Supporting Math MTSS Through SpringMath

Your guide to driving math mastery

Amanda M. VanDerHeyden, Ph.D.
Paul Muyskens, Ph.D.

SpringMath by Sourcewell

Table of Contents

CHAPTER 1

Why Math MTSS? Why SpringMath?

Education is by now more than 15 years into MTSS implementation at scale as a system improvement, eligibility determination, and instructional service delivery model. Yet most implementation efforts are heavily focused on reading, and results from many systems are disappointing.

Mathematics is a critical life skill that either opens doors or becomes a barrier to a student's future life and economic opportunities. Math learning deficits appear very early in the K-12 experience, and trajectories are unimproved without intervention (Lee, 2012). Few U.S. children attain the minimal proficiencies required to position them to enroll in or complete college education. Opportunity gaps are broad and unimproved over the last two decades, with students of color scoring lower than their white peers on standardized assessments such as the National Assessment of Educational Progress (National Center for Education Statistics, 2020).

Meanwhile, assessment and intervention research in mathematics has become increasingly sophisticated and well-funded, and has led to a body of evidence that, if attended to and followed, could produce appreciable, important, and meaningful student learning gains right away. Yet, for all the national panels, position statements, and attention paid to math instruction, math learning has remained largely stagnant over the last two decades. Students deserve better. Given that the problem of poor math achievement in the U.S. has occasioned so much attention and debate nationally, why are the results so dismally unimproved?

Math debates in the U.S. have been dominated by a heavily philosophy-driven narrative in education that promotes specific teaching tactics that are misaligned with science and evidence. The use of these tactics, at best, competes with the time and resources needed to implement more effective tactics. At worst, such tactics are actually harmful to student learning. If education were like medicine, there would be a standard of care that would guarantee that all students have access to instruction that meets certain standards with regard to efficacy. Alas, education is not like medicine, and the quality of instruction that students experience is widely variable and often worse in communities with fewer resources.

MTSS is uniquely positioned to improve "the math problem" because, when done well, it constantly holds the implemented actions to the standard of improved student learning. Thus, if a teacher wishes to use a tactic in his or her classroom that is not an evidence-based tactic, implementers can approach the implementation with curiosity but also commitment to student learning and adjust the tactic until it produces student learning gains, using the mechanisms of MTSS. In other words, the problem-solving structure of MTSS is ideal for pruning tactics in the school to select and support those that are producing the greatest gains in student learning.

Among school improvement frameworks, MTSS is likely the best choice, but it is also probably not the choice that requires the least effort. Implementing MTSS takes consistent, steady problem-solving, collecting, interpreting, and acting upon student learning data often in highly political contexts with teachers who bring their own unique philosophies, biases, and skill sets to the endeavor. It is easier to buy a new curriculum, and often such purchases make for better press than do the less flashy efforts such as increasing student engagement during math lessons by increasing the dosage of opportunities to respond. MTSS is humbling and difficult work. I have a colleague who says, “MTSS is the worst way to diagnose a learning disability, except every other way.” (Burns, 2016). There is no shortcut, but the upside is that you will raise achievement if you are willing to do the work.

1.1 Changing organizational behavior is a human enterprise

Teaching is a human enterprise. Certainly, remote learning has demonstrated that any instruction that does not begin with authentic student connection is likely to be fruitless. System change is similar in that teachers must believe that change is needed and change is possible. Teachers must trust that leaders are telling them the truth about the rationale for the change. They must understand clearly what specific actions will be expected of them in the initial stages of implementation and how the effects of the implementation will be supported and evaluated. Teachers must believe that the leaders of the effort have the sufficient skills and materials to successfully implement. Ongoing implementation support requires that teachers be provided with the results of their efforts as they implement as well as information about how to improve their implementation (VanDerHeyden, 2018).

It is not uncommon for teachers and instructional leaders to bring philosophies to math instruction that may be at odds with evidence, and it can take some time to help teachers understand how new or different practices can be incorporated into their teaching to help students learn better. In many cases, leaders will not be able to change the minds of teachers about their closely held philosophies, but what they can do is help teachers buy into a shared mission of making learning success more available to students. When the allegiance is to the learner, rather than the tactic, system change becomes possible. So, implementing math MTSS often requires ongoing discussions to place the learner at the heart of the effort. Asa Hilliard (1991) said,

> “The risk for our children in school is not a risk associated with their intelligence. Our failures have nothing to do with IQ, nothing to do with race, nothing to do with language, nothing to do with style, nothing to do with the development of unique and differentiated special pedagogies, nothing to do with the children’s families. All of these are red herrings. The study of them may ultimately lead to some greater insight into the instructional process, but at present, they serve to distract attention from the fundamental problem facing us today. We have one and only one problem: Do we truly will to see each and every child in this nation develop to the peak of his or her capacities?” (p. 36).

Cultivating the will for change is not a “one and done” effort that you complete before you begin the system change process. Rather, will is continually cultivated and rewarded through student learning gains using the processes of behavioral shaping at the system level (VanDerHeyden, 2018).

1.2 Being an effective steward of instructional resources means following science

The most predictable outcome of effective instruction is learning (Keller, 1968). This outcome is so predictable, in fact, that when learning is not occurring it means that the instruction is not effective. When learning is not occurring, many teachers and leaders are far too quick to ask, “What is wrong with the learner?” when what they should be asking is, “What is wrong with the instruction?”

Students have a right to reliable and stable access to highly effective instruction (Barrett et al., 1991). The instruction that they receive should be based upon what can be expected to work. Teachers, as stewards of instructional opportunities for students, are not simply free to use whatever instructional tactic or approach they happen to like or have the greatest familiarity with.

Is there space for teacher preferences or instructional activities that seem promising but have limited evidence to date? Of course there should be room in any profession to experiment with new ideas, but it is ethically unacceptable to consign students to unconsented experimentation as their primary method of instruction. The stakes are too high. Math is a highly cumulative endeavor. Stokke (2015) describes math as "relentlessly hierarchical." When instruction is not effective, the resulting lack of skill mastery will be magnified in future years because future learning depends upon early skill mastery. On the balance, instructional actions should be heavily weighted toward tactics of demonstrated efficacy. Experimental tactics could be tried so long as students are otherwise thriving and so long as the "new" instruction is rigorously evaluated during implementation. In fact, all systems should thoroughly evaluate their assessment and intervention programs each year (Morrison & Harms, 2018).

Several red flags can signify to implementation leaders that philosophies are prevailing over evidence in math instruction. For example, when educators believe that instruction and learning are too complicated to study, that belief is at odds with science (Carnine, 2000). When educators believe that novices learn like experts or that explicit instruction or direct instruction is harmful and at odds with creativity, such beliefs reflect profound misunderstanding of how learning occurs (Kirschner, et al., 2006). Explicit instruction is not standing in front of students and telling them the answers and how to get them. Rather, explicit instruction enables less effortful and error-prone learning that involves a high number of interactions among the teacher, the tasks, and the learner. Carefully selected and sequenced tasks that leave little guesswork in the early stages of learning help all students make the correct discriminations with the least frustration possible in the shortest period of time. Explicit instruction logically would diminish anxiety for children with weaker skills and has one of the strongest effect sizes on student achievement (Jitendra, et al., 2018).

Building fluency of responding actually creates generative, flexible, and creative skill use (Johnson & Layng, 1992). Tactics that provide opportunities to solve untaught problem types might be introduced as a generalization opportunity but are not useful or effective as an acquisition activity. Instruction is a science, and given task demands, student skill proficiencies, and goals, certain instructional actions will be effective and others will not. These actions are not mysterious. They are predictable, measurable, and teachable. The way in which a teacher connects with students is another matter. We have all known those teachers who have that special ability to connect with students. When a teacher who has the gift of knowing how to connect with students also understands the science of instruction and learning, that is the secret sauce of education. That is the teaching that we want for all our children. It is vitally important that teachers have the know-how and desire to authentically connect with students and have the skill set to engineer successful learning via the science of instruction and learning.

When educators profess a preference for tactics such as productive struggle, avoiding the teaching of algorithms, and avoiding fluency-building, those philosophies are out of sync with the research evidence and policy recommendations in mathematics (Fuchs et al., 2021; National Mathematics Advisory Panel, 2008; National Research Council, 2001) and can be expected to weaken, rather than benefit, mathematical learning. When educators understand the science of how children learn, they can understand how easy and fun it can be for students when effective learning tactics are used in classrooms. Such discussions must move beyond opinion and philosophy, however. Our joint aim is to situate specific instructional techniques for learners who can benefit from those techniques according to the Instructional Hierarchy.

1.3 SpringMath

Crafting a tool that is used in schools is a tremendous social responsibility. SpringMath has a responsibility to make adjustments to the platform in response to evolving science, to use efficacy data to drive the adjustments, to continually refine the platform itself, and to support its use in schools. SpringMath has a responsibility to limit and mitigate potential misuses. For example, when systems use SpringMath incorrectly, there should be an infrastructure to detect such errors and advise and support systems to adjust. SpringMath must remain ever vigilant to providing the platform itself in ways that promote best practices. Thus, SpringMath is built to reflect growth in sensitive ways, to build strong implementation, and to encourage effective practices. For example, we do not offer the assessments separately from the interventions because we believe that assessment without intervention is an unhelpful (albeit prevalent) practice. We are often asked by systems how they might use the tool in ways that we do not have sufficient evidence to anticipate that it would work. In those cases, we are direct about that reality. We always bear in mind that there is no right way to do the wrong thing and so we help systems understand the limits of what is possible given their specific contextual constraints, help them begin implementation, and help them optimize their implementation over time.

SpringMath is about empowering and enabling teachers to teach math more effectively. We look forward to teaching you about SpringMath so that you can create more successful math learning for your students. There is no more important work. Let’s begin.

References

Barrett, B. H., Beck, R., Binder, C., Cook, D. A., Engelmann, S., Greer, R. D., Kyrklund, S. J., Johnson, K. R., Maloney, M., McCorkle, N., Vargas, J. S., & Watkins, C. L. (1991). The right to effective education. The Behavior Analyst, 14(1), 79-82. doi.org/10.1007/BF03392556

Burns, M. K. (Chair), Stevenson, N., Miciak, J. & Maki, K. E. (2016, Feb). Identifying specific learning disabilities: What do we know and where do we go from here? Symposium presented at the Pacific Coast Research Conference. San Diego, CA.

Carnine, D. (2000). Why education experts resist effective practices (and what it would take to make education more like medicine). Washington, D.C.: Thomas B. Fordham Foundation.

Fuchs, L.S., Newman-Gonchar, R., Schumacher, R., Dougherty, B., Bucka, N., Karp, K.S., Woodward,

J., Clarke, B., Jordan, N. C., Gersten, R., Jayanthi, M., Keating, B., and Morgan, S. (2021). Assisting Students Struggling with Mathematics: Intervention in the Elementary Grades (WWC 2021006).

Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of

Education Sciences, U.S. Department of Education. Retrieved from whatworks.ed.gov/.

Hilliard, A. (1991). Do we have the will to educate all children? Educational Leadership, 49, 31-36.

Jitendra, A. K., Lein, A. E., Im, S.-h., Alghamdi, A. A., Hefte, S. B., & Mouanoutoua, J. (2018). Mathematical interventions for secondary students with learning disabilities and mathematics difficulties: A meta-analysis. Exceptional children, 84, 177-196.

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. American Psychologist, 47, 1475–1490.

Keller F. S. (1968). "Good-bye, teacher...". Journal of applied behavior analysis, 1(1), 79–89. doi.org/10.1901/jaba.1968.1-79

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75-86. doi.org/10.1207/s15326985ep4102_1

Lee, J. (2012). College for all: Gaps between desirable and actual P–12 math achievement trajectories for college readiness. Educational Researcher, 41(2), 43–55.

Morrison, J. Q., & Harms, A. L. (2018). Advancing evidence-based practice through program evaluation: A practical guide for school-based professionals. Oxford University Press.

National Center for Educational Statistics. (2020). National assessment of educational progress 2019 results for mathematics and reading. [NAEP Data Explorer]. Institute of Education Sciences. Retrieved from nationsreportcard.gov/ndecore/xplore/NDE

National Mathematics Advisory Panel (2008). Foundations for Success: The Final Report of the National Mathematics Advisory Panel, U.S. Department of Education: Washington, DC.

National Research Council. (2001). Adding it up: Helping children learn mathematics. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. Review of Educational Research, 78, 427–515. doi/10.3102/0034654308317473

Stokke, A. (2015). What to do about Canada's declining math scores? C.D. Howe Institute Commentary 427, Available at SSRN: ssrn.com/abstract_id=2613146 or dx.doi.org/10.2139/ssrn.2613146

VanDerHeyden, A. M. (2018). Why do school psychologists cling to ineffective practices? Let's do what works. School Psychology Forum: Research in Practice, 12, 1-9.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. A (2007). Multi-Year Evaluation of the Effects of a Response to Intervention (RTI) Model on Identification of Children for Special Education. Journal of School Psychology, 45, 225-256. doi.org/10.1016/j.jsp.2006.11.004

CHAPTER 2

Multi-Tiered Systems of Support (MTSS) in Mathematics

MTSS in mathematics is a powerful way to improve math proficiency in schools and districts. Not all MTSS models are the same. Rather, specific assessments, intervention procedures, and decision rules define a unique MTSS. Yet, sensitive, ongoing measurement of student learning is the fuel for decision making in any MTSS model. The process of MTSS has been referred to as "data-based decision making" and "problem solving" because that is its core purpose: to use student learning data to drive instructional actions to improve learning at the system and student level.

MTSS systems require that decision makers examine data sources to make the following decisions in a cycle that repeats iteratively until the identified problem is resolved.

Problem definition | **Problem solution** | **Solution implementation** | **Solution evaluation**

The structures of MTSS are inseparable from most school improvement frameworks, and MTSS has become a cornerstone of many professional practice and service delivery models related to general and special education (e.g., Ysseldyke, et al., 2006). When general education organizations, such as the National Council for Teachers of Mathematics, refer to formative assessment, differentiated instruction, and supplemental and targeted supports for learners who struggle in mathematics, they are, in fact, referring to the processes of MTSS.

2.1 MTSS

The MTSS process is illustrated in Figure 2.1. The core features of an MTSS model are the availability of increasingly intensive instruction (Tiers 1, 2, and 3) provided to all (Tier 1) and subsets (Tiers 2 and 3) of students using sensitive screening and progress-monitoring data for decision making.

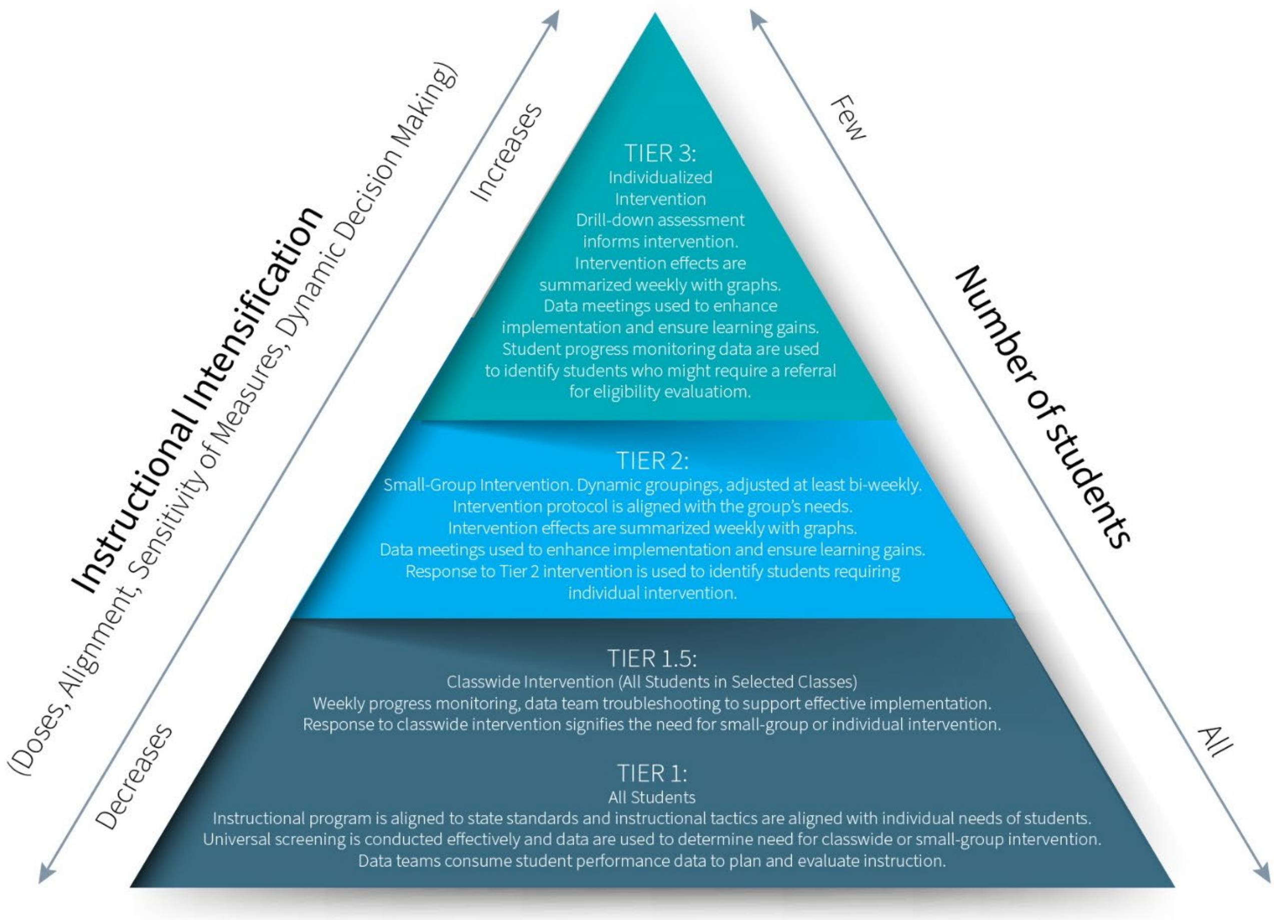


Figure 2.1. The Tiers of MTSS. Reprinted with permission from Kovaleski, VanDerHeyden, Runge, Zirkel, & Shapiro (2022). The RTI Approach to Evaluating Learning Disabilities. NY: Guilford.

