Guide to Predicting and Measuring Improvement

Introduction and Framing

Practical measurement is nothing new to teachers. Teachers engage in measurement of their own efforts on a daily basis as they observe their students' attention, understanding and work. When an instructional approach isn't working, teachers see the evidence immediately. This evidence informs everything from on-the-fly changes in the classroom to refinements in lesson plans and curricula. This feedback cycle is at the foundation of how teachers get better at what they do.

While this feedback cycle is often common practice for many individual teachers, it is less common in the shared work of overcoming the challenges that are too big for individuals to take on alone. As our larger efforts require just as much refinement and adaptation as an individual lesson plan, if not more, the role of measurement is paramount in strengthening practice. The effort to measure and learn from change *collectively* is the lynchpin of improvement science, without which the rest of the cycle of learning loses its potency.

This guide is a resource to help schools harness the experience and insight of the people closest to the change to provide the data they need to engage in continuous improvement. It is designed to help schools collect the data they need to answer their own questions and hold themselves accountable to their locally-set goals. While educators will find much of this intuitive and familiar, the larger scale of the measurement effort requires some special consideration. The following norms and practices aim to ensure that schools collect data that is rigorous and useful, without requiring unsustainable commitments of time to do so.

Starting a Research Community

The cycles of learning called for in the *Vision for School Improvement* are an approach to creating this kind of positive feedback cycle in our collective work that is based on the expertise of educators and data of their choosing. In order for this part of the cycle of learning to work well, it is especially important to establish working norms and principles. Doing so will allow the team to take advantage of immediate feedback to rapidly learn and apply lessons as they go. While there is a lot of flexibility built into this kind of measurement that will allow the school-based learning cycles team to pick the methods that suit their context and aim, the following norms enable the learning cycles team to take full advantage of the evidence they collect:

Norms of a Research Community

- **The focus of the research is the system** that produces current results. Individual teachers are not the focus of improvement, but rather parts of a larger system made up of interactions among people, tools and materials, and processes that bind the work together.
- **Trust is crucial** to working together as a research community. All discussions surrounding data collection and analysis must be low-stakes and non-threatening or members may not feel comfortable sharing their struggles and the change idea's failures. Improvement requires honest reflection on why failures occur to find areas of improvement, and without trust, these honest conversations cannot take place.
- **Transparency and sharing** of successes, challenges and setbacks are essential ingredients for iterative improvement to take place. As educators engage in cycles of learning, they may uncover challenges, small flaws, and potential areas for improvements in many of the school's work processes. Open sharing of these discoveries will create insights and shifts in the educators' thinking about the problem, which will enable improved ways of doing the work, and renewed efforts to measure the new improvements. The willingness to be transparent and share enables this feedback to drive continuous improvement.

Deciding What "Good" Looks Like

The first step to practical measurement is to settle on a common definition of what "good" looks like for each part of the Driver Diagram[1]. For instance, in the attached example diagram (page 7), it may not be immediately obvious how to define "good" for the Driver: "Students persist through difficulty," since it is necessary to decide how much persistence is required as well as how much difficulty is appropriate. If, however, the learning cycles team can agree on a definition of quality, or what "good" looks like for their students and the work they are doing, then measuring it becomes possible. Many difficulties in measurement arise where a common definition of quality is absent, so it is important to agree upon these definitions early in the process.

Using your Driver Diagram developed during the high-leverage change activity, work with your learning cycles team to come up with definitions of "good" for the Change Idea you have chosen. Use the following questions to guide your discussion:

1. What does "good" look like for our aim?
2. How can we recognize "good" when we see it? What are the look fors we can use?
3. How can "good" be measured?

Repeat the process for the rest of the key parts of the Driver Diagram (i.e. those that connect the Change Idea to the Aim). Starting from the Change Idea, move left across the Driver Diagram to the Opportunity that it connects to, and continue the process for the Driver and finally the Aim. Note, that it is possible for a Change Idea to impact more than one Opportunity, as it is possible for an Opportunity to impact more than one Driver. If this happens, repeat the process for all of these connected parts as well. See the attached example diagram (Page 8) to see what this **"reduced" Driver Diagram** can look like.

