	None	1 day
Variant 2	Participant self-selection	Delivered by external facilitators	Reduced timetable to participate in action research	Encouragement of optional participant action research	3-day block
Variant 3	Mandatory for screened/ identified stakeholders	External facilitators with gradual release model	CPD certificate	Agreed package of coaching support for all participants	3 days, spaced
Variant 4	Optional but highly recommended for prescreened stakeholders		Competition with most innovative approaches getting to go to overseas conference to present their findings	Package of coaching support for percentage of participants with expectation that they provide peer support to their colleagues	6 × 0.5 day of training with intersessional tasks
Variant 5			Hearts and minds/intrinsic/ altruism focused		
Variation Analysis					
	Research on PLD suggests that mandatory training can be just as effective as opt-in training	Our in-house team does not currently have sufficient expertise	Limited evidence in the research supporting extrinsic motivation	One-shot training with no wraparound support; generally ineffective	Spaced training seems to be more effective
Selected Variation					
	Mandatory for all	**External facilitators with gradual-release model**	**Hearts and minds/intrinsic/ altruism focused**	**Agreed package of coaching support for all participants**	**6 × 0.5 day of training with intersessional tasks**

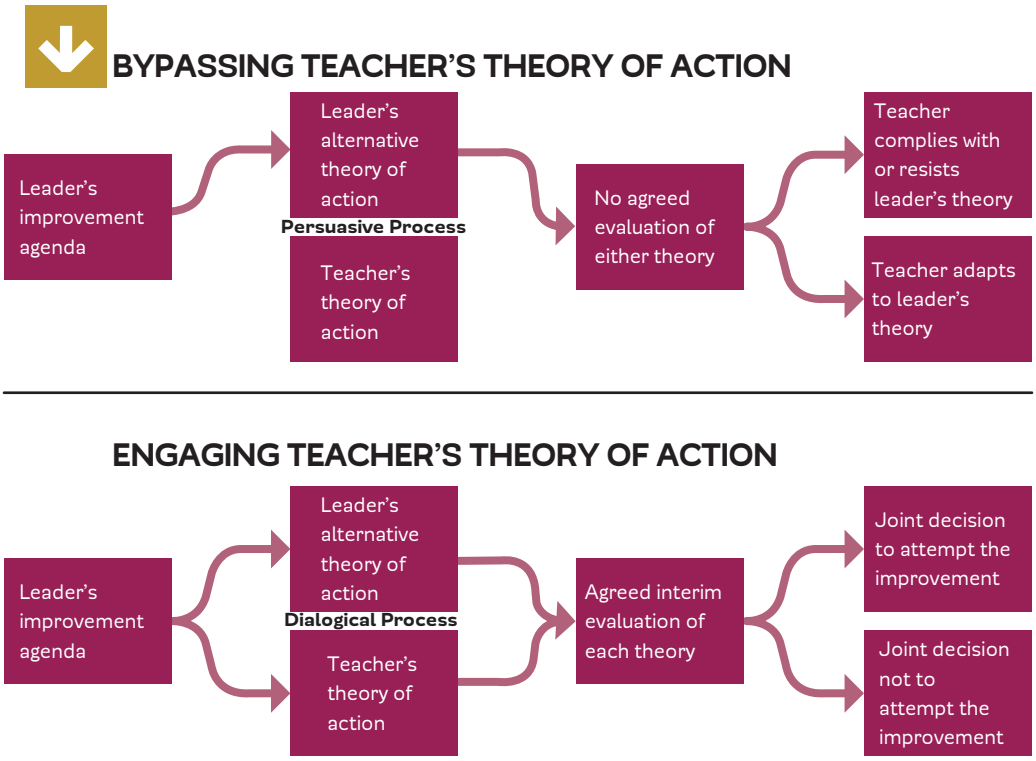
Source: Cognition Education Wiggle Tool. Copyright © Cognition Education (2020). All rights reserved.

A second crucial part to iterating your theory of improvement is to consider the potential impact of stakeholder beliefs and values on successful implementation. From the implementation science approaches in health care and more recently in education, we know that stakeholders' preexisting beliefs, motivations, and collegiality (aka collective efficacy) are a crucial determinant of whether, once you put your project logic model into action, your efforts are successful (Knoster, 1991; Robinson, 2018; Robinson, Hohepa, & Lloyd, 2009).

Viviane Robinson (2018) models the potential outcomes of engaging with versus bypassing educators' theories of action/improvement. Her argument is that improvement initiatives are more likely to be successful when they are dialogical and where they engage with teachers' (and stakeholders') theory of action. Robinson further argues that this engagement is much less likely to be successful if it is merely about gauging stakeholder reactions to a preexisting theory of improvement—in fact, she argues that this is a form of bypass!

Instead, Robinson argues that the dialogue needs to be about inquiring deeply into the local beliefs that sustain the practices that leaders are trying to improve. This includes both understanding stakeholders' espoused beliefs (i.e., what they say) and unpacking their implicit beliefs (i.e., the hunches that drive what they actually do), which they may not be fully able to describe or explain without support to tease out their thinking processes. The test of whether or not leaders are in bypass mode is whether they can clearly articulate the theories of action of those they engage with and the other parties confirm that this is an accurate representation of their beliefs (see Figure 2).

Figure 2 The Engage versus Bypass Model



Source: This is a revised version of Figure 20, page 129, Robinson, V. M. J., Hohepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Best evidence synthesis.* Wellington, New Zealand: Ministry of Education.

We agree. This is why we have recommended that the Goal Hunt involve engagement and collaboration with a wide range of stakeholders to identify an appropriate education challenge and also to collectively build and review causal driver models. This process is dialogical. It shines the light on stakeholder beliefs. And by collaboratively crafting and iterating theories of improvement, during the Opportunity Sift, we engage in yet more rich and wonderful dialogue that builds toward an agreed theory of improvement and an agreed decision to change (or not to change).

However, when it comes time to implement the jointly agreed theory of improvement, it will undoubtedly involve engagement with stakeholders who were not involved in that dialogic process, who were not party to that shared consensus, and who may (violently) disagree and disengage with the improvement process.

Therefore, another recommended process is what we call *Human Factor Analysis*. This involves recollecting all that rich dialogue that the stakeholders involved in the Goal Hunt and Opportunity Sift processes undertook—especially dialogue that generated discomfort, discord, and disagreement. Human Factor Analysis assumes that uninitiated stakeholders will react in similar ways and its purpose is to set out mitigation strategies to all of the key human-related risks.

The stakeholders involved in developing an iterating theory of improvement can use the tool in Table 16 to identify all the risks and the proposed mitigations. The outcomes can then be fed back into the theory of improvement (Table 14) to further iterate and refine this. Many of the human factor risks that are uncovered will likely be similar, if not identical, to those the G.O.L.D. Team also went through in its own process of thesis, antithesis, and synthesis.

If you are planning to scale up your G.O.L.D. initiative across multiple schools, your thinking around the human factor considerations will be crucial.