Tier 1

MTSS begins with universal screening at Tier 1. The purpose of universal screening is to identify classwide and individual student risk, and evaluate program improvements over time. Universal screening establishes a local, normative comparison group that can enable more accurate decisions about risk. When simple, norm-referenced cut scores are applied from an external vendor, decisions about whether and which students require intensified instruction in any given classroom are highly unstable and error prone. The rationale for norm-referenced identification of risk has been that schools have limited resources with which to help students and so the percent of students who can be accommodated for supplemental instruction is the percent selected from a static screening. What this process neglects is that when large numbers of students are experiencing risk, selecting some lowest-performing is unstable and error prone. Also, schools should aspire to build structures with existing resources to help all children who need help to attain important academic milestones and benchmarks. That position is fundamental to equity. Thus, to permit the best decisions about risk, rigorous, core-aligned measures and benchmark criteria are superior to norm-referenced selection of the lowest-performing students. Benchmark criteria are performance scores that forecast successful learning outcomes such as skill retention, faster learning of related content, generalization, and grade-level learning as reflected on year-end state tests.

Systems must exercise some savvy in making a decision about their level of risk. Most commercial math screening tools are not sufficiently sensitive to make this determination, and actually many tools include below-grade-level skills on the screening measures in order to "force" a more normal distribution of scores from which to identify some subset for intervention (e.g., the lowest performing 20%). This tail-wags-the-dog logic means that a student who scores in the risk range is definitely in trouble, but it also means that a student who passes such a screening may be unlikely to experience longer term math learning success (or stated another way, avoid math failure).

Savvy implementers can look at system functioning as a detective might, asking, "What percent of our students meet or exceed the proficiency criterion on the year-end test at each grade level? Is our state proficiency expectation rigorous and well-aligned with college readiness data (nces.ed.gov/nationsreportcard/studies/statemapping/)? What proportion of our students enroll and successfully complete advanced math course sequences? What proportion of our students meet or exceed the ACT college readiness benchmark in math? What do these proficiencies look like for subgroups we might define (in other words, do we see that success differs by student grouping)? Can we see performance differences in our students traceable to the prior grade level, prior school, or even specific teachers so that we might add some support for better results in those environments?"

One common MTSS pitfall in math is to believe that "our students are doing okay in math" when only 50% of students meet the year-end proficiency criterion and even fewer meet the ACT college readiness benchmarks in math. This fallacy is prevalent because many other systems around one's own system may perform similarly (after all, NAEP scores in the U.S. are abysmal) and many screening measures are opaque in terms of their content, too easy, and detect learning problems too late in the student's grade progression. Parents are often puzzled because "my student was doing fine and making A's in third grade" but then everything falls apart for the student in fifth grade. It seems to rarely occur to parents and teachers that perhaps their math-risk-detection systems failed. Another common pitfall is to conclude that the math struggles begin in the grade level where they are first detected, which, again, is usually an artifact of assessments and/or assessment procedures. Another pitfall is to assume that opportunity gaps are not present in the absence of specific examination of your data to answer this question. Subgroups of students should thrive at similar rates (or experience similar risk) and if they do not, then you have an opportunity gap. In fact, excellent program evaluation harvests MTSS data to guide program improvements (Morrison & Harms, 2018), but systems rarely seem to examine their data in this way. If you want all students to thrive in more advanced coursework in high school math, you must begin that work in grades 1-5 (Koon & Davis, 2019).

Tier 1.5

The next Tier, Tier 1.5, is classwide intervention, which becomes the second gate of screening when there is a large amount of risk in the class, grade, or school (as there often is in math). If the class median score during screening was in the frustrational range, indicating that most students in the class were likely to struggle with grade-level instruction, we recommend classwide intervention. The purpose of classwide math intervention is to (1) rapidly improve math learning for all students in the classroom in a highly efficient way, and (2) use student learning data to identify students who do not experience sufficient growth given classwide math intervention in the same environment and with the same integrity or opportunity as their same-class peers. Classwide intervention may not involve every classroom in a school or grade, but it always involves all students in a given class when that class meets the risk criterion. The technical and necessary purpose of including all students in classwide intervention is that when the risk base rate is high, the screening cannot function to accurately determine who is really at risk in that classroom, no matter how accurate the screening is in terms of sensitivity and specificity (VanDerHeyden, 2010; VanDerHeyden, 2013).

The technical reasons for the need for classwide intervention have been mathematically proofed and explained elsewhere (VanDerHeyden, 2013), but a practical example might be helpful here. In the example in Figure 2.2, the screening data for a class needing classwide math intervention are shown. All but two students in this class are scoring in the risk range on the screening.

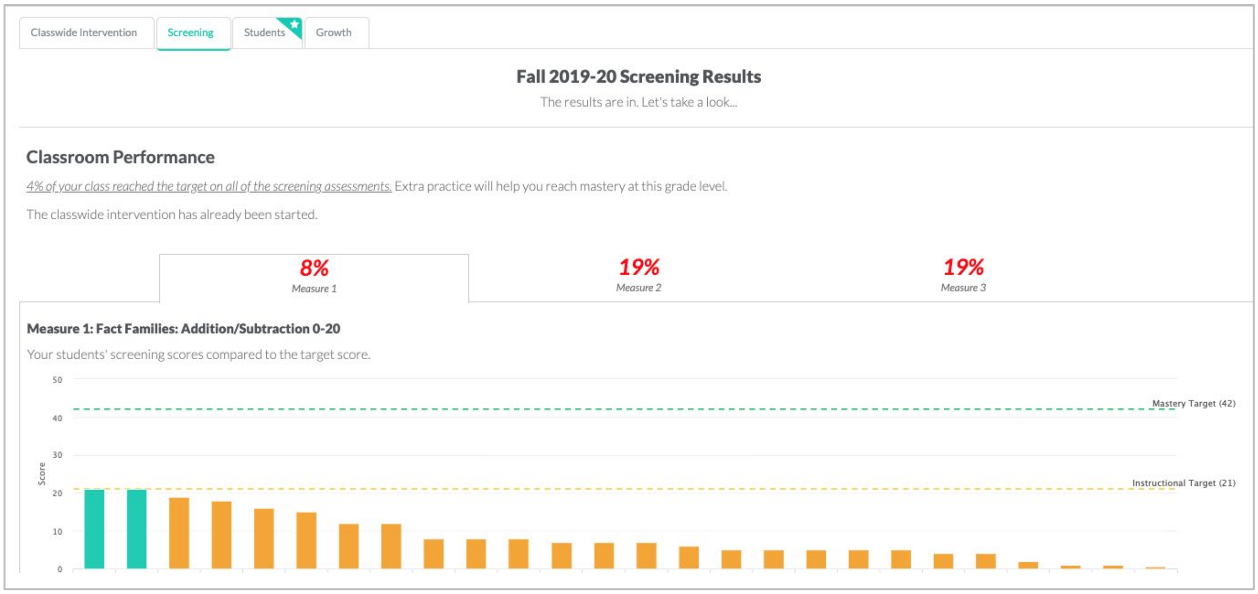


Figure 2.2. At fall screening, all but two students in this class have scored in the risk range. As a result, this class is recommended for classwide math intervention in SpringMath.

The class begins classwide intervention, and after four weeks, the median score for the class (heavier darker red line) reaches mastery (the green dashed horizontal line) as shown in Figure 2.3.

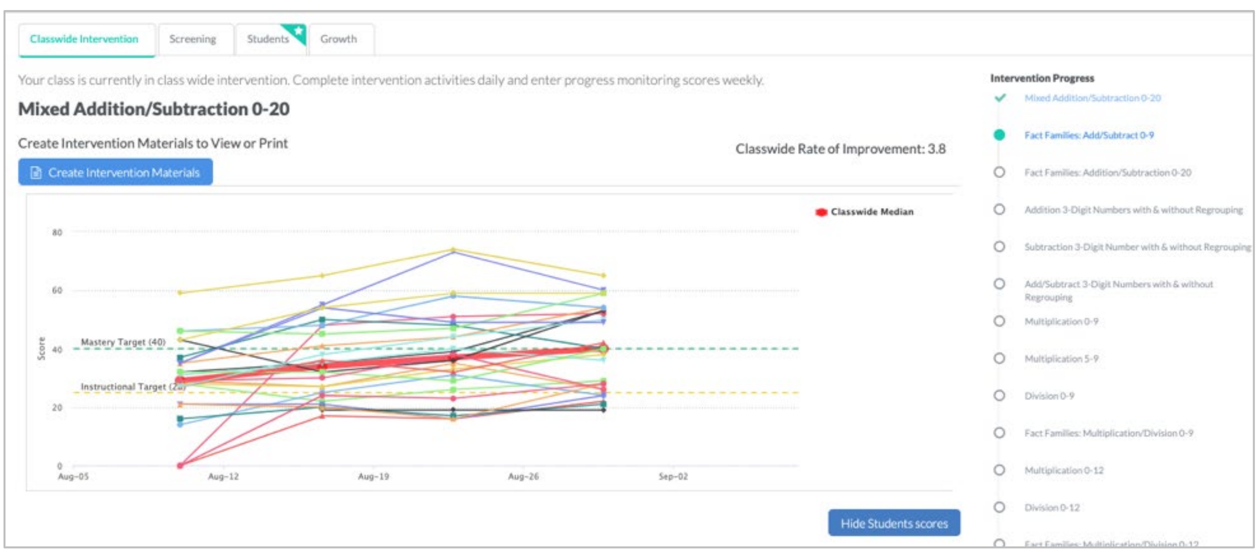


Figure 2.3. Classwide intervention begins and produces strong upward growth for nearly all students over four weeks. At week four, the class median has reached mastery and the class will move on to the next skill in the intervention skill sequence. SpringMath will now recommend students for individual intervention based on response to classwide intervention.

SpringMath identified the four students shown in Figure 2.4 as needing individual intervention. These are likely not the four students the teacher would have identified if just looking at static screening data. Let's imagine choosing the lowest 20% of students as many systems do, which would be the lowest-scoring five students in this example of a class of 26 students. In this case, the system would correctly select two of the four students who certainly needed intervention, or a 50% detection rate. In other words, two of the four students who needed intervention would have been missed. This mistake would not be uncostly due to unnecessary intervention to three of five (60%) students in the intervention group. This problem looms large in many MTSS models, bloating intervention groups while also failing to give intervention to students who need it. These data (and the mathematical proofing regarding high base rates of risk and academic screening) are the reason that classwide intervention is essential as a second screening gate in classes with high base rates of risk.

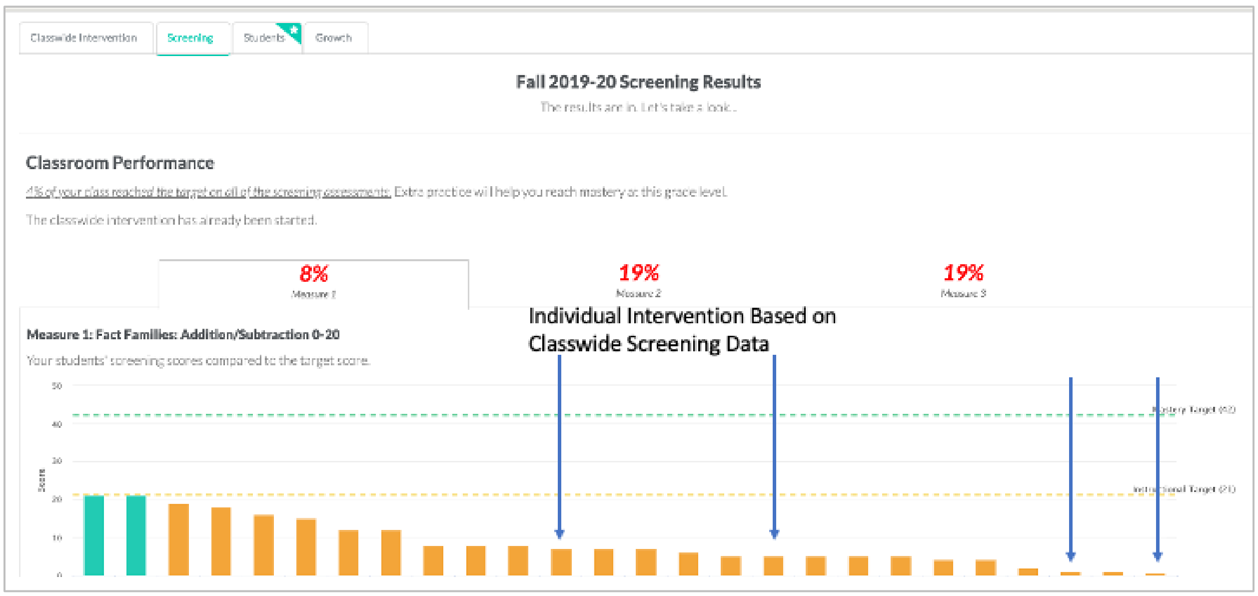


Figure 2.4. The four students identified with arrows are the students recommended for individual intervention based on their response to classwide math intervention.

As the second gate of screening when there is a lot of risk, classwide intervention has the added technical capability of lowering the base rate of risk (which results in more accurate detection of individual students at risk), making use of trend data in the risk decision (as opposed to a single static screening score), and having the benefit of more normalized distributions of scores when the risk decision rule is applied. These technical benefits improve the accuracy of academic screening. As a result, classwide intervention is now formally recognized as a necessary feature — Tier 1.5 — of MTSS (Kovaleski, et al., 2022).

Tier 2

Tier 2 is supplemental intervention, which in many MTSS models is defined as small-group supplementation of core instruction. By definition, Tier 2 is not a replacement for core instruction and rather provides supplemental instruction to those students in need of intensified instruction to attain grade-level proficiencies.

Original MTSS logic suggested that perhaps 20% of students may need supplementation (Batsche, et al., 2005), but that number was always conceptual and somewhat arbitrary. The logic of 20% reflected the resource constraints within which most systems could be expected to deliver supplemental intervention and placed the onus back on the system to improve core instruction where greater percentages of students were failing to attain expected grade-level proficiencies. Resource-driven intensification models in some ways can be highly practical, but they run the risk of bringing underwhelming results at the system level with sustained and/or widened opportunity gaps. This problem of resource constraints is one reason that classwide intervention is a powerful addition to MTSS and is now officially incorporated into MTSS (Kovaleski, et al., 2022). Whereas it is true that providing supplementation to more than 20% of students may not be manageable, it is equally true that many children scoring between the 20th and 50th percentiles locally may still need instructional intensification. It is especially likely that if a nonlocal normative criterion is used in a context where many children struggle academically, many children would be deprived of the intervention they need to avoid academic failure. Yet another benefit of classwide intervention is that it can be a model of effective universal intensification to core instruction for teachers and it provides a data source that can show systemic learning gains, which can be used to encourage continued implementation. In other words, classwide intervention can nudge improvements to core instruction, which can make Tier 2 and 3 interventions more feasible and effective.

Tier 2 intervention is often delivered in small groups. One exciting recent finding about MTSS is great news for implementers. Specifically, research has shown that the number of students receiving intervention in a group is not highly related to intervention intensification. Specifically, intervention delivered in groups of two students to one adult produced comparable effects to intervention delivered in groups of five students to one adult (Clarke, et al., 2017; Doabler, et al., 2018). Thus, the small-group format is not the key intensification variable. Rather, the key intensification variables are more explicit instruction delivered with added opportunities to respond and corrective feedback for more narrowly defined targets, including students who require the same intervention. This understanding of intensification variables is now recognized by the National Center for Intensive Intervention (NCII). NCII now invites vendors of intervention tools on their Tools Charts to submit evidence of these “ingredients” of intensification (e.g., dosage of opportunities to respond, features of explicit instruction).

Tier 3

Tier 3 intervention is the most intensive and individualized intervention that can be delivered within the school. This is a good time to remind readers that all children should receive the level of instructional intensification that they need to reach grade-level performance, and this includes children who may already be identified and receiving special education via an individual education plan, or IEP. Tier 3 intervention, by definition, is provided to students who failed to demonstrate expected response to intervention (RTI) during Tier 2 and, therefore, is provided to fewer students than received Tier 2. Tier 3 interventions are individualized, which means that they are developed based on individual, diagnostic assessment results for the individual student. Because grouping students does not reduce intensification, Tier 3 interventions can be delivered in small groups, even though they were not conceived this way in original MTSS models. The key to grouping students in Tier 3 is to ensure that each student is getting the customized intervention according to his or her diagnostic assessment, grade-level goals, and ongoing progress-monitoring data.

2.2 SpringMath and MTSS

In this section, we will explain how SpringMath is a comprehensive MTSS model and articulate how each ingredient acts in tandem with the other ingredients to drive learning improvements for all students. If

SpringMath were like baking a cake, the baker could not choose which ingredients to use and which to exclude. The ingredients are precisely specified to result in an edible cake. If the baker alters the ratio of liquids to dry ingredients, for example, the cake may be ruined. If the baker uses the prescribed ingredients in the correct amounts but changes the baking time or temperature, the cake may be ruined. Expert bakers know which variables they can alter and how to alter them to attain the desired end result, and they can do this because they understand the science underlying every step of the baking process. MTSS systems are not unlike cake baking. Expert implementers may be able to substitute ingredients (with appropriate adjustments throughout to accommodate the substitution), but most implementers should proceed with caution and follow the recommended implementation path. One pitfall of implementation in MTSS is to use some but not all of the MTSS steps and then wonder why student learning was unimproved. For example, many systems collect a lot of screening data but do not successfully interpret and act on those data to improve programs of instruction and deliver intensive interventions to the right students.

We have said that not all MTSS models are equivalent or equally effective. Specific MTSS models, like SpringMath, must specify the details of the ingredients in each layer of the MTSS process. They must operationalize each step and evaluate individual layers but also the overall effect of the integrated model when used in schools. Such evaluation takes a sustained program of research over many years, comparable to the research detailed in the later chapters of this book. In the following section, we will call out the specific MTSS features of SpringMath and discuss how implementation science is embedded in SpringMath to facilitate effective results.

SpringMath at Tier 1

SpringMath measures are rigorous, aligned to grade-level standards, and can be used to sensitively detect risk and patterns of risk where risk first appears, which is often below the grade level that presents as the first noticeable sign to leaders. SpringMath interprets the screening data to identify risk and to recommend the next important instructional action to take. Specifically, if fewer than 50% of students in a class score in the not-at-risk range on all screening skills, SpringMath will recommend and enable classwide intervention. Figure 2.5 shows students' scores on the fall screening measures at grade 4. On the first measure, less than half the class is in the not-at-risk range. Thus, a classwide intervention is recommended for this class.