Before moving on, it is necesarry to prepare for **potential unintended consequences** of the change idea. As all educators understand, school improvement efforts are not free. They cost time and energy that could be dedicated to other priorities. In other cases, there are inherent tensions in the choices educators make, such as the trade-off between increasing rigor and maintaining access and entry points for all students. In order to be mindful of the costs of carrying out a change, and to allow for later analysis and balancing of priorities, these costs should be measured as well. Spend a moment thinking of one or two key unintended consequences that may arise from the change idea.

Brainstorming Measures

Once your team has agreed on common definitions for each of the key parts of your Driver Diagram you are ready to begin brainstorming measures. The goal of the brainstorming is to think through the question:

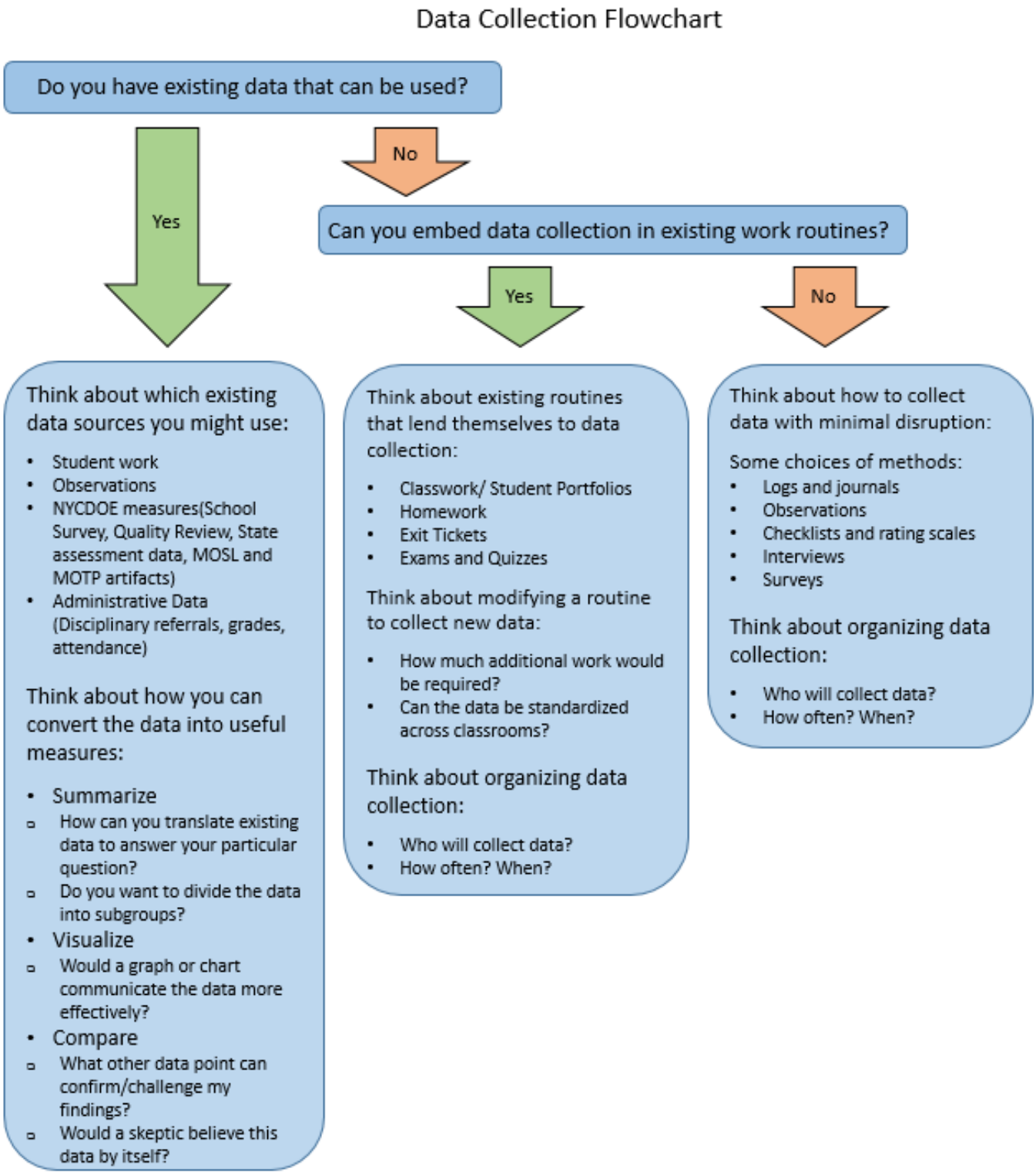
How will we know if the intended change is happening?