Table 16 Human Factor Analysis—Worked Example

No.	Human Factor Risk Description	Stakeholders	Likelihood 1–5	Severity 1–5	Impact L × S	Mitigation
1	Project design includes requirement for teachers recording and sharing videos of their lessons—they may be extremely uncomfortable doing this and/or interpret it as an accountability rather than improvement initiative.	**Teachers**	**4**	**3**	**12**	• Use of Iris Connect—so that teachers can control when and to whom they share their videos • Leadership by example—the leadership team will film an exemplar lesson and share video for review at a film club event

No.	Human Factor Risk Description	Stakeholders	Likelihood 1–5	Severity 1–5	Impact L × S	Mitigation
2	Project centers on use of a third-party training program. Our educators may not engage with it because they did not build it and it was developed overseas.	**Teachers**	**5**	**5**	**25**	• Study visit by identified teachers (opinion leaders) to other schools nearby that are using the program—they will then report back to whole teaching body on their findings • Consider option of localizing the materials, i.e., running workshops with our teachers to make modifications to materials (particularly key terms and linkage to organizational vision)
3						
4						

Designing to Scale Up

From research in the health care sector (Gawande, 2010), we know that some types of improvements scale fast and others extremely slowly. For example, health care professionals can be trained in the rationale and benefits of sanitation checklists within hours and be supported to implement and repeat each protocol about 25 times required to achieve automatic/reflexive behavior. This can then be scaled quickly across hospitals and monitored. Whereas training a surgeon to perform robo-surgery is far more complex. It requires more like 200 repetition cycles and must also confront surgeons' preferences for scalpels and their belief that they need to be in the same room as the patient during surgery.

Implementation and the successful scaling of education improvement programs confront similar challenges. Rolling out a new statewide student screening protocol to identify hard-of-hearing learners that takes educators 2 hours to learn and a few minutes per child to implement would be an example of improvement that has greater potential to scale fast. Whereas implementing a Direct Instruction (DI) program is more akin to the transition to robo-surgery: There are lots of protocols for educators to master, assuming they even believe in DI in the first place.

We know that a cognitive bias called the *Ikea Effect* can greatly influence whether implementation of a new approach is effective.

The Ikea Effect

Stakeholders place disproportionate value on products that they partly created, such as Ikea bookcases!

Within the human factor analysis mitigations, you will want to consider how you can leverage the Ikea Effect to enhance scale. By acknowledging that the local ecosystem in each school is different and identifying the implementation wiggles that can be locally adjusted without much detrimental

Figure 3 Models for Scaling Up Education Improvement

Model	Description	Outcome
Incremental Growth Model 	Assumption that the act of providing teachers with training in a specific area will always and automatically result in them understanding what they are being trained in, agreeing with it, and putting it into practice in the manner intended.	**On average, teachers will remember 25% and put less than 10% into practice.**
Cascade Model 	Variant of the incremental growth model. It makes the same assumptions but goes a step further, suggesting that each of those that have been trained in turn go on to train many others—creating a power-law effect.	**Even worse than incremental growth model: a photocopy of a photocopy of a photocopy.**
Unbalanced Growth Model 	Identifying high-performing teachers and sprinkling them like salt across a range of schools to evangelize and to increase the transmission of ideas and change in behavior.	**Mixed. Runs counter to normal career progression trajectory where teachers move to better schools as their careers progress. Also requires appropriate critical mass to avoid crowd-out.**
Cell Division Model 	Asking teachers to innovate a new approach and then, like a breakaway cell, move on to establish a new school from scratch that embodies the innovation.	**Mixed. Limits to how often we can do this before we have to start shutting down old schools. Diversity can also limit ability to evaluate impact.**

Source: Adapted from Elmore (1996).

impact, you create better potential for buy-in and even love.

We admit, however, that more collective thinking and experimentation are required to develop a foolproof approach to effectively scaling up education improvement and that we probably know more about what doesn't work than what does (Figure 3).

2.4. Develop Your Success Map

As Coach Mills knows, things that get measured get improved. That's why during the Goal Hunt (step 1.5) the process involves setting a quantifiable improvement goal. In the case of the worked example (Table 14), it was to increase student attendance to 97%. For larger-scale projects, we also recommend that you further refine the process of measurement.

This might start by reviewing and analyzing all the potential direct and indirect indicators of success (Table 17).

From this review process, you will then likely narrow down to the indicators that

- most directly link to your agreed education challenge,
- are reasonably amenable to data collection,
- are high in both validity (i.e., measure the right *thing*) and reliability (i.e., measure the *thing* in a consistent way).

Table 17 Identifying All the Potential Indicators of Success

No.	Potential Indicator	Linkage to Education Challenge	Ease of Data Collection	Validity and Reliability	Perverse Incentives?	Conclusion
1	**Student Attendance**	**Direct and strong.** Our agreed challenge is nonattendance, particularly in adolescent boys.	**Easy.** We already collect attendance data twice per day.	**High validity.** Direct measure of the education challenge. **High reliability.** Binary measure that is not open to subjective interpretation.	Need to consider whether there is any incentive for stakeholders to falsify attendance data.	
2	**Lesson Observation Data**	**Indirect and causal.** Our theory of improvement postulates that one of the reasons students are not attending is that they do not enjoy their classes. Our assumption is that improvements in lesson observation scores will correlate with increased student attendance.	**Medium.** We already undertake two lesson observations per annum but we are not currently using a structured rubric or training observers to increase inter-rater reliability.	**Medium validity.** We are not yet sure whether students are not attending because they do not enjoy their lessons. But by collecting these data and correlating them we will have a much better understanding of the drivers. **Low reliability.** The current rubrics are open to wildly different interpretation. We need to consider an alternative frequency measure.	Potential that teachers will prep and stage their "best lesson" for the observation—this will mean that what the observer sees is not representative of what takes place "on average" in that teacher's classroom.	
3	**Teacher Training Satisfaction Scores**	**Indirect.** Assumes that teachers who enjoyed the training are more likely to put it into practice.	**Easy.** We have standardized "happy-sheets."	**Low validity.** Not convinced that teacher perceptions of training are strongly connected to long-term impact.		

As part of the reviewing process, you will also want to consider whether the act of measuring creates the potential for any perverse incentives and how you will go about mitigating that potential.

It is extremely likely that as you embark on the reviewing process you will start with quite a long list of potential indicators but that as you proceed the finalized list becomes narrower and more specific. You will then want to set the baseline values for each agreed indicator and the target values for improvement across time. Table 18 gives a worked example for this exercise.

The outputs from Table 18 are then fed into your theory of improvement model and they enable you to complete the short-, medium-, and long-term outcomes components at the far right of Table 14.

The next step is to Lift Off and deliver the agreed improvement initiative.