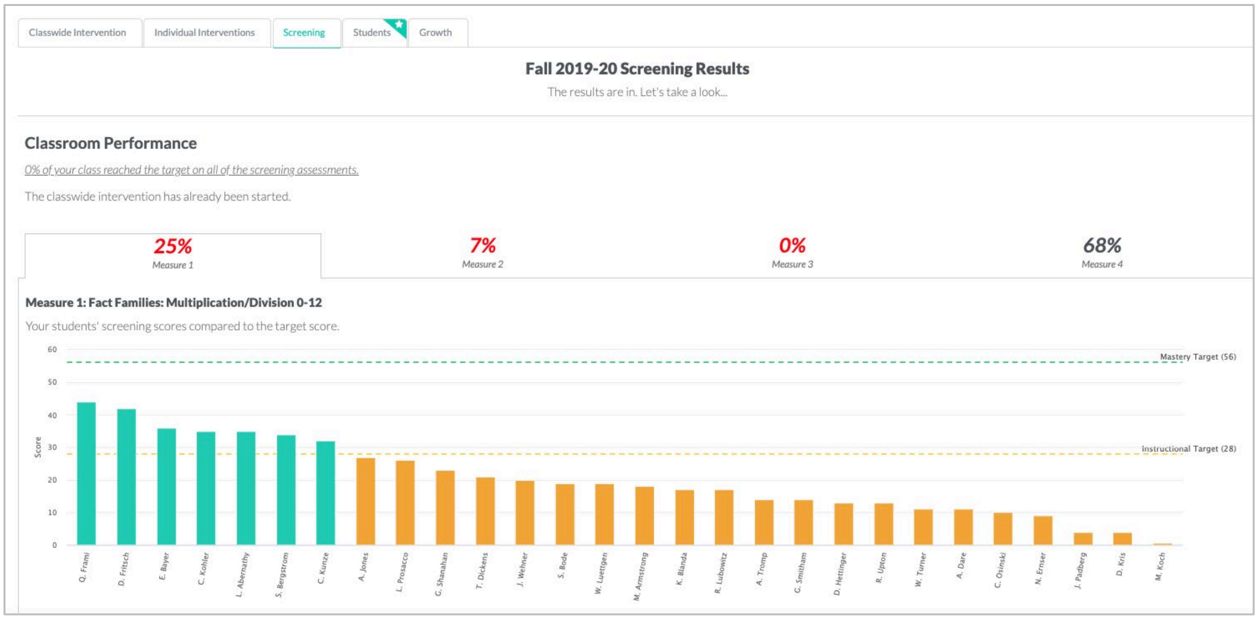


Figure 2.5. Fall screening data indicate that this class needs a classwide intervention.

If there is not a need for classwide intervention, SpringMath will identify all students in need of Tier 2 and/or 3 interventions and will recommend these students for diagnostic assessment (Figure 2.6).

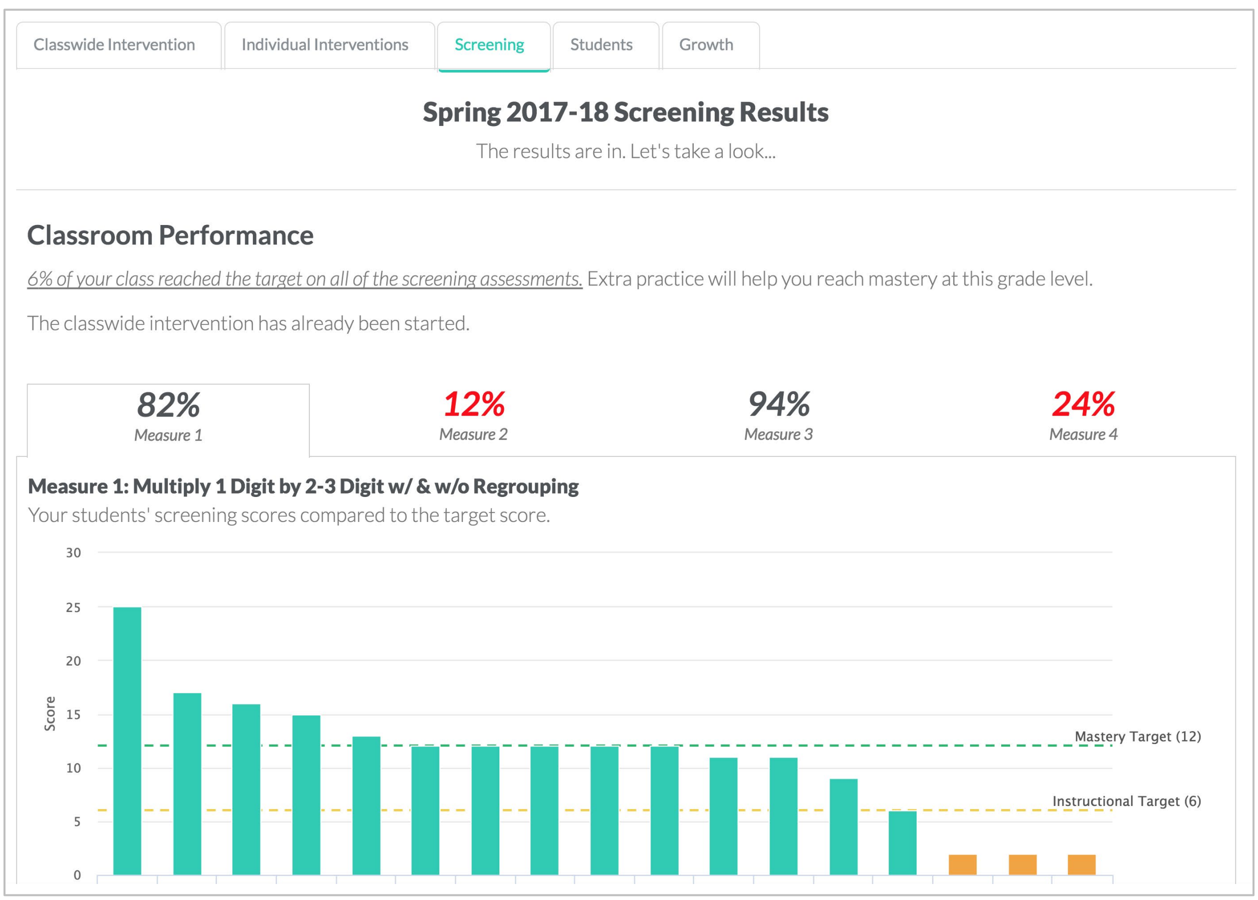


Figure 2.6. Fall screening data indicate that this class does not need a classwide intervention. The three students whose bars are yellow are recommended for diagnostic assessment and Tier 2 or 3 intervention.

For teachers for whom classwide intervention is not needed, their dashboard will open in the Individual Interventions view the next time they log in. At any time, the teacher can select the Students tab and view their roster. Clicking on a given student's name will open complete summary reports of that student's data in real time. Student pages are highly useful, comprehensive records of student performance that can be used during data team meetings, parent-teacher conferences, and as a record of MTSS for all students (Figures 2.7, 2.8, and 2.9).

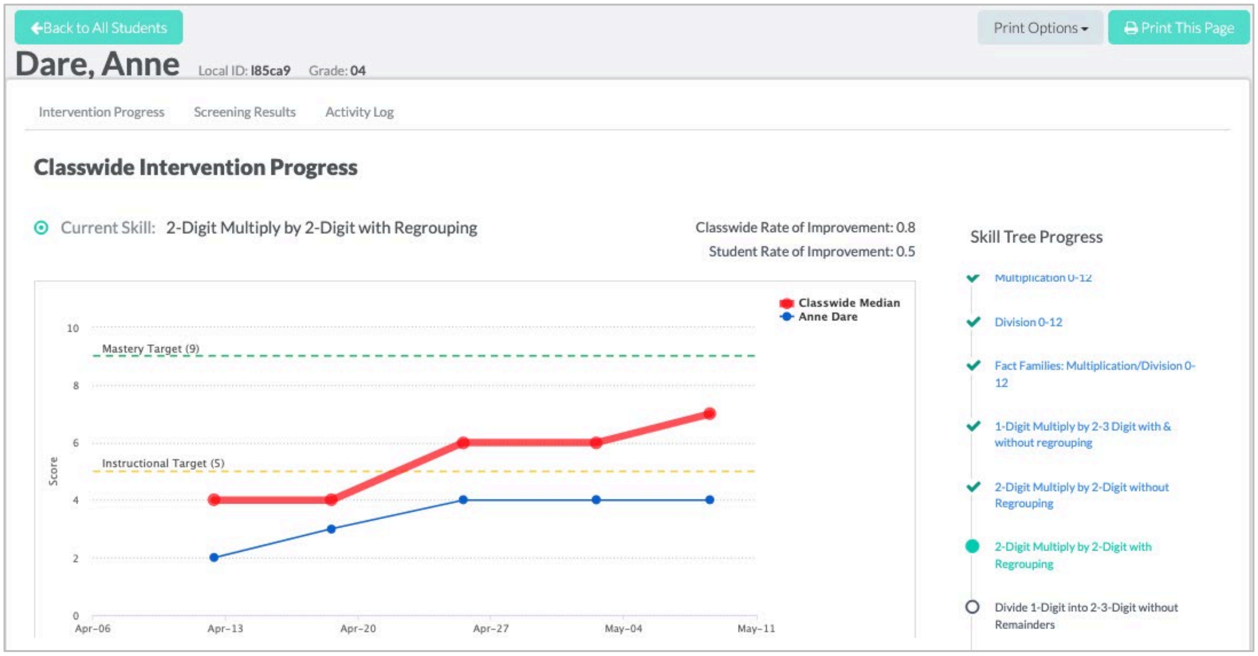


Figure 2.7. The teacher or coach can tab though skills that were mastered during classwide intervention and view Anne's progress (blue) compared to the median for her class.

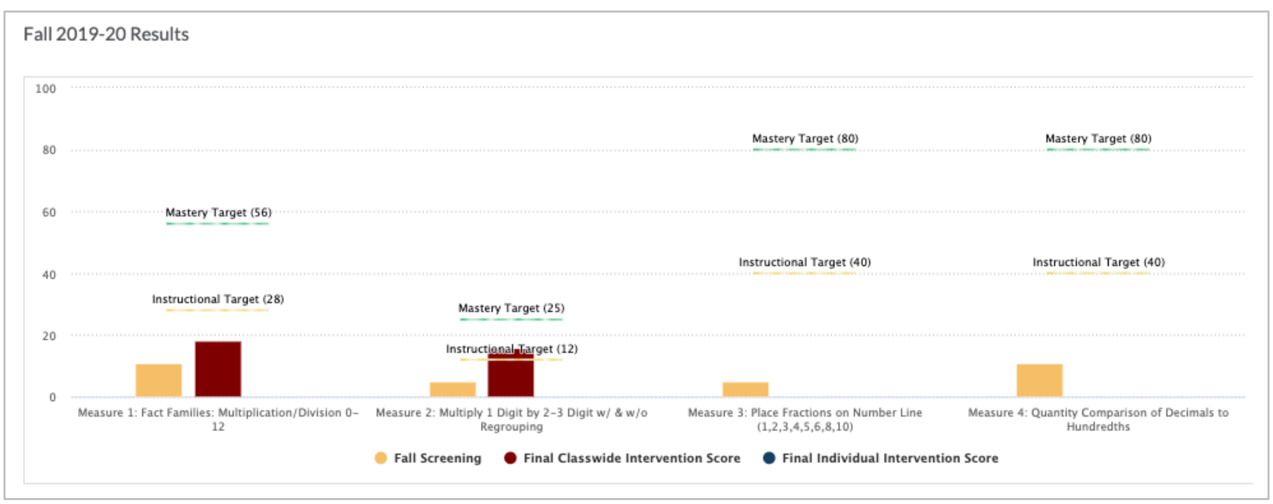


Figure 2.8. The teacher or coach can easily view screening data for Anne on fall, winter, and spring measures relative to instructional and mastery targets before (tan) and after classwide math intervention on that skill (burgundy).

Figure 2.9. The teacher and coach can view Anne's individual intervention progress on each targeted intervention skill (top panel) along with gains on the goal skill (bottom panel). These data along with those in Figures 5.7 and 5.8 are easily shared with parents via email or in printed form and can be used for eligibility determination using RTI.

SpringMath at Tier 1.5

If classwide intervention is needed, SpringMath will recommend classwide intervention and enable that path. The teacher dashboard will subsequently open in the Classwide Intervention page, which shows the sequence of skills that will be used on the righthand side. The teacher can click the button labeled Create Intervention Materials to retrieve the intervention materials for the week, including the scripted intervention protocol and all student materials. The center of the teacher dashboard displays the graph of student progress.At least once per week, the teacher enters the scores that result from conducting the classwide intervention. SpringMath graphs and interprets the results and adjusts the intervention for the following week, providing new materials that the teacher can access by clicking the Create Intervention Materials button. In Figure 2.10, this fourth-grade class has mastered about half of the skills in the sequence of skills shown in the righthand panel. The red line is the median score during classwide intervention on the current skill for which the mastery criterion has just been reached. The teacher receives a notification and can download and print new intervention materials by clicking on the button above the graph. In addition, the teacher can click on the button labelled Show Student Scores to display all student progress as shown in Figure 2.11.

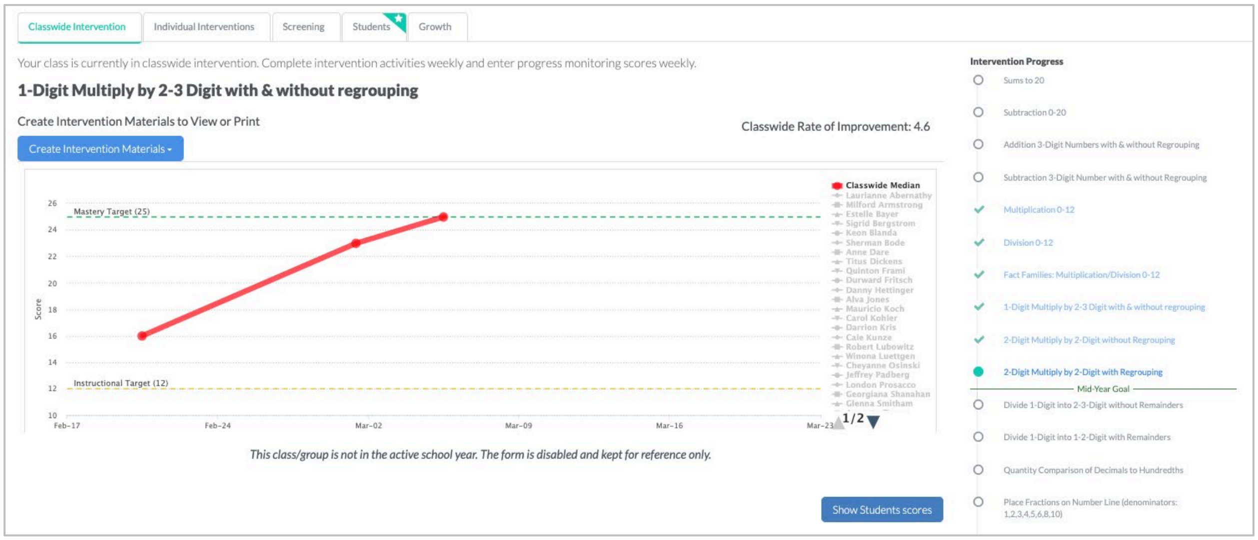


Figure 2.10. The median score for the class on a given skill in classwide intervention. The class median has just reached mastery so the class will be advanced to the next skill in the sequence.

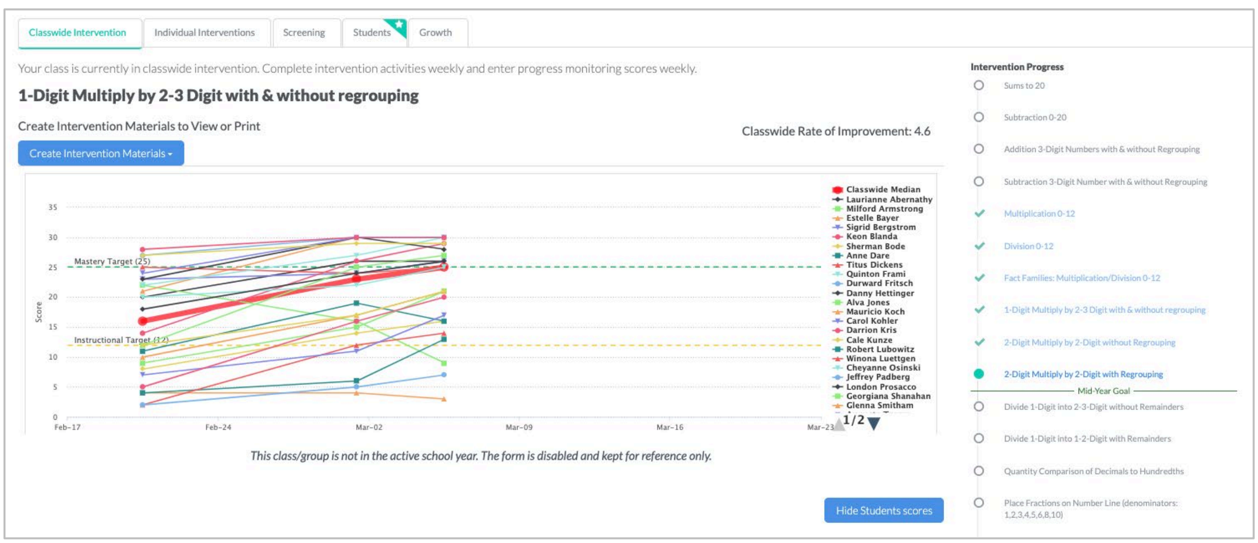


Figure 2.11. In this view, we can see all students' scores (and gains) during classwide intervention for the students whose median scores were displayed in Figure 5.7. This view makes apparent which two students are likely to be recommended for diagnostic assessment and individual intervention.

Classwide intervention is a standard protocol fluency-building intervention by design and works specifically because of the sequencing of the practice skills selected for each grade level. If the class median falls into the frustrational range (class median is below the instructional target) on any skill, SpringMath will recommend an acquisition lesson to first reteach that skill. The acquisition lesson includes scripted activities for the teacher to build both procedural and conceptual knowledge and all necessary student materials to conduct the lesson classwide. After reteaching the skill, the teacher can return to the classwide math intervention standard fluency-building protocol usually within one day. Ideally classes progress to skill mastery in about three to four weeks per skill with consistent use. When the class median reaches mastery for the skill, the entire class is advanced to the next skill in the sequence. In real time, SpringMath summarizes classwide intervention progress at the school, grade, and class level, and alerts the coach and principal to initiate a troubleshooting consultation with in-class coaching support where student learning gains signify the intervention is not being well implemented.

From the coach dashboard, coaches and other team members with coach access can view the overall progress of intervention in the school, facilitate data-team meetings, and provide in-class coaching support where it is needed. In the fifth-grade example in Figure 2.12, the coach would be recommended to visit the first two classes and the last class given their slower rate of progress and less consistent implementation. Following the coach visit, the coach can choose to "reset" the implementation metrics that are tracked to direct the coach to provide support if desired.