The first step in doing so is to list possible methods for data collection for each of the key parts of your Driver Diagram. Use the definition of what "good" looks like and answer the above question. Remember that there are often multiple ways to measure something, and choosing the best measure or combination of measures for your purposes should start with exploring multiple options.

To help with this brainstorming, the data collection flowchart on the following page provides a sequence of guiding questions. For each key part of the Driver Diagram (Aim, Driver, Opportunity and Change Idea), follow the flowchart and use the questions to think through potential measures and their merits. Continue until you

[1] For support in creating a Driver Diagram, see the Planning a High-Leverage Change with a Driver Diagram activity

have at least two or three potential measures listed for each part of the reduced Driver Diagram, including the unintended consequences.



Choosing your Measures

The next step is to choose which measures from the brainstorming to use. It may be that in the process of answering the questions in the flowchart above, you have developed a clear preference, or you may still have several potential measures to choose from. In deciding among different measures there is no one criterion that that takes priority over others, although it is important to remember that practical considerations are just as important as any other and schools should not feel shy about choosing a measure for practical reasons. After all, an unsustainable workload will undermine an improvement effort just as readily as incomplete data. The following guiding questions should help the team settle on the appropriate measures for their particular school context and aim:

- Is this measure sensitive enough to capture small changes?
- Will this measure capture the full variation of the thing it measures (i.e. variation among students and/or variation over time)?
- Does it produce data on where and for whom the impact is greatest/weakest?
- Is your measure framed in a language that is meaningful to teachers (and students)?
- How much additional work is required to collect and process this data? Who would be responsible for that extra work?
- Does your measure produce data in a timely manner?
- Is your measure formative? Can it provide insight into why your change is or isn't working as planned?
- Are your measures non-threatening and transparent?

Remember, the goal of this exercise is not to arrive at a set of perfect measures that will definitively answer the learning cycles team's questions, but rather to find a set of measures that, taken together, are good enough to provide feedback about how each part of the theory of change is working. For this reason, limitations or doubts that exist about any particular measure can be dealt with through **triangulation**, or the collection of complementary data that can paint a more complete picture. Ideally, triangulation should make use of different perspectives. For example, teacher logs on student behavior can be complemented with a student survey that asks about their perception of the change idea and its impact. In this case, the quick and frequent teacher logs reveal change over time while the student surveys offer a deeper look into variation among students. Taken together in this way, multiple sources of data offer much more actionable findings than any single measure.

The following guiding question can help your team review the measures you have identified and discuss the potential for triangulation:

- Is the kind of data that this measure produces convincing on its own? Would it convince a skeptic?
- Does it have any blind spots? If so, what measures could compensate?
- Does this measure capture multiple perspectives and points in time? Is there an easy way to capture more?

If settling on a final set of measures is difficult, remember that like the rest of the cycle of learning, these can be revised later as your learning cycles team's increased understanding of the change leads to more questions.