Table 18 Setting Baseline and Target Values for Each Selected Indicator

No.	Indicator	Instrument	Baseline Value	Target Value T1	Target Value T2	Target Value T3	Target Value T4
1	**Overall student absenteeism**	Student attendance register	**9% absence**	**7%**	**5%**	**4%**	**3%**
2	**Adolescent boys' (13–17) absenteeism**	Student attendance register	**16% absence**	**11%**	**8%**	**6%**	**5%**
3	**Teacher lesson observation score**	District lesson observation instrument	**55 average score per teacher**	**61**	**67**	**69**	**72**

Phase 3 Lift Off

The Lift Off phase is more likely to be effective where the following conditions are met:

- A project management office is established (for larger-scale/multi-institutional projects).
- Robust project management processes are deployed. This means that there is a high level of clarity about who is supposed to be doing what and by when and whether they are on target.
- Delivery is collaborative and iterative.

After having defined your education challenge, developed and assumptions-tested several theories of improvement, and identified the stronger and weaker bets for improvement—you are, finally, ready to launch. Whether you select the single best bet and commit all your resources to it or run parallel pilots for the two to five most plausible project designs will depend on the resources at your disposal.

We have remarkably little to say about the process of project or program implementation. There are several feasible approaches to implementation, like PRINCE2; Michael Barber's Deliverology; and, for smaller projects, Agile. These approaches involve setting up a project management office and utilizing tools such as those listed in Table 19.

Table 19 Standard Project Management Tools for Larger-Scale Initiatives

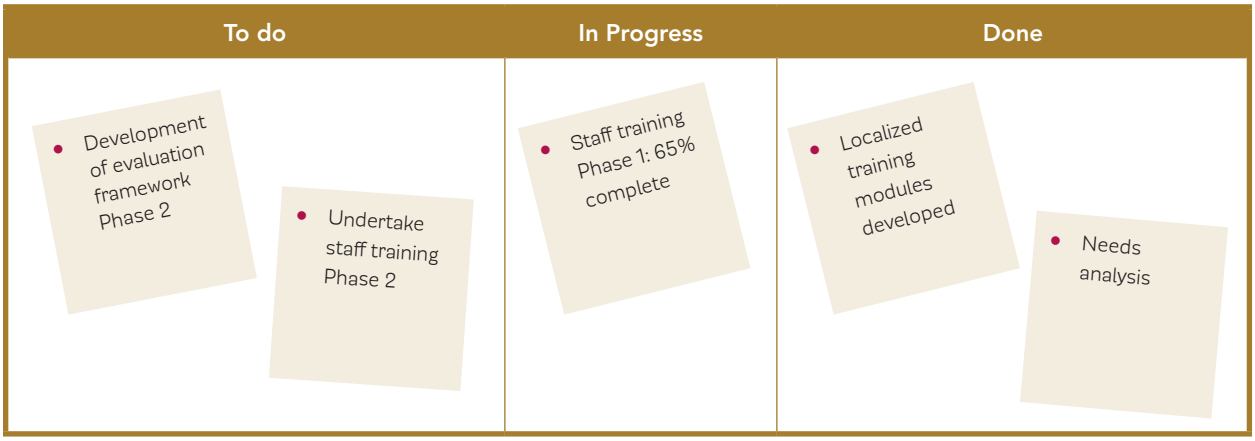
Tool	Description
Project Charter (Project Initiation Document)	This is the key project "bible," which contains the rationale for undertaking the project, budget, timelines, project activity, and breakdown structure.
Project Governance Framework	Describes who is accountable, responsible, and informed for each project component.
Risk Register	Details all project risks (i.e., what *could* go wrong) and includes risk ranking and a mitigation plan.
Issues Log	Details all live issues (i.e., things that *have* gone wrong) with suggested remediation plans for consideration by the project sponsor.
Project Progress Report	Used to report on progress to the project board.
Project Change Request	A document for formally requesting substantive change to the delivery of a project (e.g., the timelines, inputs, or outcomes) along with a rationale.
Project Plan	A table format document that lists tasks/subtasks; start-and-end dates; resource allocation/owner; percentage complete. For smaller projects a simple Kanban table that lists to do, in progress, and completed would likely suffice.

(Continued)

(Continued)

Tool	Description
Communications Plan	A document that lists key project stakeholders both within and outside the project team and which outlines key communication channels, messages, and timelines—linked to key strands of project activity.
Product Acceptance Criteria	This details the purpose of each product (output), its constituent elements, and the acceptance criteria/quality standards for signoff.
Evaluation Plan	This details the evaluation methodology, data collection tools, processes, and cycles for review and improvement.
Lessons-Learned Logs	Used as part of a regular review cycle to identify what has worked well versus what has not worked at all and how this can be improved.

Figure 4 Kanban Table—Worked Example



When utilized properly, these project management approaches work very well for scheduling inputs and measuring whether they have resulted in agreed outputs. There are a number of successful software, infrastructure, and organization redesign projects across multiple sectors that attest to the value of these project management approaches. How they are used will depend on the scale of your project.

If you are delivering an education challenge at the district, state, or national level, it is extremely likely that you will want to establish a formalized project office and deploy your selected project management methodology in full. Where delivery is taking place at an individual school, we suspect you might want to have a lighter touch and simply keep track of activities, such as through a Kanban table. Figure 4 gives a worked example of a Kanban table.

A Kanban review could involve key project stakeholders coming to the start of each day to review what they are doing, what they have done, and what they are still to do.

However, there is a major difference between a project successfully achieving all of its deliverables or outputs and this directly translating into changed lives. This is the distinction between outputs (i.e., ticking off *things* on your to-do list) and outcomes (the *things* resulting in impact). To make sure that your project achieves the expected outcomes, you need to embed evaluation from the get-go. This means that you need processes to Double Back after you Lift Off.

Phase 4a Double Back

Attempts to resolve large-scale education challenges are more likely to be successful when the following conditions are present:

- An explicit attempt is made to evaluate whether the initiative has generated sufficient impact.
- The evaluation methodology is agreed on and established before the education challenge initiative commences.
- Appropriate indicators of improvement are selected, their baseline values are measured prior to project commencement, and then these indicators are tracked over time.
- The evaluation method is designed to measure what has been done (monitoring) and whether it will result in the intended impact (evaluation), which helps to guide project improvement (iteration).
- The outcomes of evaluation are regularly used to refine/iterate the delivery of the project to enhance/amplify impact.

In other words, we need to **Double Back** in order to verify and enhance our impact (Table 20).

Resolving education challenges is an extremely complex undertaking. Solutions that have been effective in other contexts often fail to replicate in new terrains; sometimes they even make things worse. This can be because of the following:

- The causal drivers of the education challenge are different in each context, even though problems they generate are the same. This means that different interventions may have been far more effective.
- The intervention is not delivered with fidelity and/or the "human factor" is not fully accounted for, resulting in lack of buy-in or engagement with stakeholder beliefs and values.

Often, delivery failure also occurs because of the following:

- **There is no evaluation of what has been done.** This means that there is no way of knowing whether the time and resources invested have resulted in any tangible improvement. **OR**
- **There is evaluation, but it happens as an afterthought**. Without baseline data it is difficult to determine whether any improvement has occurred. **AND/OR**
- **The evaluation design is poor.** For example, it measures whether things have been done (ticking off milestones) or changes in educator/intermediate-stakeholder behaviors, rather than whether student learning/end-stakeholder outcomes have been enhanced. **AND/OR**
- **The outcomes of evaluation are not fed back into the iteration of the design/delivery process.** The project continues to be delivered "as is" even when there is no evidence of impact.