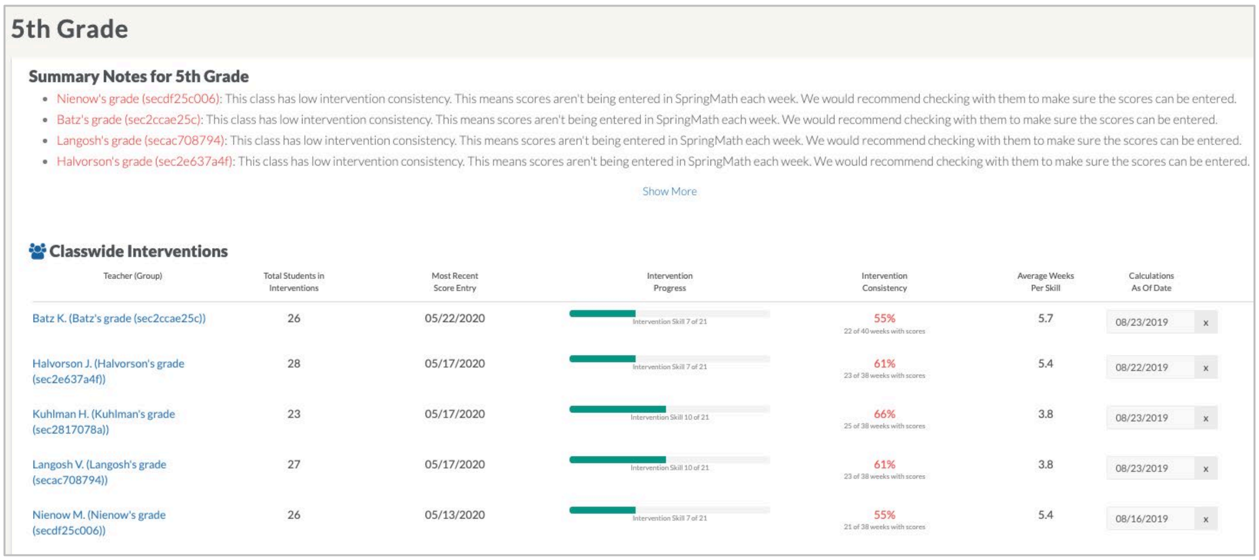


Figure 2.12. In the coach dashboard, teams can view the rate of progress across classes in the same grade and the consistency of use and view recommended actions that coaches can take to help teachers implement more effectively.

Also in the coach dashboard, teams can view the growth of students at the grade level as shown in Figure 2.13. These graphs and others are updated in real time as teachers implement SpringMath.

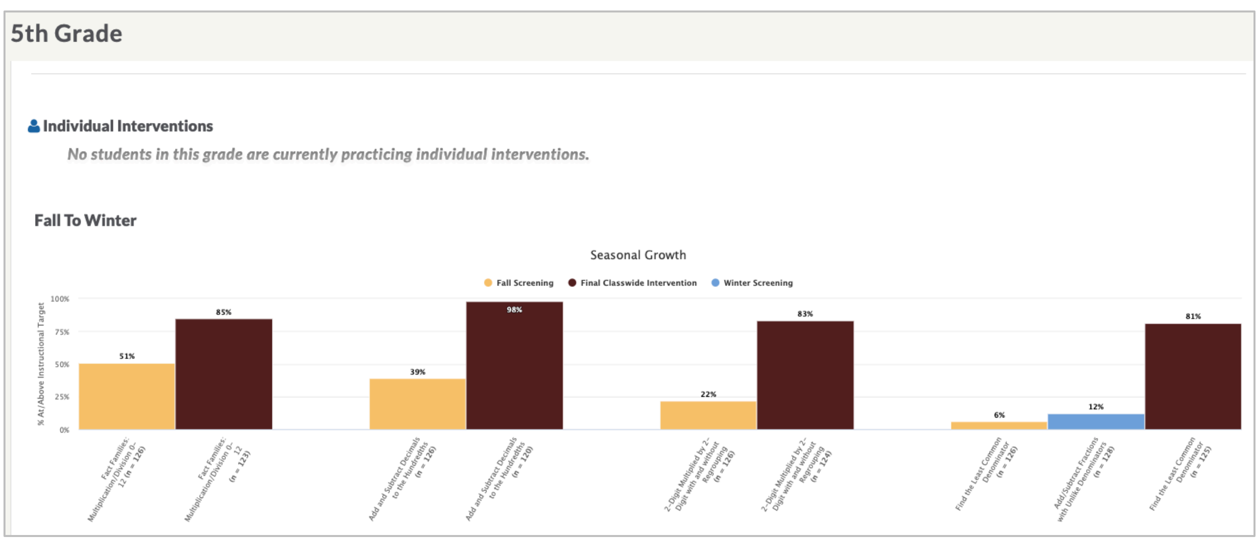


Figure 2.13. From the coach dashboard, coaches and other data team members can view the growth that has occurred in grade 5 from fall screening (yellow bars) to the final classwide intervention session for the same skill (brown bars) in terms of percent of students not at risk.

From the coach dashboard, the coach can click into any teacher's dashboard to facilitate live or virtual consultation. A teacher's growth data are provided in the teacher dashboard in a similar way so that teachers always have access to their own data in the same way that a coach would, except that coach views also include data aggregated across classes within grades. For example, a fourth-grade teacher can view his or her progress with intervention implementation and growth as shown in Figure 2.14.

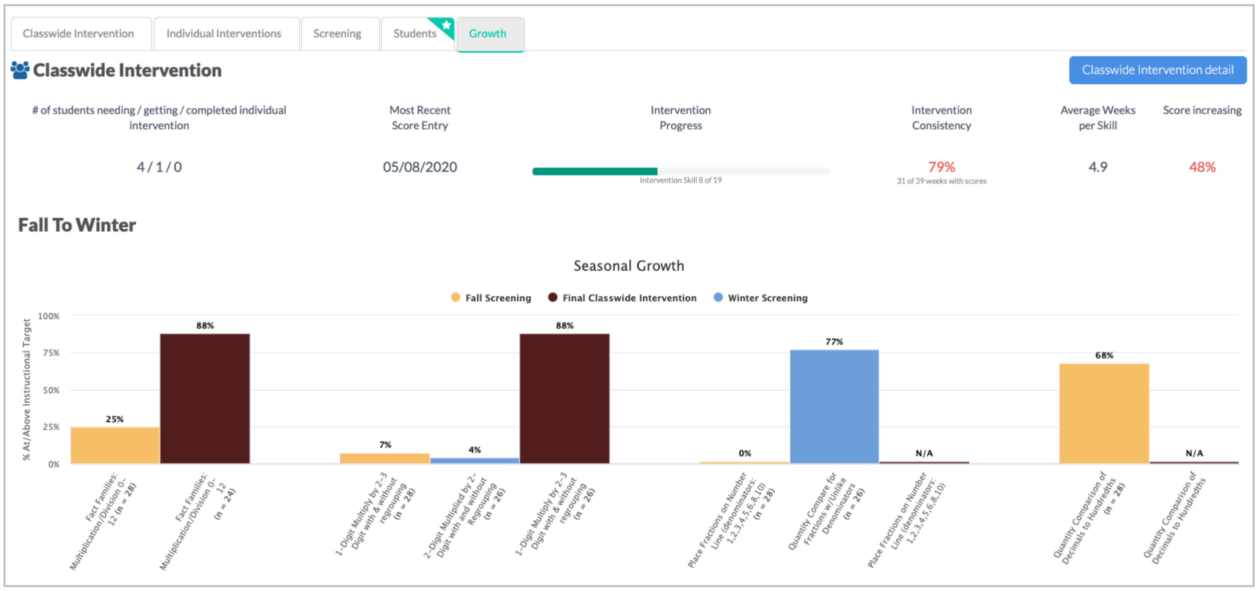


Figure 2.14. The teacher's view of his or her intervention implementation data and growth summary is under the Growth tab.

After four weeks of classwide intervention and in every week after, SpringMath will apply a decision rule to recommend students who need intervention intensification via Tier 2 or 3 intervention. The rule is applied when the class median score reaches mastery. There are a few caveats pertaining to integrity and missing data, but generally the rule identifies any student who remains in the frustrational range when the class median has reached mastery and recommends that student for intervention. Students recommended for intervention will be listed in the Students tab in the teacher dashboard and under the teacher's name in the coach dashboard. Notification will be provided in the teacher's classwide intervention dashboard that a student or students have been recommended for individual intervention. If more than one student is recommended, SpringMath uses student data to prioritize recommended students from greatest to least need. Once the teacher selects students for intervention, the students will appear in the teacher's Individual Interventions view as shown in Figure 2.15. Everything the teacher needs to implement individual intervention beginning with the diagnostic assessment and followed by intervention packets (including the scripted lesson and student materials) and the student's progress-monitoring graphs will be available on the Individual Interventions page.

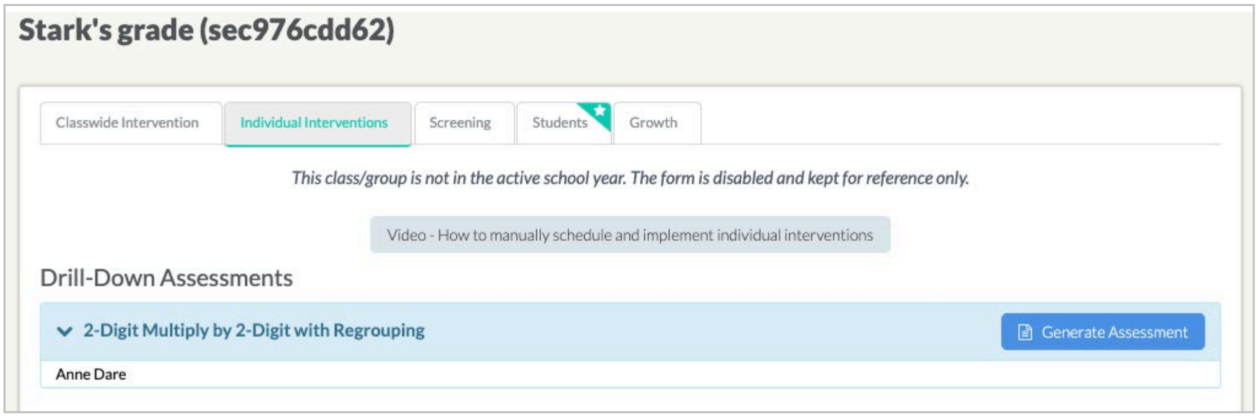


Figure 2.15. In this example, Anne Dare has been recommended for individual intervention based upon her scores during screening and classwide intervention. Her data will now appear in the Individual Interventions tab, where the teacher will be prompted to complete each step, beginning with the diagnostic assessment. Her data will also be in her page under the Students tab and she will continue to be included in classwide intervention.

The rule used by SpringMath to recommend students for individual intervention is empirically derived and continually evaluated to ensure its accurate functioning (these data are reported in Chapter 4). However, if a system has additional resources or does not want to wait for SpringMath to identify students for intervention, a coach can schedule any child in the school for individual intervention under the Students tab by clicking the three dots next to the student's name and then selecting "Schedule Intervention" from the drop-down menu.

SpringMath at Tiers 2 and 3

Given the promising data about group size not being a key dimension of intensification and the efficacy of classwide intervention to reduce overall intervention need, SpringMath combines Tiers 2 and 3 interventions for implementation. Specifically, any student who is identified for Tier 2 or 3 intervention is given a complete diagnostic assessment, which takes about 20 minutes. When more than one student has been recommended for intervention, the diagnostic assessments can be administered in small groups to improve efficiency (SpringMath will recommend this automatically). The diagnostic assessment begins with grade-level screening expectations and samples backward through incrementally prerequisite skills to identify the "right" intervention skill target and instructional tactic for each student as detailed in Chapter 3. Following diagnostic assessment, SpringMath identifies children within a class or grade who need a very similar intervention and recommends them as a small group for the week (see Figure 2.16). SpringMath analyzes student progress in real time and adjusts recommended groupings accordingly. Every student receives his or her own practice materials and unique progress-monitoring assessments, and every student's data is summarized independently showing his or her progress on the targeted intervention skill and the generalization or goal skill.

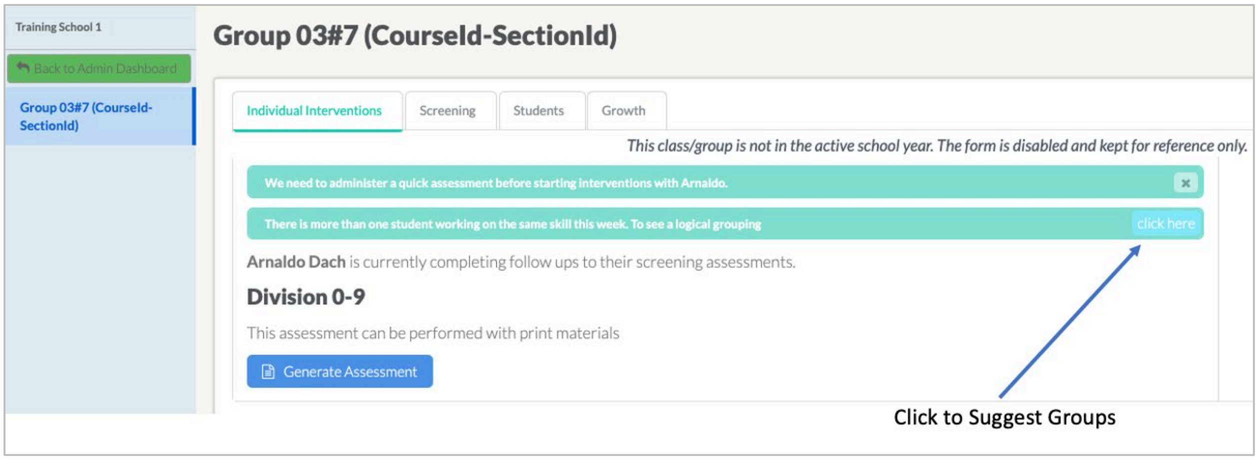


Figure 2.16. When more than one student needs the same intervention, SpringMath will recommend that those students receive intervention as a small group. These groupings are adjusted each week and updated in the teacher's dashboard. Individual student data can be viewed on the Individual Interventions page and also in the student's page under the Students tab.

As with classwide intervention, summaries of individual intervention progress are held in each student's page in the teacher dashboard, and the coach dashboard summarizes individual interventions in real time.

SpringMath decision making

SpringMath uses empirically derived and experimentally tested decision rules for all decisions at all Tiers. Following each teacher action, SpringMath provides a data summary and recommends the next key action along with any needed materials to take that action. Automated data interpretation and provision of all needed materials effectively removes two of the most common barriers to correct MTSS implementation, namely failing to interpret and then act on collected data.

SpringMath was specifically built to include antecedent and consequent supports for effective implementation. SpringMath's Theory of Change (see Chapter 3) specifies that the automated data interpretation and provision of all needed materials to take each recommended MTSS step will improve implementation integrity and drive stronger learning gains. Thus, implementation science is built into SpringMath. We will cover implementation support in greater detail in Chapter 7.

References

Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovaleski, J. F., & Prasse, D. (2005). *Response to intervention: Policy considerations and implementation*. National Association of State Directors of Special Education.

Clarke, B., Doabler, C. T., Kosty, D., Kurtz Nelson, E., Smolkowski, K., Fien, H., & Turtura, J. (2017). Testing the efficacy of a kindergarten mathematics intervention by small group size. *AERA Open, 3*(2), 1–16. doi.org/10.1177/2332858417706899

Doabler, C. T., Clarke, B., Kosty, D., Kurtz-Nelson, E., Fien, H., Smolkowski, K., Baker, S. K. (2018). Examining the impact of group size on the treatment intensity of a Tier 2 mathematics intervention within a systematic framework of replication. *Journal of Learning Disabilities, 52*(2), 168-180. doi.org/10.1177/0022219418789376

Koon, S., & Davis, M. (2019). Math course sequences in grades 6–11 and math achievement in Mississippi (REL 2019–007). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from ies.ed.gov/ncee/edlabs

Kovaleski, J., VanDerHeyden, A. M., Runge, T., Zirkel, P., & Shapiro, E. (2022). *The RtI Approach to Evaluating Learning Disabilities, 2nd Edition*. NY: Guilford.

Morrison, J. Q., & Harms, A. L. (2018). Advancing evidence-based practice through program evaluation: A practical guide for school-based professionals. Oxford University Press.

VanDerHeyden, A. M. (2010). Determining early mathematical risk: Ideas for extending the research. *School Psychology Review, 39,* 196-202. doi.org/10.1080/02796015.2010.12087773

VanDerHeyden, A. M. (2013). Universal screening may not be for everyone: Using a threshold model as a smarter way to determine risk. *School Psychology Review, 42,* 402-414. doi.org/10.1080/02796015.2013.12087462

Ysseldyke, J., Burns, M., Dawson, P., Kelley, B., Morrison, D., Ortiz, S., Rosenfield, S., & Telzrow, C. (2006). *School psychology: A blueprint for training and practice III.* Bethesda, MD: National Association of School Psychologists.

CHAPTER 3

Overview of SpringMath

3.1 Background

SpringMath is a complete, schoolwide multi-tiered systems of support (MTSS) designed to accelerate math achievement for every student in a school. It includes screening and progress-monitoring assessments, classwide and individual interventions, automated data interpretation and recommended actions, and a coach dashboard to facilitate effective implementation. SpringMath is based on Intervention Adviser (Education Research & Consulting, Inc., 2013), which was also developed by Dr. Amanda VanDerHeyden. SpringMath includes a novel assessment process validated through research, decision trees to automate data interpretation and deliver the right interventions to teachers, evidence-based interventions, and implementation science built into all stages of use.

3.2 Theory of change

There are three active ingredients to the SpringMath theory of change shown in Figure 3.1: (1) alignment of student need with intervention content or skill (Solomon et al., 2022; VanDerHeyden, Broussard, & Burns, 2019; VanDerHeyden, Codding, & Martin, 2017), (2) alignment of student need with instructional or intervention tactic (Burns, Codding, Boice, & Lukito, 2010; Burns, VanDerHeyden, & Jiban, 2006; Haring & Eaton, 1978; VanDerHeyden, 2013), and (3) increased correct intervention use (VanDerHeyden, McLaughlin, Algina, & Snyder, 2012; Witt, Noell, LaFleur, & Mortenson, 1997). Providing students with the evidence-based interventions that target their specific measured learning needs results in higher quality instruction and corresponding improvements in math achievement. In addition, the assessments provided by SpringMath accurately identify specific student needs, allowing for targeted instruction. When combined with frequent, sensitive, progress-monitoring assessments, SpringMath allows for much more efficient intervention than a generic and imprecise approach to intervention that attempts to address a broad range of student needs. Decision accuracy is maximized because SpringMath automates data interpretation at each stage of decision making (screening, intervention, program evaluation). Finally, the intentional efficiency of the tool in recommending the necessary actions, providing all needed materials to take the recommended action, summarizing the effects of such action, and tracking actions across the school to facilitate system-level troubleshooting increases student achievement (Codding, VanDerHeyden, & Chehayeb, in submission; VanDerHeyden, Codding, Martin, Desai, Maki, & McKevett, in submission; VanDerHeyden et al., 2012).