Predicting Change

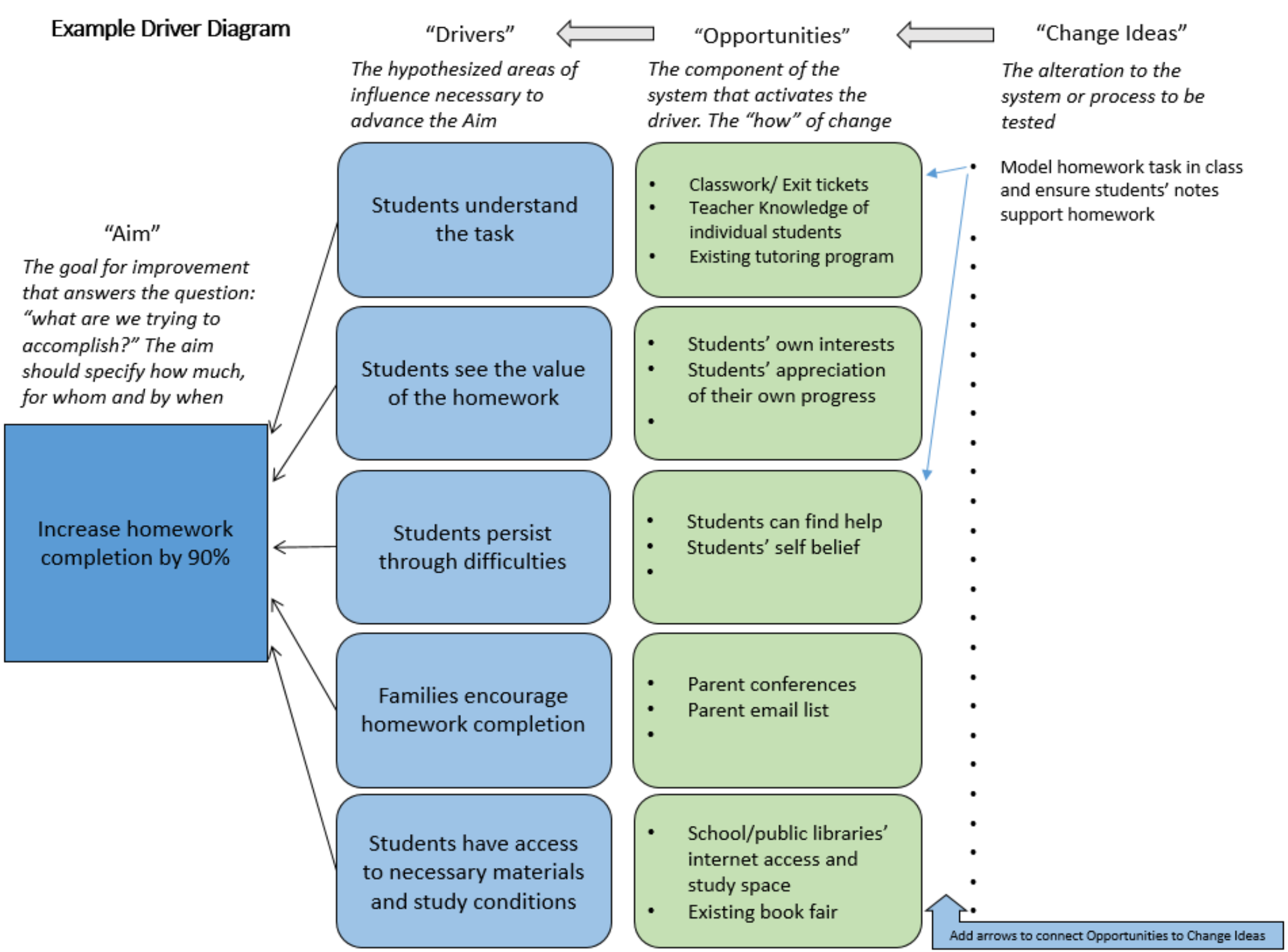
The final step in your preparations for measurement is prediction. This step creates a benchmark that allows the team to learn from predictions that don't play out exactly as planned. Unlike external forms of accountability however, the team's internal accountability has the sole purpose of turning a failed prediction into a question that can inform the team's next cycle or change idea. These failed predictions are just as valuable as successes

because they uncover assumptions and biases we all have about the system we are trying to fix and the solutions that are needed.

For this reason, making predictions is a necessary step in the process, and will allow the team to compare their thinking at the beginning of the cycle with their thinking at the end.