From the work of Kurt Gödel and Alan Turing, we know that there exist many *undecidable problems* (Hofstadter, 1999). These are basically

questions that we are unable to answer in advance, questions that we can only answer by doing, observing, and recording. Whether theory of improvement A, B, or C or wiggles 1, 2, or 3 will have more success in achieving a given education outcome are examples of *undecidable problems*. We can make an intelligent guess, based on past data, but we will only know for sure once we have rolled the dice and begun to watch the pieces in play.

To systematically watch and reflect on each of these twists and turns, you need to have a robust set of project evaluation tools and protocols. We already embedded these evaluation processes during the earlier stages of the G.O.L.D. methodology (Table 20).

The key reasons for doubling back are to understand the following:

- *Whether* the education challenge has been sufficiently addressed. Is student attendance increasing in line with our targets?
- *Why* what you have implemented worked (or did not work). Are there aspects of our theory of improvement and our delivery activities that could be enhanced to deliver even greater impact?

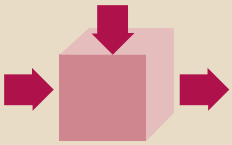
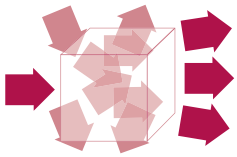
Some of the approaches to evaluation that you could undertake include the evaluation types shown in Table 21.

Table 20 Embedding Double Back or Evaluation Throughout G.O.L.D.

Embedded Double-Back Processes

Goal Hunt	Opportunity Sift	Lift Off
1.1–1.3: In identifying and ranking education challenges, you are drawing on local and global data to *evaluate* which are most worthy of resolution. **1.4:** By modelling and validating causal drivers, you are laying the groundwork for step 2.1. **1.5:** By defining what success looks like, you have taken a baseline reading and established a quantifiable improvement goal to *evaluate* against.	**2.1**: By sketching design opportunities and mapping these against global and local data, you are *evaluating* which have the highest probability of impact. **2.2–2.3**: By iteratively developing your theory of improvement, you are explicitly pre-evaluating the activities and wiggles most likely to generate impact. **2.4:** By developing a detailed inputs, outputs, and outcomes map, you build the work you undertook in step 1.5 into a full map of success to measure against.	Through use of • Project Charter • Project Plan • Risk and Issues Registers • Product Acceptance Criteria you are able to measure your inputs, e.g., whether you are delivering all your milestones on time and to budget and whether your products meet the agreed acceptance criteria. You will also develop an **Evaluation Plan**, which is a detailed statement of how, when, why, and by whom for evaluation—which is linked to steps 2.2–2.4. The ultimate purpose of your evaluation activity is to double back in order to double up the impact!

Table 21 Approaches to Double Back/Evaluation

Evaluation Type	Description
Black Box Evaluation  **Key Question:** Does it work?	• Pre- and post-intervention data are collected (e.g., quantifiable achievement scores, attitudinal survey data, coded observations). • Calculation of effect size to determine whether there was an improvement in post-intervention score. • Can also include a control group to add counterfactual dimension and give greater confidence that any improvement was not due to chance or placebo effect. • **BUT** does not give deeper insights into why/which dimensions of the project design created impact. These remain inside the black box. • Also generally requires at least 12 months of intervention activity before post-test evaluation is likely to show improved outcomes. In the intervening period, all parties are flying blind.
Grey Box Evaluation  **Key Question:** Does it work and why do people think it works?	• Combines black box techniques with qualitative data collection including focus groups, interviews, and open-text surveys to gather participant data to make inferences about which aspects of the intervention model were more/less effective. • **BUT** the qualitative data collected centers on participant *perceptions*, which may or may not focus on the appropriate causal variables, i.e., potential for sunk cost fallacy, bandwagon effect, anecdotal fallacy.
Clear Box Evaluation  **Key Question:** What works for whom, in what contexts, to what extent, and how?	• Principal focus is on forensic analysis of every aspect of the project that could be iterated/wiggled and what the consequences would be of tweaking each. • Strong linkage to the processes in Opportunity Sift—wiggle testing, where we detail all of the elements of each intervention that can be varied. • Emphasis on thought experiment; i.e., what would happen if we increased the duration of training, included coaching, delivered entirely via video conference, etc.? • Asks: What works for whom, in what context, how?

Source: Adapted from *Evaluation Practice Toolkit*. (n.d.). Developed by King's College London Clinical Education and Patient Safety Research Group in partnership with Health Education England.

Our recommendation is that, where possible, you consider a Clear Box evaluation. Education gold is knowing what works for whom, in what context, and how. On larger-scale programs, it is likely that you will want/need to co-opt trained evaluation professionals to your G.O.L.D. Team and some of the mixed methods tools you might consider using for evaluation include those described in Table 22.

Table 22 Tools for Evaluation

Evaluation Tools	Description	Pitfalls
Lesson Observation	Watching and coding lessons at regular intervals using a standardized observation rubric	Challenges with inter-rater reliability, i.e., observers often cannot agree on what they see
Student Achievement Data	Getting all students to sit standardized assessment tests at regular intervals and measuring progress	Learner success is wider than improvements in student assessment tests
Teacher and Leader Mindframes Surveys	Getting teachers and leaders to take psychometric surveys at regular intervals that measure their attitudes about teaching and leadership	Not always clear that changes in survey outcomes reflect changes in attitudes or that changes in attitudes result in changes in practice
Student and Parent Voice	Asking students and parents to reflect on their perceptions of change within the school	Placebo effect, i.e., merely being asked can be enough for students and parents to believe something has changed; it may well be that the only thing that's changed is that they have been asked
Participant Satisfaction Surveys	Asking participants in training/support interventions to rate the quality of what they have experienced and the likelihood that they will put it into practice	Perceptions of effectiveness/habit change are not always the same as reality, and people often commit to implementing things that they then fail to carry out

By collecting and regularly reviewing formative evaluation data at every stage of project implementation, we can begin to develop hypotheses about elements of the project that are working well and those that are not working at all.

In traditional implementation cycles, the evaluation data are harvested and reviewed at the end. This culminates in a weighty external evaluation report that details lessons learned. But our key message is that it's a sin to wait. The *undecidable problem* means that we have absolutely no idea whether the lessons learned in context A would be relevant in the slightest for context B. So if we are going to bother collecting evaluation data, we should put them to work as quickly as possible in the context in which they are being collected.

By reviewing the data in real time, we can engage in what Deborah Eyre (2007/2016) calls *structured tinkering* and what Pritchett, Samji, and Hammer (2013) call *crawl and learn*. In other words, we can wiggle on the go by varying different project features and then reviewing the impact that this has on the evaluation data. Through this continued process of varying and reviewing, we can gradually inch closer and closer to the best possible theory of improvement and best possible implementation wiggles in our given context. We will have built the best sprinting shoes the world has ever seen, although whether these will help athletes in the newest Olympic event of Sport Climbing remains to be seen.