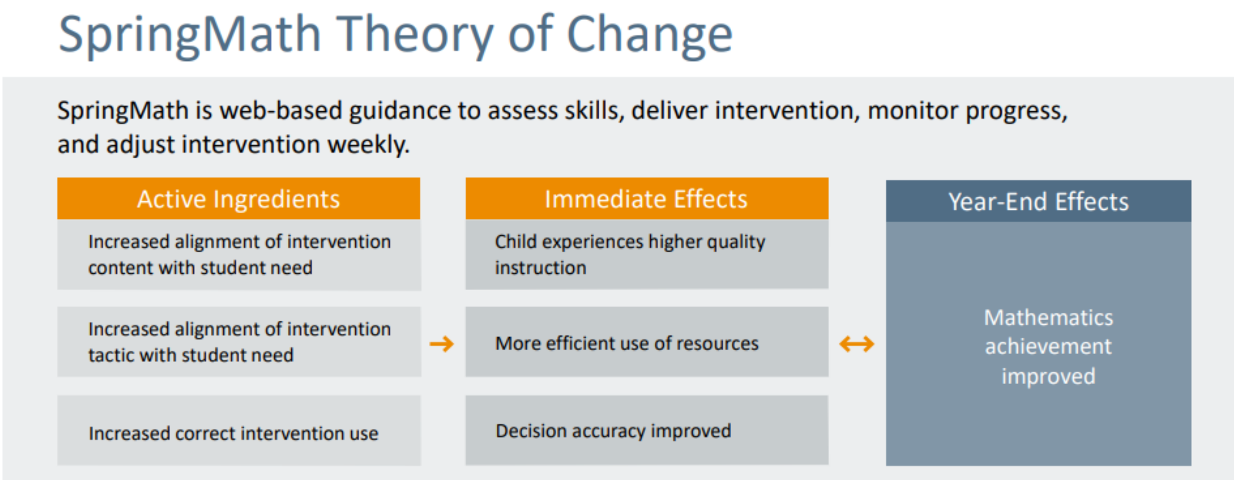


Figure 3.1. The SpringMath Theory of Change specifies that SpringMath enables better alignment of intervention content and tactic with student need using a sequence of very sensitive measures in tandem with intervention trials and increased correct intervention use, which results in higher quality, more efficient, and more accurate instructional decision making, which in turn causes improved math achievement.

3.3 How it works

SpringMath uses best practices in assessment, intervention, and implementation science. Universal screening occurs three times per year. As shown in Figure 3.2, if the median score on the screening measures is below the instructional target, the class is directed into a classwide intervention. If the median is at or above the instructional target, students in need of individual interventions will be identified. Once the interventions have begun, progress-monitoring measures are administered at least weekly, although teachers may monitor progress and enter scores more often to advance more quickly to new interventions. Based upon student performance, the interventions are adjusted for the following week. Students may move up to the next skill, continue working at their current level, or have the type of intervention adjusted to meet their instructional level (i.e., intervention may change from an acquisition intervention to a fluency-building intervention based on student learning gains).

Classes in classwide intervention may have students recommended for individual intervention based upon their individual response to classwide intervention, which can be viewed in each student's page under the Students tab in the teacher dashboard. Once all interventions are completed, the interventions are discontinued.

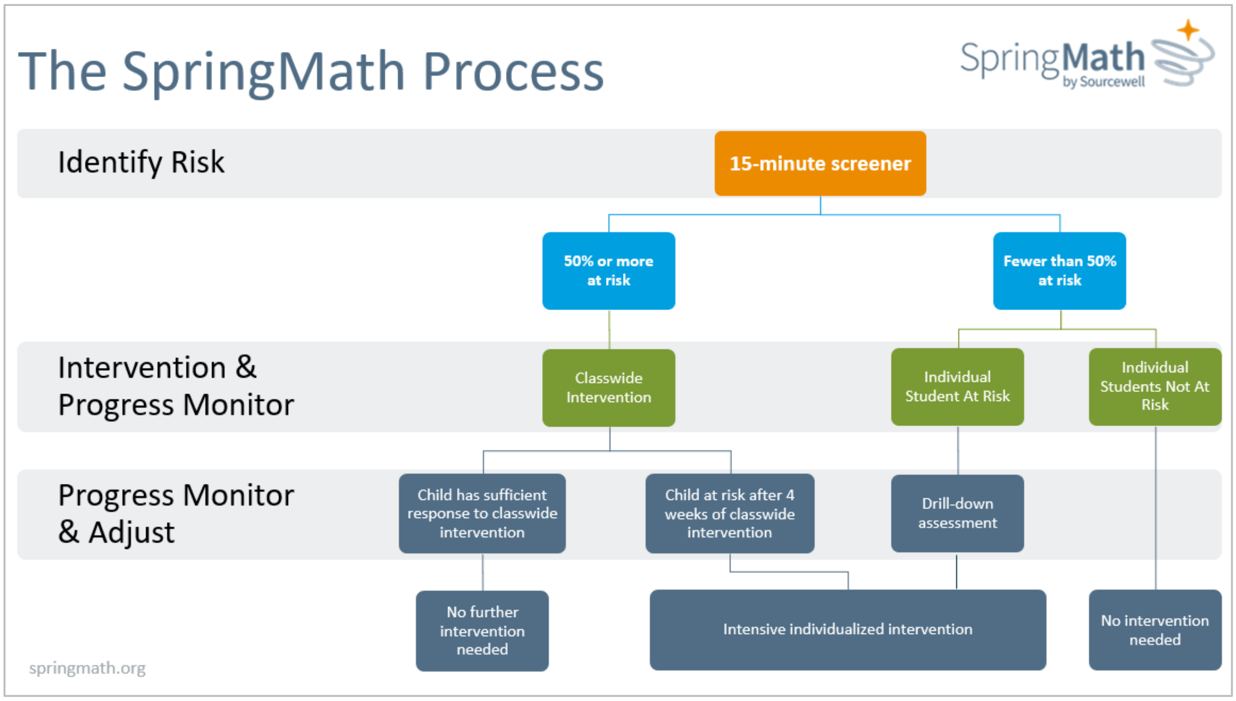


Figure 3.2. SpringMath is a comprehensive MTSS system that begins with screening, delivers classwide intervention where it is needed, and uses available data to sensitively determine which students need individual intervention.

SpringMath includes:

Assessments

- Teachers administer brief universal screening assessments to their students as a group.
- Assessments vary in length from one to four minutes.
- Based upon the results of the screening assessments, SpringMath produces summary reports and recommends either classwide or individual intervention for any student who may be struggling.
- SpringMath assesses weekly progress of students in response to all levels of intervention (classwide, individual) and summarizes progress at the student, class, grade, and school level.

Classwide intervention

- SpringMath provides intervention packets targeting specific sequences of skills that are unique to each grade level.
- Students work in pairs to complete specific activities for 15-20 minutes per day.
- Each week students are assessed on the skill they are being taught.
- Class progress is graphed and when the class has achieved the target score, the class moves on to the next skill.
- Individual students who experience less success with classwide intervention are recommended for individual intervention.
- The coach dashboard organizes implementation and learning gains by class and advises teams to provide consultation and in-class coaching where it is needed.

Individual intervention

- Diagnostic assessments are administered following proprietary decision trees (Education Research & Consulting, 2013) to determine exactly which skills require intervention and what type of intervention should be provided (acquisition or fluency-building).
- Individual intervention packets are generated focusing on the identified skill and containing the evidence-based ingredients specific to the student’s measured learning needs (i.e., acquisition instruction or fluency-building instruction).

- Progress is monitored on the targeted skill and a more distal goal skill and graphed weekly. Interventions adapt based on student progress to bring the student to mastery on the grade-level goal skill.

Reporting

- Actionable reports provide teachers and coaches the information they need to monitor progress and administer the interventions with fidelity.
- The coach dashboard organizes implementation and learning gains by class, recommends students for small groups, and advises teams where consultation and in-class coaching are needed for more effective intervention implementation.

3.4 Implementation features

Screening

- The screening skills were selected with the goal of forecasting readiness for successful algebra completion rather than predicting and avoiding mathematical failure at grade 3 or 5. These skills are well aligned with the Common Core state standards and recommendations of the National Math Panel (2008). The screening assessments are administered in the fall, winter, and spring, and the measures change to reflect the progression of skills over the course of the school year.
- If the median score for the class is below the instructional target, classwide math intervention is provided.

Classwide intervention

- Classwide interventions follow a static sequence of skills unique to each grade level.
- Classwide interventions are primarily designed to build fluency and take about 15 minutes per day.
- During classwide intervention, students pair up and work together to complete specific activities using materials provided in a printable packet. The essential elements of this intervention are:
  - High dosage of opportunities to respond on an instructional-level skill
  - Increased academic engagement
  - Corrective feedback and student-conducted error correction
  - Motivation to beat the last best score
  - The ability to accommodate the needs of diverse learners

Individual intervention

- Automated identification of students in need of individual intervention either through screening or based on response to classwide intervention. Prioritization of highest need students (based on data) in individual intervention recommendations.
- Intervention skill and tactic aligned with student need based on assessment data.
- All intervention materials needed to deliver individual intervention and the weekly progress-monitoring assessment(s) are provided in a printable packet. Scripted activities are provided to build conceptual understanding and procedural skill.
- Gains summarized in dashboards and intervention adjusted based on growth.

Screening

SpringMath begins with universal screening at fall, winter, and spring. Typically, the fall and winter screenings are used to identify classes for classwide intervention and/or to identify individual students for individual intervention. The spring screening is most commonly used for program evaluation.

At each grade level, for fall, winter, and spring, there are three or four screening assessments. Screening skills reflect a skill that should have been taught to the student prior to the screening, that the student is expected to have acquired, and acquisition of which is necessary for the student to benefit from the instruction that is forthcoming at that grade level. Weighty skills or skills that are most strongly connected to mastery of algebra by eighth grade are afforded more in-depth assessment and instruction (e.g., proportional understanding for grades 3, 4, 5, and 6). Screening skills are aligned with the Common Core standards by grade level. However, in several cases, developmentally appropriate instructional options are provided at early grades on skills not explicitly listed in the Common Core to facilitate successful mastery of Common Core skills that appear at a later grade.

Screening skills were selected with the goal of forecasting readiness for successful algebra completion rather than predicting and avoiding mathematical failure at grade 3 or 5. This subtle shift in purpose of screening causes the screening assessments within SpringMath to be more rigorous than many schools may be accustomed to. As a result, SpringMath measures demonstrate excellent sensitivity, which is the most important feature of an accurate screening tool. Many math screening tools lack sensitivity. In other words, they include measures that may be thought of as “too easy.” If a child fails a “too easy” measure, then the child is certainly in trouble and at risk of failure (this test characteristic is called “specificity”). But when a child passes a “too easy” measure, the child may still have a high likelihood of actual math failure. SpringMath measures have been designed to be rigorous, to forecast readiness for successful algebra learning in grade 8, and to be sensitive screeners.

Screening measures in SpringMath

	Fall	Winter	Spring
Kindergarten	• Count Objects to 10, Circle Answer • Identify Number, Draw Circles to 10 • Quantity Comparison with Dots to 10 • Missing Number 0-10	• Count Objects, Write Number to 20 • Identify Number, Draw Circles • Quantity Comparison with Dots to 20 • Missing Number 0-20	• Change Quantity with Dots to 10 • Missing Number 0-20 • Sums to 5 Kindergarten • Subtraction 0-5 Kindergarten
Grade 1	• Sums to 6 • Subtraction 0-5 • Quantity Comparison 20-99	• Sums to 12 • Subtraction 0-9 • Fact Families Addition & Subtraction 0-5	• Sums to 20 • Subtraction 0-20 • Fact Families Addition & Subtraction 0-9 • Quantity Comparison 101-999
Grade 2	• Sums to 20 • Subtraction 0-20 • Fact Families Addition & Subtraction 0-20 • Quantity Comparison 1001-9999	• 2-digit Addition without Regrouping • 2-digit Subtraction without Regrouping • Quantity Comparison Sums & Differences to 20 • Create Equivalent Addition & Subtraction Problems Using Place Value & Decomposition)	• 2-digit Addition with Regrouping • 2-digit Subtraction with Regrouping • Create Equivalent Addition & Subtraction problems Using Associative Property & Near Easy
Grade 3	• Fact Families Addition & Subtraction 0-20 • 3-digit Addition with & without Regrouping • 3-digit Subtraction with & without Regrouping	• Multiplication 0-9 • Division 0-9 • Fact Families Multiplication & Division 0-9	• Multiply 1- by 2-3 digit without Regrouping • Divide 1-digit into 2-3 digit without Remainders • Quantity Comparison Fractions with Like Denominators • Place Fractions on Number Line I

Grade 4	• Fact Families Multiplication & Division 0-12 • Multiply 1- by 2-3 with & without regrouping • Place Fractions on Number Line II • Quantity Comparison Decimals to Hundredths	• Multiply 2-digit by 2-digit with & without Regrouping • Add & Subtract Mixed Numbers with Like Denominators & Regrouping • Quantity Comparison for Fractions with Unlike Denominators	• Add & Subtract with Decimals to Hundredths • Convert Decimals to Fractions & Fractions to Decimals • Quantity Comparison Fractions, Decimals, Whole Numbers • Create Equivalent Multiplication Problems by Factoring
Grade 5	• Fact Families Multiplication & Division 0-12 • Add & Subtract with Decimals to Hundredths • Multiply 2-digit by 2-digit with & without Regrouping • Find Least Common Denominator	• Convert Improper Fractions to Mixed Numbers • Add & Subtract Fractions with Unlike Denominators • Quantity Comparison Fractions, Decimals, Whole Numbers	• Simplify Fractions • Multiply & Divide Decimals • Multiply & Divide Proper & Improper Fractions • Quantity Comparison with Whole Numbers, Fractions, Decimals, & Percents
Grade 6	• Add & Subtract Fractions with Unlike Denominators • Order of Operations • Multiply 2-digit by 2-digit with Decimals • Multiply & Divide Mixed Numbers	• Distributive Property of Expression • Collect Like Terms • Find Percent of a Whole Number • Mixed Fraction Operations	• Substitute Whole Number to Solve Equations • Mixed Decimal Operations • Graph Points in a Coordinate Plane • Quantity Comparison with Negative Numbers
Grade 7	• Solve Algebraic Proportions • Solve Missing Value in a Percentage Problem • Mixed Operations Integers	• Order of Operations • Mixed Inverse Operations • Complex Fractions	• Solve 2-step Equations • Translate Verbal Expressions into Math Equations • Solve 2-step Equations with Fractions
Grade 8	• Distributive Property to Simplify Expressions • Collect Like Terms to Simplify Expressions • Solve for Slope & Intercept using Linear Function	• Mixed Operations with Exponents • Order of Operations II • Point on a Line	• Linear Combinations to Solve Equations • Substitute Equation to Solve Linear Equations • Comparison Method to Solve Linear Equations
Grades 9-12	Classwide Intervention Data Are Used to Determine Need for Individual Intervention*		

*High school students have a substantial amount of data available on their math performance, and generally schools do not require additional data to identify high school students who may struggle in math. Most high schools offer course sequence tracks for which students in need of remedial support have already been identified. Thus, additional academic screening is not needed.

In grades K-8, three to four measures are given at each screening occasion and reflect current grade-level skills that differ by season (grades 9-12 screening measures are administered per year in each grade). If there is a high base rate of risk at screening (defined as median score in the frustrational range), then the class is routed into classwide intervention. Weekly progress-monitoring data are then harvested using the SpringMath rules engine to support correct classwide intervention use and to identify children who need intensified instruction in that context. In grades 9-12, the screening process begins with classwide intervention, and individual student responding during classwide intervention is the basis for advancing to individual intervention.

Following screening, students can be recommended for individual intervention in three different ways.

If the median score on the screening measure is above the instructional target, a classwide intervention is not necessary and students who scored the below the instructional target on the screener will be recommended for individual assessment and intervention if needed (VanDerHeyden, Witt, & Naquin, 2003; VanDerHeyden & Witt, 2005; VanDerHeyden, 2010; VanDerHeyden, 2013).

If the class goes into classwide intervention, SpringMath will monitor the progress of students based upon their weekly scores. SpringMath will identify students whose progress is not sufficient, and they will be recommended for individual intervention. When a class has entered classwide intervention, classwide intervention responsiveness is more accurate than static screening for determining who needs individual intervention (VanDerHeyden, Broussard, & Burns, 2019).

Coaches can manually assign students to individual intervention from the Students tab in their dashboard.

Classwide intervention

The classwide intervention skill sequence varies by grade level, beginning with below-grade-level foundation skills and progressing to essential grade-level skills. The whole class advances to the next skill based upon mastering the first skill and so on. The classwide intervention is a standard protocol fluency-building intervention, which has been evaluated for efficacy (Codding et al., 2016; VanDerHeyden et al., 2012).

A significant body of evidence supports the effectiveness of classwide math intervention. Specifically, the efficacy is driven by (1) delivering a high dosage of opportunities to respond for all students during a short instructional interval using (2) the right level of task difficulty in a sequence of skills that builds over time. By design, the classwide intervention is a fluency-building intervention. It is important to note that the efficacy of classwide math intervention is not caused by the peer tutoring arrangement, but rather the peer tutoring arrangement is the format that enables all children to experience a high dosage of opportunities to respond during the same interval of time. Simply having students work together is not the same as what occurs during classwide math intervention. Sometimes people might think that "peer tutoring" means that one student is teaching another student, and that is not correct. In the academic literature, peer tutoring involves a specific way of matching students into working pairs, providing the right level of task difficulty for practice, providing a high dosage of practice opportunities called "opportunities to respond," having each student take on the role of the worker and the helper, and ongoing progress monitoring to increase task difficulty as students' skills improve.