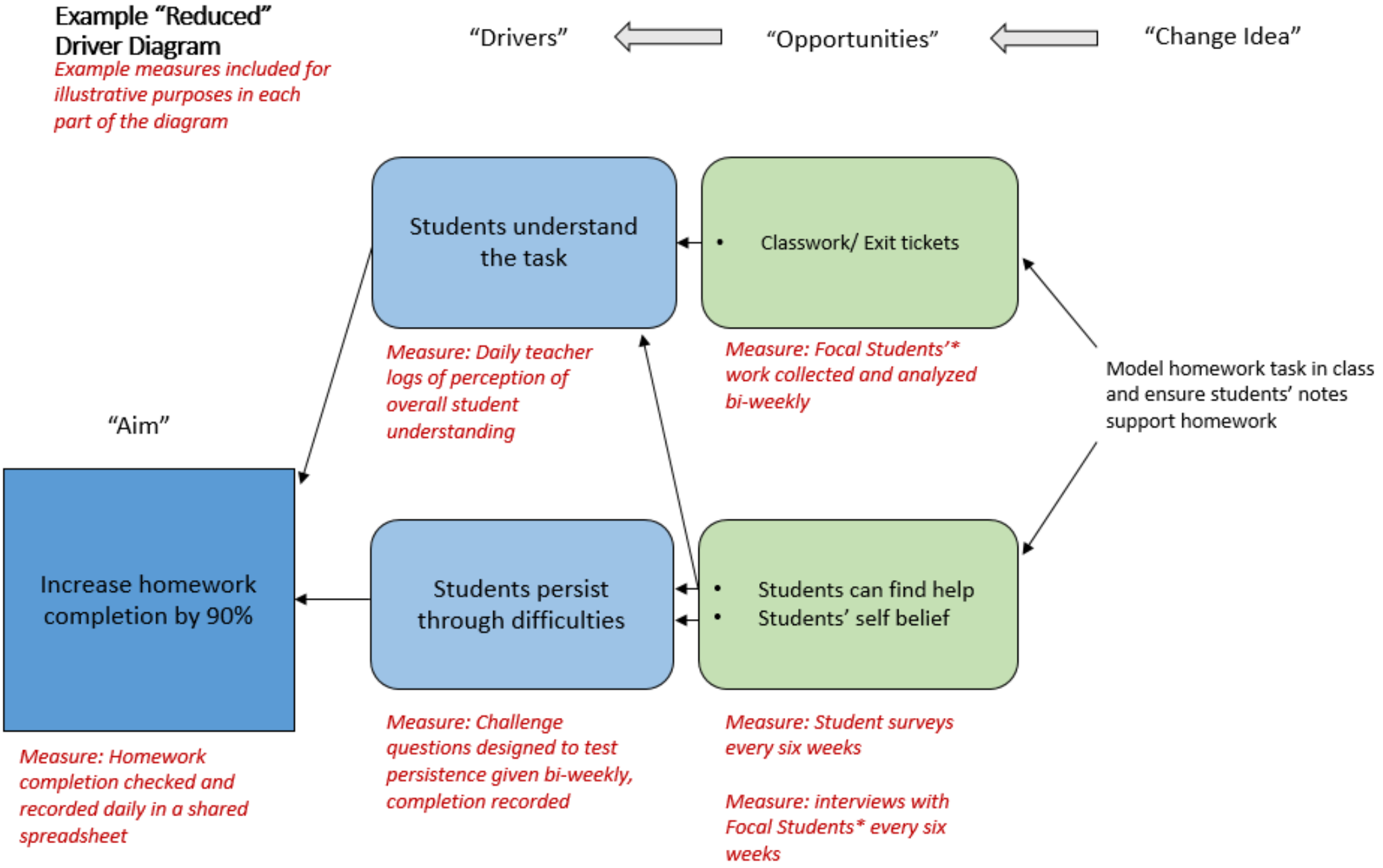
Working List of Measures

Type of measure	What will you measure?	When and how often will you measure?	What change do you predict you will see in this measure?
Aim (summative):	How will you know when you have reached your aim?		
Aim (predictive):	What are early indicators that will let you know if you are getting closer to your aim?		
Driver:	What are the areas that will make a difference to that aim?		
Opportunity:	How will you know if there is improvement in these areas?		
Change Idea:	How will you know if the Change Idea is being implemented as planned?		
Unintended Consequence:	What are the costs or potential negative consequences of the change?		


Example Driver Diagram
"Drivers"
"Opportunities"
"Change Ideas"
The hypothesized areas of influence necessary to advance the Aim
The component of the system that activates the driver. The "how" of change
The alteration to the system or process to be tested
"Aim"
The goal for improvement that answers the question: "what are we trying to accomplish?" The aim should specify how much, for whom and by when
Increase homework completion by 90%
Students understand the task
Students see the value of the homework
Students persist through difficulties
Families encourage homework completion
Students have access to necessary materials and study conditions
Classwork/ Exit tickets
Teacher Knowledge of individual students
Existing tutoring program
Students' own interests
Students' appreciation of their own progress
Students can find help
Students' self belief
Parent conferences
Parent email list
School/public libraries' internet access and study space
Existing book fair
Model homework task in class and ensure students' notes support homework
Add arrows to connect Opportunities to Change Ideas

Example "Reduced" Driver Diagram

Example measures included for illustrative purposes in each part of the diagram



*For more on the Focal Student strategy for data collection, see the Guide to Practical Measures, or Data Wise in Action (Boudett & Steele, 2007), pages 78-80, 84.