Phase 4b From Double Back to Double Up

We suspect that many readers of this paper might say, "We can see how G.O.L.D. could be used to drive more effective improvement at individual schools but we are struggling to see how this impact can be scaled up with fidelity in multiple settings." In this final section, we want to address the question of scale head-on.

Many researchers and implementers have defined scale as the widespread use of the program or intervention (Adelman & Taylor, 1997; Fullan, 2000; Sloane, 2005; Stringfield & Datnow, 1998). Fullan (2000), for example, defines scale as a *minimum* of 50 schools and 20,000 students. We prefer to be less prescriptive and to simply think about scale in terms of multiple sites or settings for improvement and to see scale more in terms of fidelity *of* impact rather than consistent use of a specific intervention.

We also like to think about scaling (or doubling) up from the following two perspectives: (1) grain size and (2) transmission mechanism.

1. Grain Size

Grain size is the magnitude of change required in each site or setting for improvement to be implemented. We distinguish between fine grain, coarse grain, and rock.

Fine Grain

Fine grain is a change that is relatively easy to implement, such as the screening protocols for identifying students with hearing difficulties that we discussed earlier. This extent of change can be easily described, requires limited training for educators to be able to follow the process with fidelity, and needs relatively few repetition cycles before it becomes ingrained. And the diagnosis upside for learners with hearing difficulties is bountiful.

Coarse Grain

Considerably harder to implement, coarse grain is a magnitude of change where educators require much more support and where their preexisting beliefs may act as a block to adoption of the improvement program. However, coarse grain change is relatively self-contained and does not require an educator to change every aspect of what they do simultaneously. A good example is our work with Iris Connect, a platform for educators to video record their lessons for self- and peer reflection. The platform and tools are highly structured and the effect size gain (of $d = 0.88$ from video analytics) makes it a useful tool in contexts where the goal is to improve the professional competence and efficacy of teachers. However, to maximize impact, school leaders need to give thought to how the improvement program will be implemented, and teachers require time and support in order to hone their reflective-practitioner skills. This is why we have wiggle tested a range of deployment models to identify the ones that have highest probability of unlocking school-wide adoption, across networks of schools.

(Continued)

(Continued)

1. Grain Size

Rock

Rock is a magnitude of change that is extremely hard to implement. It faces a confluence of challenging variables, including these:

- Flying in the face of prevailing teacher beliefs and values in the settings where G.O.L.D. Teams attempt deployment.
- The amount of change to practice and processes is extremely high. This might be as severe as expecting educators to discard everything they currently do and to reboot with an entirely new "operating system."
- High levels of ambiguity in how to implement the change or what success looks like. Instruction manuals, protocols, or checklists are not provided to support educators in implementing change in their contexts.

An example of such change might be scripted teaching approaches. While the delivery protocols are highly structured (after all it involves teaching by following a prewritten script), they also involve changing every aspect of what a teacher does in the classroom, and very few teachers are likely to see themselves as actors that follow a script.

The finer the grain and the less it flies in the face of educators' beliefs, the higher the probability of achieving scale—although often this low-hanging fruit does not generate impact.

2. Transmission Mechanism

Transmission mechanism deals with the approach to convince educators to engage with new ways of thinking and working. Here we distinguish between viral transmission, replication, and adaptation (Morel, Coburn, Catterson, & Higgs, 2019).

Viral Transmission

Viral transmission occurs when a useful idea, process, or set of tools gradually permeates large swaths of the education sector through word of mouth, positive feedback, or research from early adopters and figures of authority to gradually and organically become an accepted or even default method. An example would be the way Bloom's Taxonomy has moved from limited use to widespread adoption in curriculum development, instruction, and assessment, although educators use Bloom's in a host of different ways (Schneider, 2014).

The core features of viral transmission are that it is largely unplanned, it just spreads through network effects, and the fidelity/returns from implementation are extremely variable.

Replication

Replication is about specifically designing a precisely structured intervention, with the explicit purpose of replicable and repeatable implementation. This might be done via the following steps:

- Piloting different methods of implementation in a range of contexts
- Evaluating the benefits of each
- Locking an agreed method and then building comprehensive protocols and instructions for use that can be followed almost like a cooking recipe
- Providing a package of specifically designed gradual-release support, including
  - spaced learning and reflection opportunities,
  - modelling,
  - feedback, and
  - peer mentoring and/or expert coaching.

2. Transmission Mechanism

Unlike viral transmission, replication is specifically and deliberately planned with the intention of achieving fidelity. This means that educators use the approach, tools, or strategies exactly as intended by the designers. However, novice teachers are likely to be more open to replication approaches than are experienced teachers. The latter are more likely to feel de-professionalized where they are asked to discard their prior habits and implement a new approach without deviation (Morel et al., 2019).

A subcategory is what we call meta-replication. This is not about replicating an intervention. Instead it centers around replicating a process like Visible Learning G.O.L.D. that is used to diagnose education challenges, develop interventions, implement, and evaluate. However, even G.O.L.D. does not present a single way but more a set of suggested processes and protocols, which brings us to our final category.

Adaptation

Adaptation is about the scaled implementation of an intervention while giving explicit encouragement or endorsement to local educators to make suitable adaptations so that it is more relevant and useful to their local context (Clarke & Dede, 2009; Fishman, 2005; Means & Penuel, 2005; Sisken, 2016; Wiske & Perkins, 2005). This might include explicit use of our Wiggle Tool to explore all the different ways that an intervention can be iterated. Or adaptation might emerge more organically as a consequence of implementing in a new context. The assumption is that these local innovations enhance effectiveness either through making the intervention fit better or simply by generating greater buy-in because local stakeholders think of it as *theirs*.

G.O.L.D. would perhaps fit into the category of meta-adaption: A set of suggested processes (which can be varied) that have been fashioned to help educators discover local needs and then design, implement, and evaluate locally suitable interventions. Our belief is that there might be more mileage in scaling up G.O.L.D.-like approaches than in scaling up a specific intervention that may be suitable for some contexts but not for others. We then get fidelity of impact rather than uniform implementation of approaches that add little value in the local context.

Conclusion

It's both *what* you do and *how* you do it—that's what gets results. And in the current age we need results more than ever. We can't leave student progress to chance by implementing random measures and giving up halfway. Granted, education systems have made phenomenal progress with education reform over the last 170 years but often the last few hundred yards to the finish line are the hardest. The ground gets steeper and steeper and at the very end, it's almost vertical. It can take more effort to travel those few yards than the cumulative energy required to run all the miles that came before.

Success requires thorough planning, and the G.O.L.D. method supports this through four phases (Table 23).

Every child deserves a year's growth for a year's input, now more than ever. Let's win this race together!

Table 23 The Four Phases of G.O.L.D.