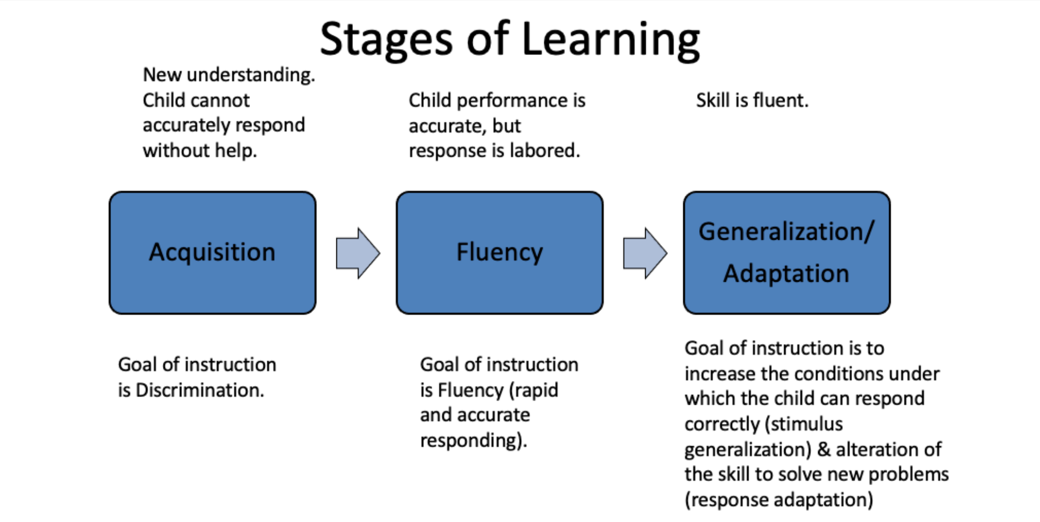
In SpringMath, when a class scores below the instructional target, the teacher is provided with an acquisition lesson to reteach the skill to the class as a whole before continuing with fluency building. The acquisition lesson includes scripted activities and materials to build procedural knowledge and conceptual understanding. The purpose of this step is to prevent the misapplication of a fluency-building intervention when the class has not yet learned the skill. If the acquisition lesson is recommended, then the teacher can deliver that lesson for one to two sessions and then return to the classwide intervention for fluency building (i.e., the standard classwide intervention).

Individual intervention

Once children are identified for individual intervention, SpringMath administers diagnostic assessment to select the intervention that is aligned to the student's learning needs in terms of the targeted skill and instructional tactic. Proprietary decision trees that direct all actions within SpringMath guide the decision-making process based upon the assessment results (Education Research & Consulting, 2013). Grounded in the science of the Instructional Hierarchy, these diagnostic assessments connect the child to the right difficulty level and tactic for intervention, which is central to our theory of change.

The Instructional Hierarchy (Haring & Eaton, 1978), illustrated in Figure 3.3, states that learning progresses predictably through each of the following stages: Acquisition, Fluency, and Generalization/Adaptation. During each stage, learner proficiency differs and the goals of instruction and the specific tactics used necessarily must differ for optimal learning. During Acquisition, children are in the frustrational range of performance where errors are highly likely and the goal of instruction is to establish correct and independent performance and understanding. During the Fluency stage of learning, student performance is in the instructional range, where learning gains will accelerate rapidly if students are provided with the correctly aligned fluency-building strategies (e.g., high-quality practice with a high dosage of opportunities to respond). Errors are rare in the instructional range.

It is necessary to measure fluency directly using brief timed assessments because most children in the instructional range are accurate, and untimed accuracy scores will no longer be meaningful to characterize continued learning. In other words, accuracy cannot grow beyond 100%. The goal of instruction during fluency building is to make the responding easier and more automatic. In the Generalization/Adaptation stage of learning, performance is in the mastery range. Student learning in this stage is characterized as flexible, enduring, likely to be remembered, useful, and generalizable. Mastery performance is strongly associated with a student's ability to solve novel problems (generalization) and experience faster learning of related but more complex skills (Burns et al., 2006).



Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23–40). Columbus, OH: Merrill.

Figure 3.3. The Instructional Hierarchy is the framework for aligning instructional tactic with student proficiency to deliver intensified instruction.

Teachers are often unfamiliar with the Instructional Hierarchy, but once understood, it is a powerful way to intensify instruction at all Tiers of MTSS. All students should receive acquisition, fluency building, and generalization instruction during core instruction every day, which means necessarily different skill targets for each. In other words, it is not possible for the same skill to reflect acquisition, fluency, and generalization stage learning for a given student and class. Conceptual and procedural instruction should be used in tandem every day at every stage of learning, as well, and this process is called "interleaving," which has been shown to result in more robust mathematical skill mastery and achievement (Rohrer et al., 2019). A common misunderstanding among math teachers is that they must first establish conceptual understanding before teaching procedural skill, which is not possible. Procedural and conceptual skills develop in concert over time, and teachers should dedicate instructional opportunity to both every day for skills at all stages of learning (Rittle-Johnson, 2017; Rittle-Johnson, Schneider, & Star, 2015; Schneider, Star, & Rittle-Johnson, 2011; Star, 2005).

Intervention tactics used in SpringMath

Many teachers believe in a particular strategy and try to use it regardless of student proficiency. The only instructional tactic that works similarly at all stages of learning is explicit instruction (Chodura, Kuhn, & Holling, 2015; Morgan, Farkas, & Maczuga, 2015). Other popular tactics (e.g., inquiry-based learning, productive struggle) have not been shown to be effective in rigorous experimental research and, in fact, have been shown to be ineffective (Kirschner, Sweller, & Clark, 2006). When ineffective tactics are provided, overall learning is harmed because ineffective tactics rob valuable instructional time and resources and are unlikely to produce strong and stable learning gains for most children. On the other hand, some of these popular tactics could perhaps be part of a comprehensive program of instruction if used with greater precision. For example, inquiry-based learning and productive struggle might be useful generalization tactics where students have mastered a skill and are presented with a challenging, untaught iteration. As another example, modeling a problem solution through drawing or manipulatives might be a useful acquisition tactic, especially if the algorithm is taught in concert with a focus on helping students understand how and why the algorithm works. In any case, adding explicit instruction to core instruction and selecting interventions that use explicit instructional techniques are important parts of MTSS.

Classwide math intervention is designed as a fluency-building intervention to supplement core instruction. Thus, the active ingredients of classwide math intervention are the delivery of a high dosage of opportunities to respond in a short interval of time using instructional-level tasks. The peer tutoring format used in classwide math intervention is specifically used to increase the opportunities to respond experienced by all students in a short interval of time. Other formats could accomplish the same goal, for example, use of response cards and choral responding, but the classwide peer tutoring format is the one that SpringMath uses. Classwide math intervention has been demonstrated to produce gains on year-end test scores (Codding et al., 2016; VanDerHeyden & Burns, 2005; VanDerHeyden et al., 2012), to close opportunity gaps by race (sometimes called equity gaps) (VanDerHeyden & Codding, 2015), and to produce a very strong return on investment as indicated by very low incremental cost effectiveness ratios (Barrett & VanDerHeyden, 2020).

Diagnostic assessment to select individual intervention for the student

Decision trees that begin with the grade and season screening skills are specified. Each screening skill is the first skill in a given tree and samples back through prerequisite skills, specifying frustrational, instructional, and mastery-level performance for each skill and grade level. Thus, all 145 skills are contained in hierarchical and independent decision trees for which the screening is the “dropping in” point, closely connected to grade-level content. The diagnostic assessment follows the decision tree to identify the correct intervention skill target and intervention tactic (acquisition or fluency building).

The diagnostic assessment identifies the right skill target and tactic for the student according to the Instructional Hierarchy. This logic is shown in Figure 3.4, which shows that the diagnostic assessment samples back through incrementally prerequisite understandings to determine where the skill gap is and whether the student needs acquisition instruction or fluency-building instruction. Each skill is assessed using a technically adequate measure and a specific range of cut scores that indicate whether the student’s performance is in the mastery, instructional, or frustrational range.

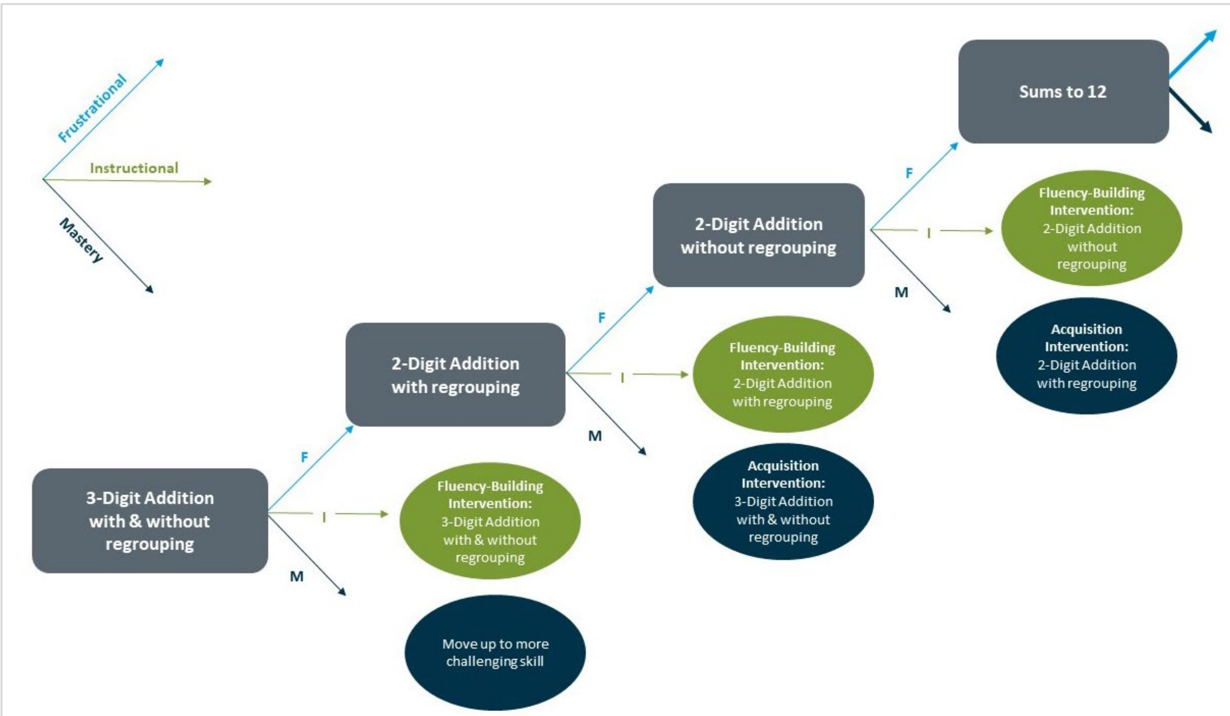


Figure 3.4. Overview of the diagnostic, sampling back technique to identify starting intervention skill and tactic.

Weekly assessment of the skill target and the generalization target (the screening skill) is conducted to monitor progress and advance the intervention content and tactic until the student reaches mastery on the grade-level skill, at which time the student can be exited from intervention.

In SpringMath, all approximately 145 skills have acquisition intervention protocols and fluency-building protocols. The formats we use for acquisition are cover-copy-compare, guided practice, bingo, and a few others unique to kindergarten and first grade. The formats we use for fluency building include timed trials and response cards. SpringMath individual interventions do not target mastery-level performance, but instead embed generalization problems and opportunities in the conceptual understanding part of the fluency-building intervention protocols.

Individual intervention protocols and materials in SpringMath

All materials are provided with each intervention protocol, including scripted instructions, practice materials to build procedural skill, scripted activities and materials to build conceptual understanding, and the progress-monitoring assessment(s).

The interventions are provided to the students daily. Classwide intervention takes about 15 minutes per day, and individual intervention requires about 20 minutes per day. Findings strongly support the use of a shorter, more frequent intervention dosage, which is what SpringMath uses (Codding et al., 2016). Codding et al. (2016) randomly assigned students to different dosages of classwide math intervention in SpringMath: one control condition where students received no intervention but did receive weekly progress monitoring; a once-per-week 48-minute session; twice-per-week 24-minute sessions; and four times per week 12-minute sessions. Thus, the overall investment of time was the same in all treatment conditions. All students got 48 minutes per week of intervention, but the group that got the 48 minutes in 12-minute doses over four days showed double the growth of students in the next-best condition, which was the 24-minute sessions twice-per-week. Thus, the dosage recommended for SpringMath interventions is daily or a minimum of four days per week.

SpringMath Intervention Protocols

Intervention formats

Designed to establish correct understanding and responding.

- Whole-class guided practice lessons
- Guided practice
- Cover-copy-compare
- Bingo
- Fill-in bingo (kindergarten & grade 1 only)
- Before and after game (kindergarten only)
- Scatter and rearrange (kindergarten only)

Fluency-building intervention formats

Designed to make responding easier and to facilitate generalization.

- Standard classwide math intervention
- Timed trial
- Response card

All interventions are selected to align with measured student learning needs according to the Instructional Hierarchy and driven by sensitive, accurate assessment of student learning provided by the SpringMath assessments.

Procedural knowledge and conceptual understanding are targeted daily in all individual interventions in an interleaved fashion (Rittle-Johnson, 2017; Rohrer et al., 2019).

Weekly progress-monitoring data are used to adjust the intervention and provide new materials to the teacher each week as shown in Figure 3.5. The focus of the intervention can be to build either acquisition of new skill understanding or fluency. Once intervention begins, weekly progress-monitoring scores are used to advance the content and tactic toward grade-level proficiency (we call this the "goal skill").

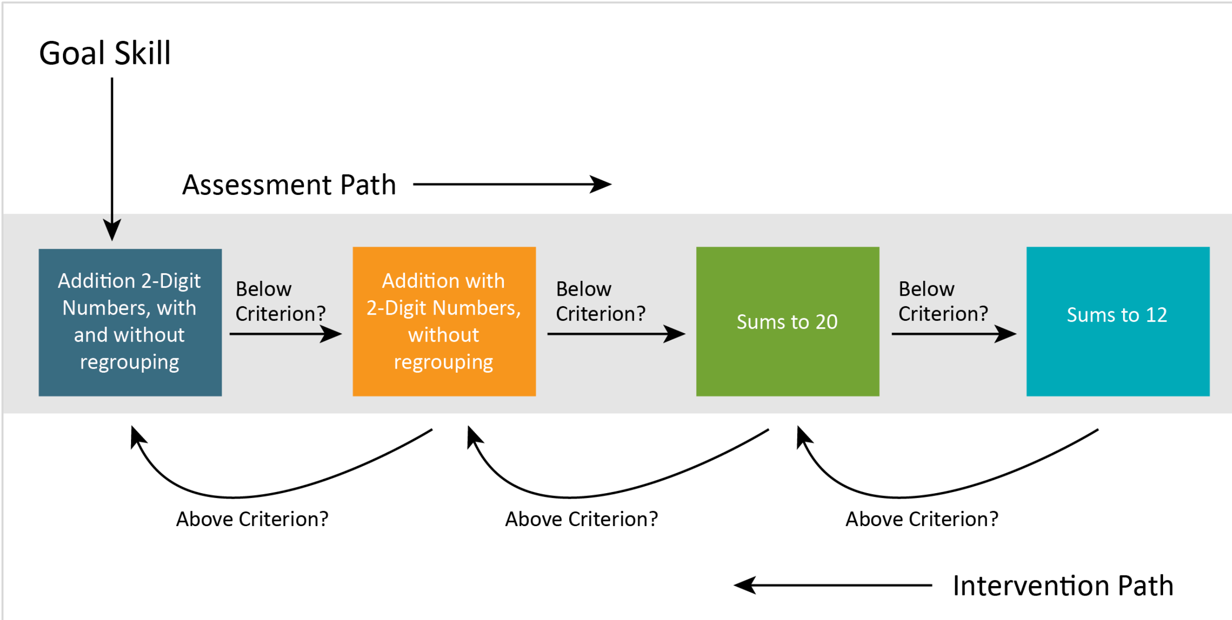


Figure 3.5. Once a child is assigned to a specific intervention tactic and a specific level of task difficulty (based on the diagnostic assessment), weekly progress-monitoring assessment is given on the intervention skill and the goal of the school and students are advanced back toward the starting skill until they reach mastery on the goal skill.

Intervention implementation support

Implementation support is provided to ensure that interventions occur with integrity and at the optimal dose. Antecedent supports are provided, including: efficient, easy-to-administer assessments, automated data interpretation, efficient mechanisms of intervention delivery (e.g., classwide intervention), scripted protocols for intervention, and provision of all needed materials to take the action recommended by SpringMath at each step. Onboarding training is required to position teachers for effective implementation, and ongoing coaching is strongly encouraged. In addition (and most importantly), SpringMath provides consequent support for implementation by automating growth charts in the teacher and coach dashboards so that teams can see the results of their instructional efforts immediately (Noell et al., 2005).

Specific antecedent and consequence supports for implementation built into SpringMath

Antecedent supports for implementation	Improve efficiency to minimize number needed	• Use of screening rules to minimize over-identification of students for more intensive instruction. • Use of implementation support in classes using classwide intervention to avoid over-identification of students for more intensive instruction.
	Minimize complexity	• Assessments are brief, with standard instructions printed on each. Scoring is answers correct and answer keys are provided. • Assessment data interpretation is automated. • Intervention protocols are scripted. • All materials needed to implement intervention are provided.
	Improve teacher skill to Implement	• Training is embedded in SpringMath with brief "just-in-time" videos explaining how to complete each recommended step. • Support portal contains materials, training videos, and guidance on specific questions that might arise. • Onboarding training is required and positions teams to implement. • Virtual and on-site training support is available at an added cost.
Consequent supports for implementation	Automated data interpretation recommending next action	• Following each score entry, performance is summarized, and the next action is recommended.
	Frequent, sensitive feedback on learning gains	• The coach dashboard shows gains at the school, grade, class, and student level. These data are harvested and updated with assessment that occurs as part of the intervention. • The teacher dashboard shows gains at the class and student level. These data are harvested and updated with assessment that occurs as part of the intervention.
	Guidance to enable performance feedback	• Sensitive implementation metrics are tracked and summarized in the coach dashboard so coaches can understand what actions are occurring and with what results. • Recommendations to provide in-class coaching or consultation with the teacher are suggested for classes and students who are not showing expected progress. • Coach and teacher dashboards can be used to guide data team meetings.
	Program evaluation	• Program evaluation is automated and provides effects by dosage of SpringMath to enable more effective implementation.

Proprietary metrics (Education Research & Consulting, 2013) drive recommendations in the coach dashboard to provide in-class coaching support where it is needed to improve intervention effects schoolwide (Joyce & Showers, 1981; Fixsen & Blasé, 1993; VanDerHeyden et al., 2012). These metrics are sensitive and reflect dosage of intervention and results across grade levels. They are provided via the coach dashboard, which allows school leaders to know where intervention is being used, with what results, and to see the results in real time. To provide consistent metrics and facilitate data team discussions, a parallel version of this dashboard is provided for teachers for their own classrooms. In other words, the coach can view all teachers' information and each individual teacher can view the same information but only specific to his or her own classroom.