Goal Hunt	Opportunity Sift	Lift Off	Double Back
1. Find/diagnose an education challenge worth solving (or go home) 2. Map the causal drivers 3. Agree what success looks like	1. Reverse engineer from causes to opportunity sketches 2. Build as many competing theories of improvement as possible and select the most plausible 3. Produce detailed outcome maps and feed this back into the theory of improvement	1. Establish project management office 2. Undertake delivery using project management tools 3. Monitor implementation	1. Evaluate at regular intervals to ask: *what works for whom, in what contexts, to what extent and how?* 2. Use data from evaluation to collaboratively review whether program is achieving intended effect and what (if any) changes should be made during implementation 3. Use the Double Back as a mechanism to Double Up
Ensures we are entering the right race	**Equips us with the tools to cross the finish line**	**The race begins!**	**We can systematically change our pace, our technique, and even our running shoes while we are on the track—keeping what works and discarding the rest.**

References

Aarons, G. A., Green, A. E., & Miller, E. (2012). Researching readiness for implementation of evidence-based practice: A comprehensive review of the evidence-based practice attitude scale (EBPAS). In B. Kelly & D. F. Perkins (Eds.), *Handbook of implementation science for psychology in education* (pp. 150–164). New York, NY: Cambridge University Press.

Adelman, H. S., & Taylor, L. (1997). Toward a scale-up model for replicating new approaches to schooling. *Journal of Educational and Psychological Consultation*, *8*(2), 197–230.

Barber, M., Kihn, P., & Moffit, A. (2011). *Deliverology 101: A field guide for educational leaders*. Thousand Oaks, CA: Corwin.

Berman, P., & McLaughlin, M. W. (1975). *Federal programs supporting educational change: Vol. IV the findings in review*. Santa Monica, CA: RAND.

Bishop, R., O'Sullivan, D., & Berryman, M. (2010). *Scaling up education reform: Addressing the politics of disparity*. Wellington, New Zealand: NZCER Press.

Blase, K., Kiser, L., & Van Dyke, M. (2013). The hexagon tool: Exploring context. Chapel Hill: National Implementation Research Network, FPG Child Development Institute, University of North Carolina at Chapel Hill.

Breakspear, S. (2017). Embracing agile leadership for learning: How leaders can create impact despite growing complexity. *Australian Educational Leader*, *39*(3), 68–71.

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

Clarke, J., & Dede, C. (2009). Design for scalability: A case study of the River City curriculum. *Journal of Science Education and Technology*, *18*(4), 353–365.

Cohen, D. K., & Moffitt, S. L. (2009). *The ordeal of equality: Did federal regulation fix the schools?* Cambridge, MA: Harvard University Press.

Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. *American Educational Research Journal* *44*(2), 298–339.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms 1880–1990* (2nd ed.). New York, NY: Teachers College Press.

Cuban, L. (1998). How schools change reforms: Redefining reform success and failure. *Teachers College Record, 99*(3), 453–477.

Darling-Hammond, L., & Snyder, J. (1992). Curriculum studies and traditions of inquiry: The scientific tradition. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 41–77). New York, NY: Macmillan.

Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, *66*(1), 1–27.

Elmore, R. F., & McLaughlin, M. W. (1998). *Steady work: Policy, practice, and the reform of American education*. Santa Monica, CA: RAND.

Evaluation Practice Toolkit. (n.d.). Retrieved from https://www.kcl.ac.uk/study/learningteaching/kli/research/articles/docs-tom-simpson/kcl-li-clinical-evaluation-practice-toolkit-interactive-aw.pdf

Eyre, D. (2007/2016). Structured tinkering: Improving provision for the gifted in ordinary schools. *Gifted and Talented International*, *22*(1), 31–38. doi:10.1080/15332276.2007.11673483

Fishman, B. J. (2005). Adapting innovations to particular contexts of use: A collaborative framework. In C. Dede, J. P. Honan, & L. C. Peters (Eds.), *Scaling up success: Lessons from technology-based educational improvement* (pp. 48–66). San Francisco, CA: Jossey-Bass.

Ford, M. (2017). *The rise of the robots: Technology and the threat of mass unemployment*. London, England: Oneworld.

Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change*, *1*, 5–28.

Fullan, M., & Scott, G. (2014). *New pedagogies for deep learning whitepaper: Education plus*. Seattle, WA: Collaborative Impact SPC.

Galton, F. (1907). Vox populi. *Nature*, *75*(1949), 450–451. doi: 10.1038/075450a0

Gawande, A. (2010). *The checklist manifesto: How to get things right*. New York, NY: Picador.

Harari, Y. N. (2019). *21 lessons for the 21st century*. Toronto, Canada: CELA.

Hargreaves, A., & Fullan, M. (2012). *Professional capital transforming teaching in every school*. New York, NY: Teachers College Press.

Hattie, J. (2015). *What doesn't work in education: The politics of distraction*. London, England: Pearson.

Hattie, J., & Hamilton, A. (2020). As good as gold? Why we focus on the wrong drivers in education. *Gold Paper*. Thousand Oaks, CA: Corwin.

Hattie, J., & Zierer, K. (2018). *Ten mindframes for visible learning: Teaching for success*. London, England: Routledge.

Hofstadter, D. R. (1999). *Gödel, Escher, Bach: An eternal golden braid*. New York, NY: Basic Books.

Jefferson, M., & Anderson, M. (2017). *Transforming schools: Creativity, critical reflection, communication, collaboration*. London, England: Bloomsbury Academic.

Kania, J., & Kramer, M. (2011). Collective impact. *Stanford Social Innovation Review*, Winter.

Kelly, B. (2012). Implementation science for psychology in education. In B. Kelly & D. F. Perkins (Eds.), *Handbook of implementation science for psychology in education* (pp. 3–12). New York, NY: Cambridge University Press.

Kelly, B., & Perkins, D. (Eds.). (2012). *Handbook of implementation science for psychology in education*. New York, NY: Cambridge University Press.

Knoster, T. (1991). *Factors in managing complex change*. Material presentation at TASH conference. Association for People with Severe Disabilities, Washington, DC.

LeMahieu, P. G., Nordstrum, L., & Gale, G. (2017). Positive deviance: Learning from positive anomalies. *Quality Assurance in Education*, *25*(1). doi:10.1108/QAE-12-2016-0083

Means, B., & Penuel, W. R. (2005). Scaling up technology-based educational innovations. In C. Dede, J. P. Honan, & L. C. Peters (Eds.), *Scaling up success: Lessons from technology-based educational improvement* (pp. 176–197). San Francisco, CA: Jossey-Bass.

Meyers, D. C., Durlak, J. A., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps in the implementation process. *American Journal of Community Psychology*, *50*(4), 462–480.

Morel, R. P., Coburn, C., Catterson, A. K., & Higgs J. (2019). The multiple meanings of scale: Implications for researchers and practitioners. *Educational Researcher 48*(6), 369–377. doi: 10 .3102/0013189X19860531

OECD. (2019). Secondary graduation rate. Retrieved from https://data.oecd.org/students/secondary-graduation-rate.htm

Ohno, T. *(1988). Toyota production system: Beyond large-scale production.* Portland, OR: Productivity Press.

Pascale, R., Sternin, J., & Sternin, M. (2010). *The power of positive deviance: How unlikely innovators solve the world's toughest problems*. Boston, MA: Harvard Business Review.

Pressman, J. L., & Wildawsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland* (3rd ed.). Berkeley: University of California Press.

Pritchett, L., Samji, S., & Hammer, J. (2013). It's all about MeE: Using structured experiential learning ("e") to crawl the design space. (CGD Working Paper 322). Washington, DC: Center for Global Development. Retrieved from http://www.cgdev.org/publication/its-all-about-mee

Rivlin, A. M., & Timpane, P. M. (1975). *Planned variation in education: Should we give up or try harder?* Washington, DC: Brookings Institution.

Robinson, V. M. J. (2018). *Reduce change to increase improvement.* Thousand Oaks, CA: Corwin.

Robinson, V. M. J., Hohepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why. Summary of the Best Evidence Synthesis* (*BES*). Wellington, New Zealand: Ministry of Education.

Rowan, B., Correnti, R., Miller, R. J., & Camburn, E. M. (2009). *School improvement by design: Lessons from a study of comprehensive school reform programs*. Philadelphia: University of Pennsylvania Graduate School of Education, Consortium for Policy Research in Education.

Sahlberg, P. (2015). *Finnish lessons: What can the world learn from educational change in Finland?* New York, NY: Teachers College Press.

Schneider, J. (2014). *From the ivory tower to the schoolhouse: How scholarship becomes common knowledge in education*. Cambridge, MA: Harvard Education Press.

Sisken, L. S. (2016). Mutual adaptation in action. *Teachers College Record*, *118*(13), 1–16.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15–21.

Sloane, F. C. (2005). The scaling of reading interventions: Building multilevel insight. *Reading Research Quarterly*, *40*(3), 361–366.

Stringfield, S., & Datnow, A. (1998). Introduction: Scaling up school restructuring designs in urban schools. *Education and Urban Society*, *30*(3), 269–276.

Timperley, H., Kaser, L., & Halbert, J. (2014). *A framework for transforming learning in schools: Innovation and the spiral of inquiry*. Victoria, Australia: Centre for Strategic Education.

UNESCO Institute for Statistics (UIS). (2018). *One in five children, adolescents and youth is out of school*. (UIS Fact Sheet No. 48 UIS/FS/2018/ED/48). Retrieved from http://uis.unesco.org/sites/default/files/documents/fs48-one-five-children-adolescents-youth-out-school-2018-en.pdf

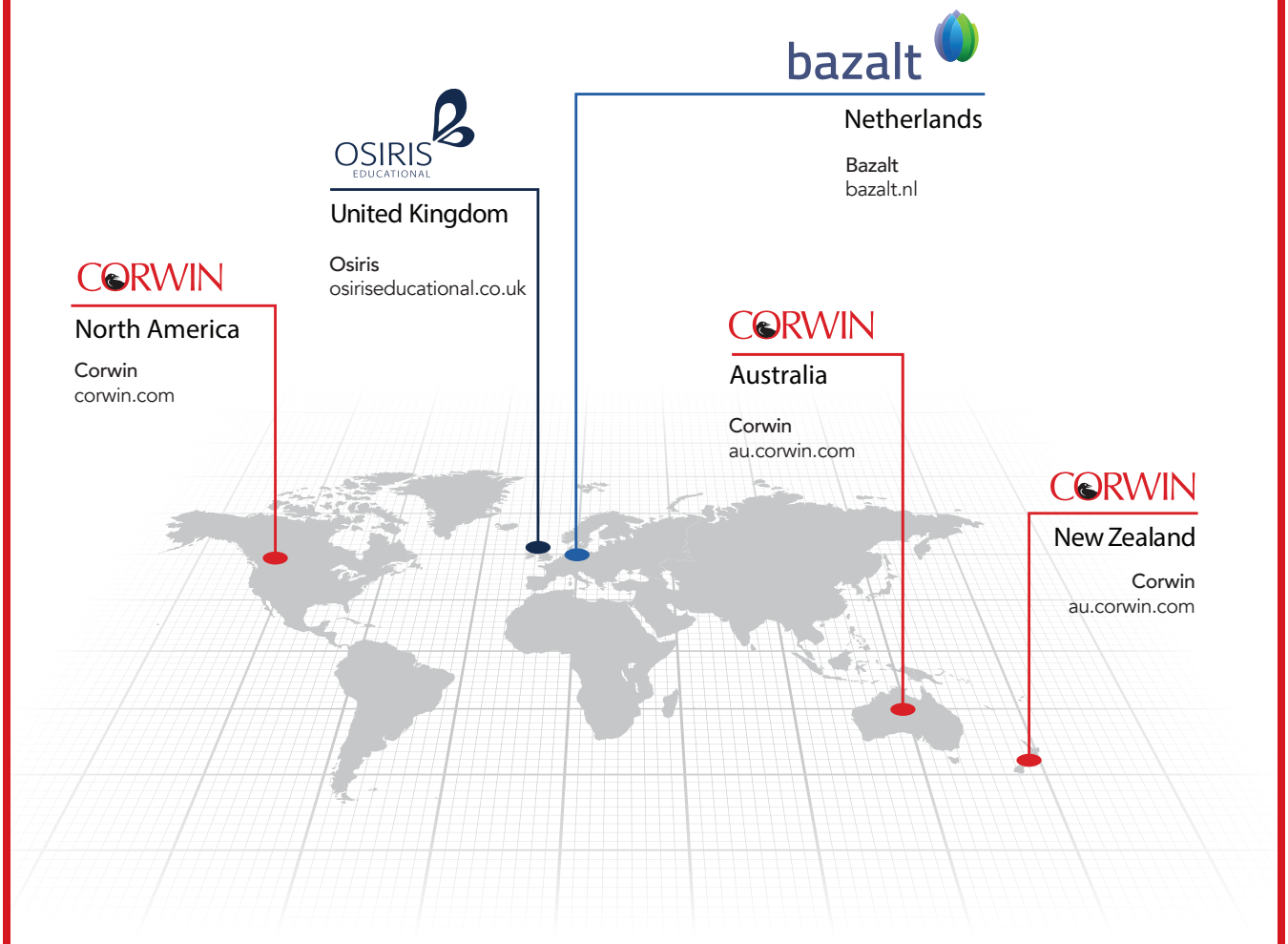
Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., . . . Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology, 41*(3), 171–181.

Wiske, M. S., & Perkins, D. (2005). Dewey goes digital: Scaling up constructivist pedagogies and the promise of new technologies. In C. Dede, J. P. Honan, & L. C. Peters (Eds.), *Scaling up success: Lessons from technology-based educational improvement* (pp. 27–47). San Francisco, CA: Jossey-Bass.