Questions answered by the coach dashboard include:

- What actions are underway (e.g., screening, classwide intervention, individual intervention)?
- What are the results right now?
- Where is support needed?
- Are proximal indicators headed in the right direction?
- What are the barriers we can troubleshoot?

An example of a dashboard is provided in Figure 3.6:

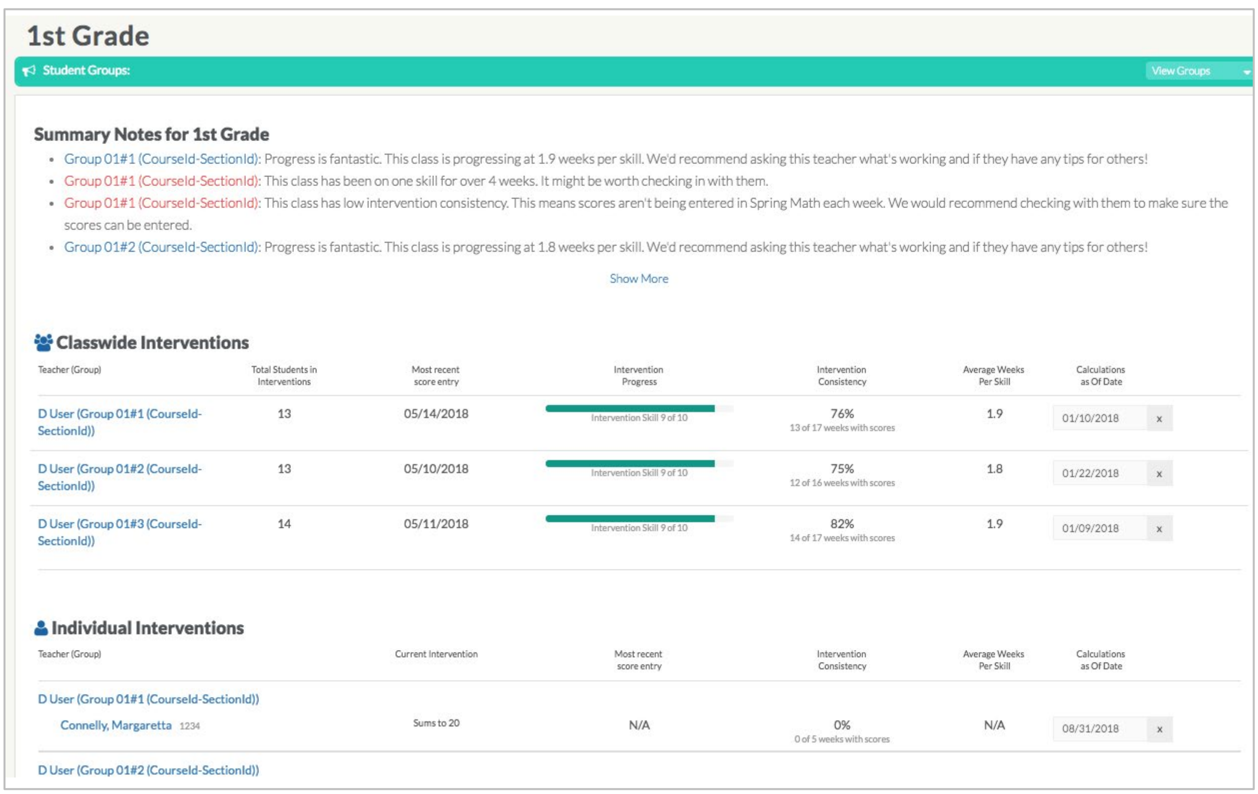


Figure 3.6. The coach dashboard summarizes progress within and across classes and grades showing rate of skill mastery and identifying classes in need of coaching support for both classwide and individual intervention.

The teacher dashboard provides instant, graphed feedback to the teacher, reflecting gains resulting from the use of intervention as shown in Figure 3.7. This feedback loop to the teacher is a critical element of implementation designed to drive stronger intervention use over time.

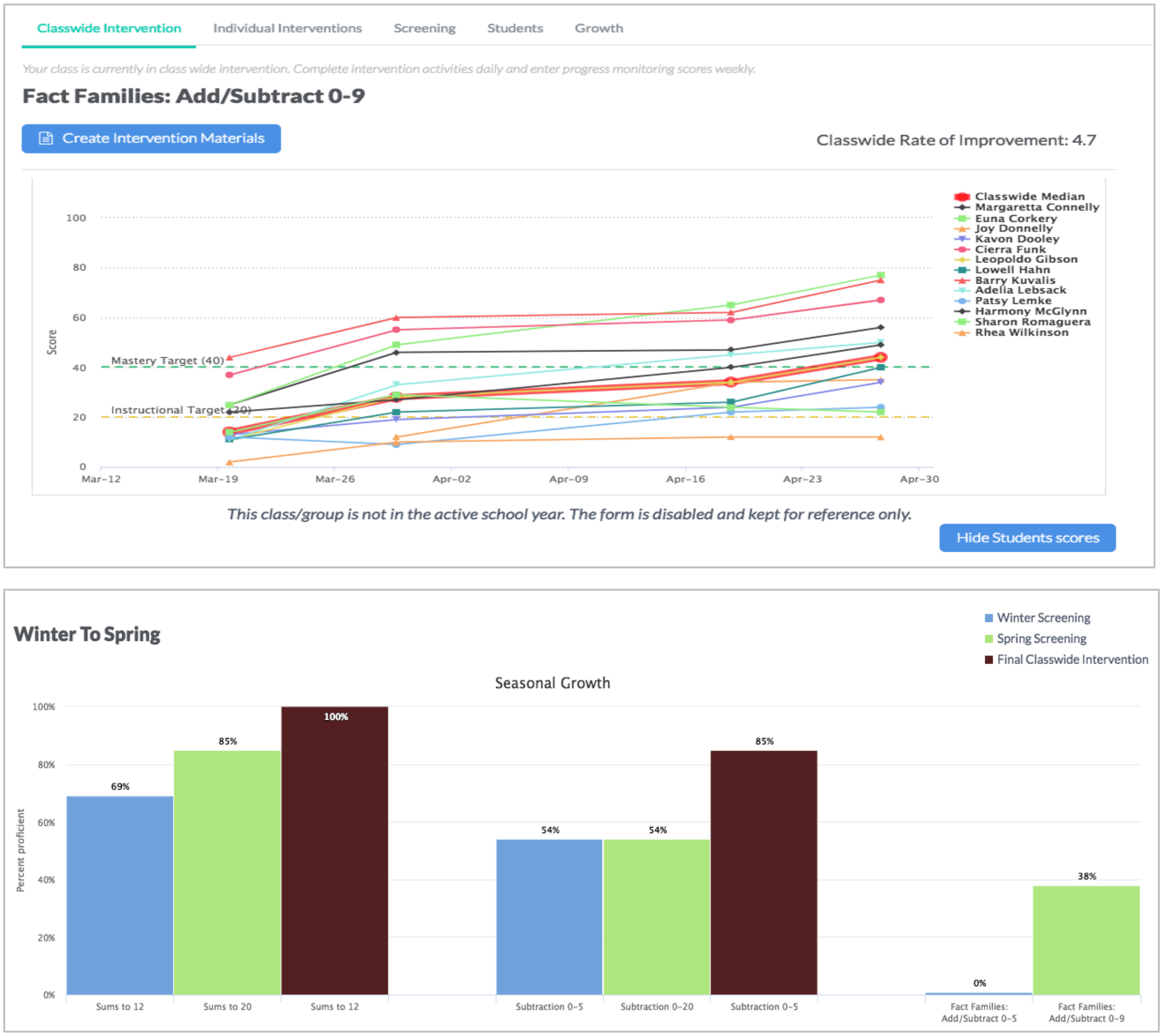


Figure 3.7. Growth is shown in the top panel for all students in a class receiving classwide intervention. These gains tell the teacher that the intervention is working to improve performance for all students and makes apparent which students may be lagging behind and recommended for individual intervention. The bottom panel reports the percent of students not at risk on related screening assessments from winter (blue) to spring (green) and on the winter screening skill in the final classwide intervention session for that skill if the class has received classwide intervention. These graphs provide the teacher with immediate feedback on the growth caused by classwide intervention. Gains are generally detectable in the top panel view within two weeks. When upward gains are not observed, in-class coaching and troubleshooting to improve intervention delivery are generally successful.

The proprietary metrics in the coach dashboard (Education Research & Consulting, 2013) are used in all supplemental training services that are available from Sourcewell. These services are available for a fee and include a range of support from virtual training sessions to on-site coaching and support by our trained coaches. An onboarding session is required for new subscribers and is provided for a fee. All additional training services are optional but highly recommended for new users.

References

Barrett, C. A., & VanDerHeyden, A. M. (2020). A cost-effectiveness analysis of classwide math intervention. Journal of School Psychology, 80, 54-65. doi.org/10.1016/j.jsp.2020.04.002

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. School Psychology Review, 39, 69-83.

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. (2006). Assessing the instructional level for mathematics: A comparison of methods. School Psychology Review, 35, 401-418.

Chodura, S., Kuhn, J.-T., & Holling, H. (2015). Interventions for children with mathematical difficulties: A meta-analysis. Zeitschrift für Psychologie, 223, 129-144.

Codding, R., VanDerHeyden, Martin, R. J., & Perrault, L. (2016). Manipulating Treatment Dose: Evaluating the Frequency of a Small Group Intervention Targeting Whole Number Operations. Learning Disabilities Research & Practice, 31, 208- 220. doi.org/10.1111/ldrp.121206

Education Research & Consulting, Inc. (2013). SpringMath (version 2.2.1) [Computer Software. Sourcewell. springmath.org

Fixsen, D. L., & Blasé, K. A. (1993). Creating new realities: Program development and dissemination. Journal of Applied Behavior Analysis, 26, 597-615. doi.org/10.1901/jaba.1993.26-597

Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), The fourth R: Research in the classroom (pp. 23–40). Columbus, OH: Merrill.

Joyce, B. R., & Showers, B. (1981). Transfer of Training: The Contribution of "Coaching." Journal of Education, 163, 163-172. doi.org/10.1177/002205748116300208

Kirschner, P. A.; Sweller, John, & Clark, Richard, E. (2006): Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. Educational Psychologist, 41, 75-86.

Morgan, P. L., Farkas, G., & Maczuga, S. (2015). Which instructional practices most help first-grade students with and without mathematics difficulties? Educational Evaluation and Policy Analysis, 37, 184-205.

National Mathematics Advisory Panel (2008). Foundations for Success: The Final Report of the National Mathematics Advisory Panel, U.S. Department of Education: Washington, DC.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L., Resetar, J. L., & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. School Psychology Review, 34, 87–106. doi.org/10.1080/02796015.2005.12086277

Rittle-Johnson, B. (2017). Developing mathematics knowledge. Child Development Perspectives, 11, 184-190. doi.org/10.1111/cdep.12229

Rittle-Johnson, B., Schneider, M., & Star, J. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. Educational Psychology Review, 27, 587-597. doi.org/10.1007/s10648-015-9302-x

Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C. N. (2019, May 16). A randomized controlled trial of interleaved mathematics practice. Journal of Educational Psychology. Advance online publication. dx.doi.org/10.1037/edu0000367

Schneider, M., Star, J. R., & Rittle-Johnson, B. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. Developmental Psychology, 47, 1525-1538. doi.org/10.1037/a0024997

Star, J. (2005). Reconceptualizing procedural knowledge. Journal for Research in Mathematics Education, 36, 404-411.

Solomon, B., G., VanDerHeyden, A. M., Solomon, E. C., Korzeniewski, E. R., Payne, L. L., Campaña, K. V., & Dillon, C. R. (2022). Mastery Measurement in Mathematics and the Goldilocks Effect. School Psychology.

VanDerHeyden, A. M. (2010). Determining early mathematical risk: Ideas for extending the research. School Psychology Review, 39, 196-202. doi.org/10.1080/02796015.2010.12087773

VanDerHeyden, A. M. (2013). Universal screening may not be for everyone: Using a threshold model as a smarter way to determine risk. School Psychology Review, 42, 402-414. doi.org/10.1080/02796015.2013.12087462

VanDerHeyden, A. M., & Broussard, C. (2019). Construction and Examination of Math Subskill Mastery Measures. Advance Online Publication. Assessment for Effective Intervention. doi.org/10.1177/1534508419883947

VanDerHeyden, A. M., Broussard, C., & Burns, M. K. (2019). Classification agreement for gated screening in mathematics: Subskill mastery measurement and classwide intervention. Assessment for Effective Intervention. Advance Online Publication. doi.org/10.1177/1534508419882484

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. Assessment for Effective Intervention, 30, 15-31. doi.org/10.1177/073724770503000302

VanDerHeyden, A. M. & Codding, R. (2015). Practical effects of classwide mathematics intervention. School Psychology Review, 44, 169-190. doi.org/10.17105/spr-13-0087.1

VanDerHeyden, A. M., Codding, R., Martin, R. (2017). Relative value of common screening measures in mathematics. School Psychology Review, 46, 65-87. doi.org/10.17105/SPR46-1.65-87

VanDerHeyden, A. M., Codding, R. S., Martin, R., Desai, S., Maki, E., & McKevett, N. (in submission). Use of a web-based mathematics intervention tool to promote implementation and student learning. Manuscript submitted for publication.

VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. American Education Research Journal, 49, 1251-1284. doi.org/10.3102/0002831212462736

VanDerHeyden, A. M. & Witt, J. C. (2005). Quantifying the context of assessment: Capturing the effect of base rates on teacher referral and a problem-solving model of identification. School Psychology Review, 34, 161-183. doi.org/10.1080/02796015.2005.12086281

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. School Psychology Review, 32, 204-227. doi.org/10.1080/02796015.2003.12086194

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. Journal of Applied Behavior Analysis, 30, 693-696. doi.org/10.1901/jaba.1997.30-693

CHAPTER 4

Assessment and Decision Making in SpringMath

4.1 Purpose of SpringMath measures

Assessments in SpringMath are used to determine whole-class and individual student risk. They are also used to conduct the diagnostic assessment to build the right intervention for the student. Finally, they are used to monitor progress, which then informs weekly intervention adjustments and triggers in-class implementation support when needed to assure successful intervention outcomes.

SpringMath measures were built using the science of curriculum-based measurement (Deno & Mirkin, 1977) and were constructed to sample specific skills, yield scores that are reliable for diverse groups and lead to valid decisions, and to allow for brief, repeated assessment to monitor progress over time.

4.2 Mastery measurement

Formative assessment can have a powerful effect on student learning. Formative assessment transforms the learning space by shifting the responsibility from the student *to receive* what is being taught to the teacher *to effectively teach* skills and understandings. Whether or not the student learns when taught becomes the most critical feedback to the teacher. When successful learning can be measured and successful learning is the measure of effective teaching, then teaching can become more meaningful. In fact, we might say that the learner is never wrong, since the learner's response is feedback to the teacher (Keller, 1968).

CBM as a behavioral observation and a score

The notion of learning being the barometer of instruction comes from the field of behavioral psychology. Skinner (1953) discovered rate of responding as a metric that could be used to evaluate environmental variables that could lead to therapies designed to promote greater freedom, adaptation, and learning. The field of precision teaching focused this science on how people learn. One of precision teaching's primary contributions was the addition of *accuracy* to Skinner's rate of responding. Prior to the advent of CBM, precision teaching researchers were conducting timed readings and math computations, recording words read correctly per minute or digits correct per minute. These counts of behavior correct per minute were referred to as *fluency*, which was defined as accuracy plus speed (Binder, 1996) and eventually became the response metric of CBM (Deno & Mirkin, 1977). Before CBM was widely used, precision teachers were screening entire classes of students, using those data to monitor and adjust instruction, and demonstrating programmatic gains in fluency. Haughton (1984) identified the first benchmark criteria, which he called R/APS for skill retention and application performance standards (Haughton, 1984). This work heavily influenced Deno and Mirkin (1977), who were seeking ways to measure student learning so that students could be provided with special education under the newly enacted Education of the Handicapped Act, Public Law No. 94–142 (1975), later named the Individuals with Disabilities in Education Act or IDEA.

With roots in Skinner's science of behavior, precision teaching researchers tended to view these metrics as idiographic behavioral observations that could be used to determine which instructional techniques benefitted specific students. Each student could be viewed as an experiment where the specific techniques of use to that student could be discovered using the principles of applied behavioral analysis. Deno & Mirkin referred to this work as data-based program modification and the metrics driving the instructional modifications were fluency on low-inference direct skills measures.

CBM researchers saw the potential for behavioral observations to become more generally useful if viewed as scores of demonstrated reliability, validity, and practical utility. Thus, CBM researchers turned their attention toward validating cut scores for decision making concerning who should get intervention, for example, and whether the intervention had worked. CBM researchers also turned their attention to the actual materials (e.g., reading passages) and procedures (e.g., timing and number of trials) used to collect these data points, which had received much less scrutiny from the precision teaching researchers.

Through a prolific program of research largely led by Deno and colleagues, CBMs became standard units that were sensitive to instructional changes (as opposed to differences in the difficulty of task materials or the measurement procedures or conditions). CBM permitted the evaluation of learning gains over time and short-term changes in learning caused by instructional changes, like reduced task difficulty, repeated practice, or the use of rewards (Noell et al, 1998). CBM was game changing in education because it created a standardized, reliable, and valid score from which instruction could be evaluated and adjusted for better learning (Stecker, et al, 2005).

CBM and the failed search for a general outcome measure (GOM) in mathematics

Beginning in the 1990s, words read correctly per minute became standard grist for studying and evaluating ingredients and even full recipes of instruction for students in reading. The idea of CBM, as set forth in an especially influential article, was that CBM functioned as a general outcome measurement system whereby a number of individual skills could be taught during intervention sessions, and CBM as a general outcome measure could reflect linear gains over some period of time to determine whether students were learning as expected (Fuchs & Deno, 1991).

The capacity to model learning in a linear way and to compare that learning to some expected rate of learning made measurement of learning generalizable and enabled data-based decision making in schools. CBM enabled screening and evaluation of programs of instruction at the district, school, class, and student level. CBM and MTSS created a radical shift in schools whereby teachers, parents, and students could reasonably evaluate current progress against desired progress and expect leaders to adjust programs of instruction to improve learning for all students. CBM and data-based decision making subsequently became the basis for response to intervention and multi-tiered systems of support.

Researchers attempted to construct CBMs in math using a curricular-sampling approach, meaning that researchers identified three to five skills that are reflective of the most important mathematical learning that is expected to occur during a school year. Those problems are then provided on a single probe. The rationale for curricular sampling is that as instruction progresses (and learning occurs), gains will be observed in a linear fashion as children master skills included on the measure. But the technical characteristics of these measures have always been weaker than those reported for reading CBM (Foegen et al., 2007), and an oral reading fluency equivalent in math has never been found.