CORWIN

To learn more

and get involved in the Visible Learning$^{plus®}$ Global Network

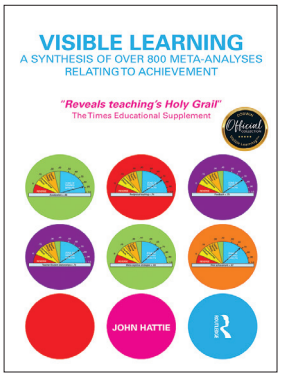


CORWIN Visible Learning$^{plus®}$

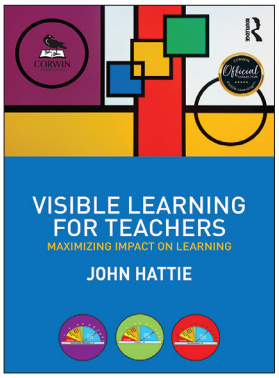
Contact Us: **www.visiblelearningplus.com** | **www.corwin.com**
Email: **visiblelearning@corwin.com** | Twitter: **@VisibleLearning**

CORWIN

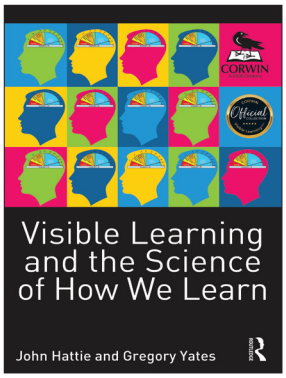
Build your Visible Learning® library!



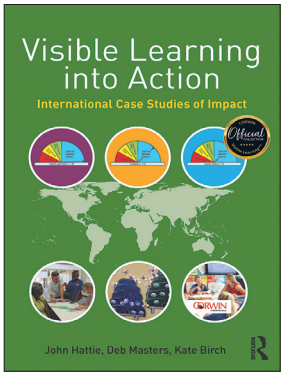
VISIBLE LEARNING



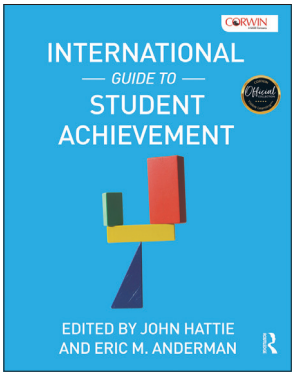
VISIBLE LEARNING FOR TEACHERS



VISIBLE LEARNING AND THE SCIENCE OF HOW WE LEARN

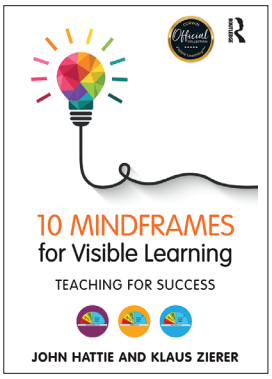


VISIBLE LEARNING INTO ACTION

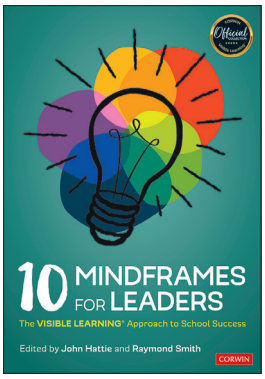


INTERNATIONAL GUIDE TO STUDENT ACHIEVEMENT

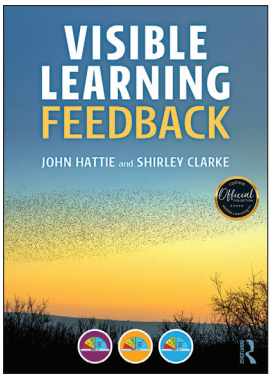
Visit corwin.com/vlbooks



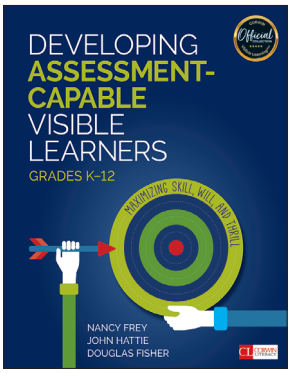
10 MINDFRAMES FOR VISIBLE LEARNING



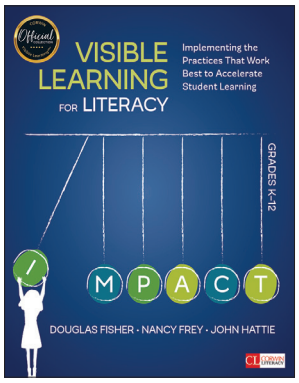
10 MINDFRAMES FOR LEADERS



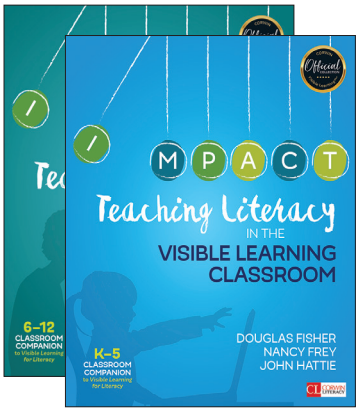
VISIBLE LEARNING FEEDBACK



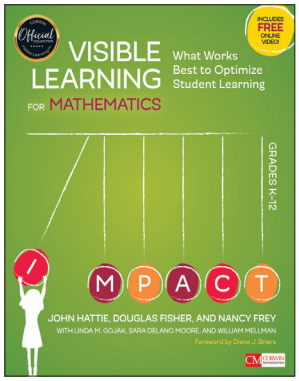
DEVELOPING ASSESSMENT-CAPABLE VISIBLE LEARNERS, Grades K–12



VISIBLE LEARNING FOR LITERACY, Grades K–12

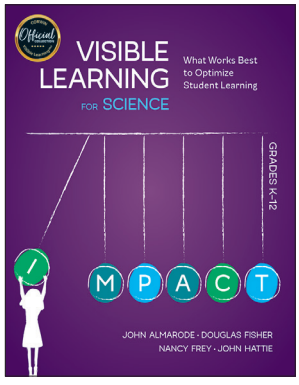


TEACHING LITERACY IN THE VISIBLE LEARNING CLASSROOM, Grades K–5, 6–12

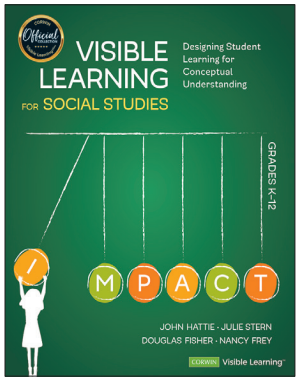


VISIBLE LEARNING FOR MATHEMATICS, Grades K–12

TEACHING MATHEMATICS IN THE VISIBLE LEARNING CLASSROOM, Grades K–2, 3–5, 6–8, & High School



VISIBLE LEARNING FOR SCIENCE, Grades K–12



VISIBLE LEARNING FOR SOCIAL STUDIES, Grades K–12

Helping educators make the greatest impact

CORWIN HAS ONE MISSION: to enhance education through intentional professional learning.

We build long-term relationships with our authors, educators, clients, and associations who partner with us to develop and continuously improve the best evidence-based practices that establish and support lifelong learning.

Ignite the global passion for learning

COGNITION EDUCATION GROUP is a leading provider of education consultancy, professional learning, teacher recruitment, early years and primary tutoring, e-learning and publishing services. Headquartered in New Zealand and operating around the world, our focus is to build the capability and expertise of educators and leaders to improve educational outcomes for all.