The problem of sensitivity and mastery measurement as an alternative

Two problems have become apparent with regard to multiple-skill measures intended to function as general outcome measures in math, and both are problems of sensitivity. CBMs are used during screening to identify students who may be at risk for poor learning outcomes and in need of intervention. With multiple-skill math measures, only a small number of problem types (i.e., no more than four to five) can be used on a single measure, so the measures are constructed choosing four to five operational skills that are important for children to master during the school year. Therefore, children are assessed on problem types that they have not yet been taught how to do, especially at the beginning of the monitoring period. Thus, at the beginning of the monitoring period, score ranges will be severely constrained. These constrained scores destroy the capacity to use the scores to make screening decisions, which is especially important during the first half of the monitoring period, when supplemental intervention could most usefully be added to prevent or repair detected gaps.

Similarly, in terms of progress monitoring, because the measure is intended to model growth during the year, the gains between each assessment occasion will be so minimal that characterizing the success of instructional changes relative to "typical" growth becomes technically very difficult if not impossible. In concrete terms, a student in intervention might successfully improve the ability to multiply single-digit whole numbers and identify common factors, but a mixed-skill measure that includes mostly fraction operations is unlikely to detect these gains and therefore is not useful as a basis for adjusting the intervention or interpreting its effects. This problem has been noted in research studies with subsequent recommendations to collect many weeks of data before a decision can be made about whether the specific intervention is working for the student. Given that the school year generally lasts about 36 weeks, it is impractical to require 10 weeks or more of data collection, as that would only allow the intervention to be adjusted one to three times during an entire school year and would have caused students to experience perhaps 10 weeks of an ineffective intervention before the teacher could safely conclude the intervention was, in fact, ineffective for the student.

In the example below, the amount of growth that would be required for the student to reach mastery is shown by the blue line; the amount of growth required for the student to perform at least above the at-risk range is shown in orange. The first intervention did not produce sufficient growth for the student to avoid academic risk; in other words, the intervention was not effective for the student. A second intervention was effective, yet by the time an effective intervention was installed, 10 weeks of instruction had already passed and six more weeks were required to determine that the second intervention was working for the student.

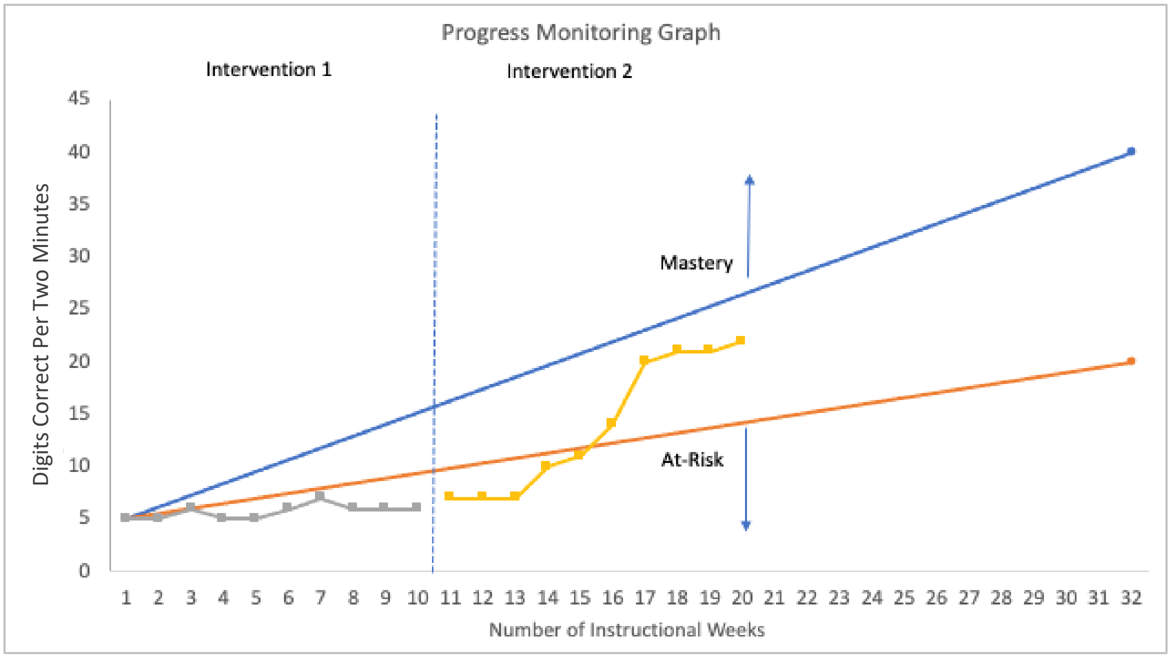


Figure 4.1. In this example, a general outcome measure is used to model growth or response to intervention. Too much time is required to determine whether the intervention has worked. General outcome measures in math are not sufficiently sensitive for progress monitoring in MTSS.

Mastery measurement, Goldilocks, and SpringMath

Given the limitations of sensitivity with math CBM, researchers began to study direct measurement of more specifically defined skills that were being taught in a known sequence (VanDerHeyden & Burns, 2008; VanDerHeyden & Burns, 2009; VanDerHeyden et al., 2017). Sampling a single skill allowed researchers to measure skill mastery with greater precision and permit more fine-tuned decision making about progress even if it changed the way progress was measured. In other words, if the right skills could be measured at the right moments of instruction, these data might be meaningful to formative assessment decisions. Such measures would have to change more frequently, resulting in a series of shorter-term linear trends. Summative metrics like rate of skill mastery could function as key decision metrics. We think of mastery measurement in math as a series of “Goldilocks” measures (i.e., the right measure at the right time) closely connected to grade-level learning in mathematics. In other words, the key is that a well-constructed, technically equivalent measure of a single skill is used when that skill is being taught to model learning from acquisition to mastery.

In the example below, a typical flow is shown for skills and instructional tactic. In the sample case here, a student begins intervention for fluency building for multiplication facts 0-12. After reaching mastery, the intervention shifts to establish a new skill, which is multiplying with fractions. After reaching the instructional range, the intervention shifts to building fluency for multiplying with fractions. After reaching mastery for multiplying with fractions, the intervention shifts yet again to establish division of fractions. These instructional adjustments can be made much more quickly than would be possible with the use of less sensitive measures (e.g., multiple skill measures as shown in the last figure). Rapid intervention adjustment allows the optimized intervention intensity for the student. Multiple-skill measures should also reflect gains but with less steep gains.

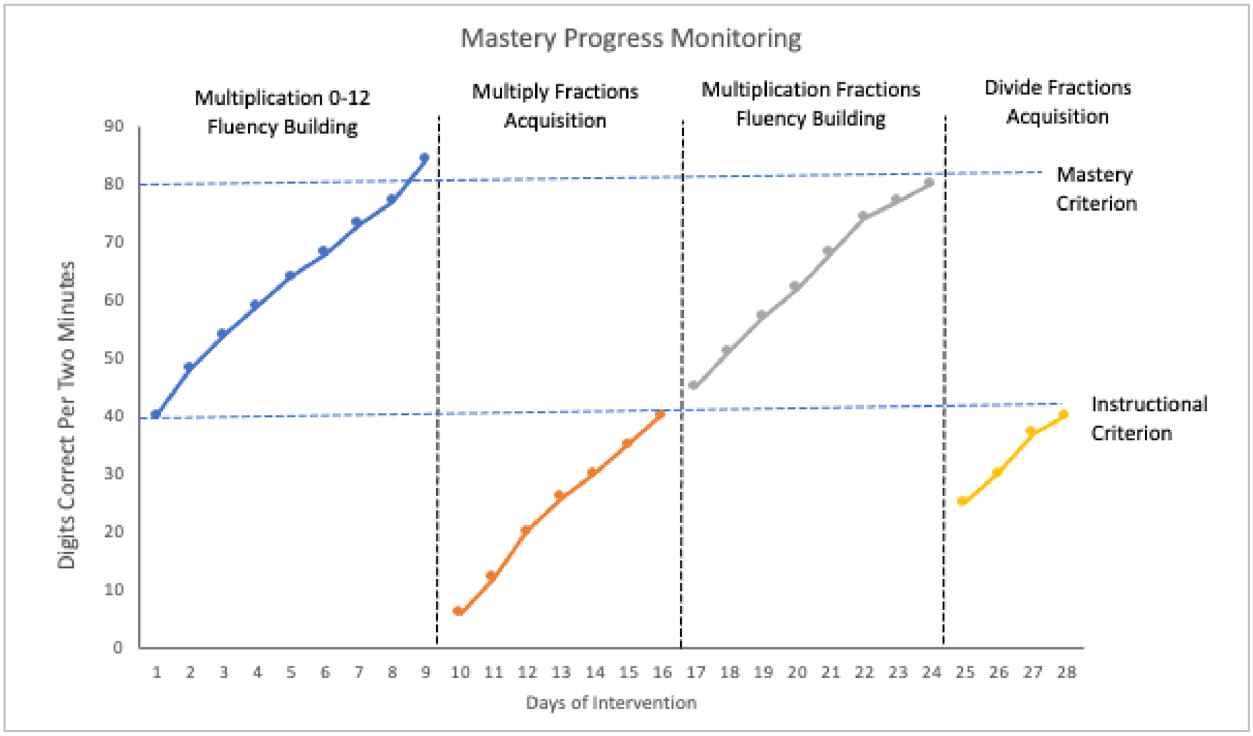


Figure 4.2. Mastery measures reflect growth during intervention much more quickly and facilitate more rapid adjustment of the intervention to promote student learning. SpringMath uses mastery measurement to monitor and adjust the intervention.

A series of recent studies has demonstrated that such an approach can meet conventional standards of technical adequacy, including reliability (VanDerHeyden et al., 2020; Solomon et al., 2022) and classification accuracy (VanDerHeyden et al., 2017; VanDerHeyden et al., 2020). This approach also provides a closer connection to instruction, permitting more sensitive feedback to the teacher about the instructional effects and thus allowing the teacher to make more rapid and fine-tuned adjustments to instruction to improve learning. For example, Burns et al. (2006) used mastery measurement in mathematics to predict skill retention and more rapid learning of more complex associated skills. This study suggested a method and a set of decision rules to determine skill mastery during instruction. In other words, teachers could assess students in the class using a two-minute classwide measure and know whether students required additional acquisition instruction, fluency-building opportunities, or advancement to more challenging content. Subsequently, such measures have been used to specify optimal dosages of instruction (Codding et al., 2016; Duhon et al., 2020), and rates of improvement have been empirically examined in large-scale reviews to characterize the dimensions of intervention conditions that can affect rates of improvement on mastery measures (e.g., dosage of intervention, skill targeted) (Solomon et al., 2020). Mastery measures also yield datasets that can be used to determine classwide and individual student risk (i.e., screening) and to evaluate programs of instruction more generally.

The most exciting news from emerging measurement research in mathematics is the possibility of using technically strong (reliable, generalizable) measurements in highly efficient ways to drive instructional changes in the classroom the next day. In contrast with norm-referenced rules that simply tell decision makers how students perform relative to other students, mastery measurement data can be used to tell us how likely a student is to thrive given specific instructional tactics. Knowing that a student performs below the 20th percentile tells us nothing about which instructional tactics are likely to benefit that student and whether the student is likely to respond to intervention when given optimal instruction. But knowing that a student performs at a level that predicts the student will remember what has been taught, will experience more robust and more efficient learning of more complex associated content, and can adapt the skill or use the skill under different task demands is highly useful to instruction. From the teacher's perspective, knowing that the student is lower performing than other students means nothing if the assessment does not help the teacher solve the problem. Mastery measurement in math is necessary for teachers to know whether instruction is working for students and what instruction is needed by their students. SpringMath uses mastery measurement to drive all decisions.

4.3 Assessment administration and scoring

All measures are provided by SpringMath when needed for certain decisions (e.g., screening, progress monitoring, intervention selection). Each time a measure is provided, an answer key is also provided to facilitate scoring. All measures in SpringMath are timed. Kindergarten measures are administered for one minute. Measures in grades 1-6 are administered for two minutes. Measures in grades 7-12 are administered for either two or four minutes.

Standardized administration directions are provided on each measure. For all measures, the assessor works the first problem on the page with students to ensure the students understand how to correctly respond, tells students to work for the entire timed interval, to work the problems in order, asks if there are any questions, and then tells students to begin working. When the time is up, the teacher asks students to stop working and collects the papers.

All measures are scored as answers correct. The digits correct criteria for decision making within the tool have been equated to answers correct criteria to simplify the scoring burden for teachers. The decision criteria for determining student skill proficiency are specific to each skill because some skills require many more digits correct to arrive at a correct answer (thus fewer answers correct would be needed to demonstrate skill proficiency). These decision rules are coded into SpringMath from the proprietary decision trees to determine risk, to identify the correct intervention, and to determine intervention success (Education Research & Consulting, 2013).

Scorers should score for answers correct according to the answer key provided by SpringMath. For 10 measures, it is possible for an answer to be correct that does not exactly match the answer key. For example, b^{-1} can also be written as $1/b$. Both answers are correct. For these measures, the standardized directions instruct the scorer to give credit for both acceptable answers (even though the answer key shows only a single answer). For all other measures, the scorer should follow the answer key exactly.

4.4 Content of measures

SpringMath includes the skills that most effectively forecast readiness for algebra success by grade 8. Across grade levels, we provide extensive coverage via assessment and intervention for Counting and Cardinality, Operations and Algebraic Thinking, Number and Operations in Base Ten, Number and Operations with Fractions, The Number System, Expressions and Equations, and Ratios and Proportional Relationships. We do not provide assessment and instruction for Measurement and Data, Geometry, and Statistics and Probability. One source of evidence for content validity is the alignment study with the Common Core standards (Read the full report at sourcewell.co/CCSSAlignment).

The following table summarizes key skills that are assessed at each grade level.

Key skills assessed at each grade level

Instructional grade level(s)	Skill strands
Kindergarten	• Object-number correspondence in counting, rote counting, number naming • Ordinal understanding of numbers • Quantity discrimination with dot sets • Changing quantities to make 10 • Adding and subtracting using numbers 0 to 5
Grade 1	• Sums to 20 • Subtraction 0-20 • Solving fact families addition & subtraction 0-20* • Quantity comparison with numbers to 20-99 and 100-999
Grade 2	• Sums to 20 • Subtraction 0-20 • Fact families add/subtract 0-20 • Quantity comparison with numbers 1,000-9,999 • Quantity comparison for sums and differences to 20 • Add and subtract 2-digit numbers using place value understanding • Add and subtract 3-digit numbers using place value understanding* • Create equivalent addition and subtraction problems using associative property, decomposition, and near-easy problems
Grade 3	• Fact families add/subtract 0-20 • Adding & subtracting 3-digit numbers using place value understanding • Multiplication 0-12* • Division 0-12* • Fact families multiplication/division 0-12* • Multiply 1-digit by 2-3-digit numbers using place value understanding • Divide 1-digit into 2-3-digit dividend using place value understanding • Compare fraction quantities for fractions with like denominators • Place fraction quantities on a number line using denominators 2, 4, 8
Grade 4	• Fact families multiplication/division 0-12 • Multiply 1-digit by 2-3-digit numbers using place value understanding • Multiply 2-digit by 2-digit numbers using place value understanding • Divide 1-digit into 2-3-digit dividend using place value understanding • Quantity comparison with decimals to the hundredths • Add and subtract with decimals to the hundredths • Place fractions on a number line with quantities > 1 and challenging denominators of 1, 2, 3, 4, 5, 6, 8, 10 • Quantity comparison with fractions with unlike denominators • Create equivalent multiplication expressions using common factors • Add and subtract mixed numbers with like denominators and regrouping • Convert fractions to decimals and decimals to fractions • Quantity comparison fractions, whole numbers, and decimals

Grade 5	• Fact families multiplication/division 0-12 • Multiply 2-digit by 2-digit numbers using place value understanding • Divide 2-digit into 3-4-digit dividend using place value understanding • Add and subtract with decimals to the hundredths • Create equivalent fractions that share least common denominator • Convert improper fractions to mixed numbers • Convert mixed numbers to improper fractions • Add & subtract fractions with unlike denominators • Quantity comparison for whole numbers, fractions, and decimals • Simplify fractions • Multiply & divide with decimals • Multiply & divide proper and improper fractions • Quantity comparison whole numbers, fractions, decimals, and percentages
Grade 6	• Order of operations (excluding exponents) • Add and subtract fractions with unlike denominators • Multiply 2-digit numbers by 2-digit numbers with decimals to the hundredths • Multiply and divide with mixed numbers • Distributive property of expression • Collect like terms • Find percent of a whole number • Add, subtract, multiply, and divide with fractions and mixed numbers • Substitute whole numbers to solve equations • Add, subtract, multiply, & divide with decimals to tenths and hundredths • Graph points in a coordinate plane • Quantity comparison with negative numbers
Grade 7	• Add, subtract, multiply, and divide with fractions and mixed numbers • Solve for missing value in two equivalent algebraic proportions • Find percent of a whole number • Solve equations with percentages • Solve missing value in a percentage problem • Quantity comparison with negative numbers • Add, subtract, multiply, and divide integers of varied sign • Order of operations • Inverse operations for addition, subtraction, multiplication, and division • Complex fractions • Convert decimals to fractions and fractions to decimals • Solve 2-step equations • Translate verbal expressions into math equations • Solve 2-step equations with fractions
Grade 8	• Add, subtract, multiply, and divide with fractions and mixed numbers • Add, subtract, multiply, and divide integers of varied sign • Distributive property to simplify expressions • Collect like terms to simplify expressions • Simplify expressions • Solve for slope & intercept using linear function • Collect, distribute, and factor to simplify exponent expressions with positive and negative exponents attached to variables, whole numbers, and fractions expressed as equations using addition, subtraction, multiplication, and division • Order of operations II • Solve for slope & intercept using linear function • Point on a line • Linear combinations to solve equations • Substitute equation to solve linear equations • Comparison method to solve linear equations

Grades 9-12	• Mixed operations • Create equivalent fractions that share least common denominator • Simplify fractions • Add and subtract fractions with unlike denominators • Multiply and divide proper and improper fractions • Convert fractions to decimals and decimals to fractions • Add and subtract with decimals to the hundredths • Multiply and divide with decimals • Multiply 2-digit by 2-digit with decimals • Order of operations (excluding exponents) • Find percent of a whole number • Algebraic proportions • Complex fractions • Add and subtract with integers • Add, subtract, multiply, and divide integers of varied sign • Solve equations with percentages • Calculate missing value in percentage problem • Translate verbal expressions into math equations • Mixed inverse operations – add, subtract, multiply, divide • Collect like terms to simplify expressions • Distributive property to simplify expressions • Add and subtract with exponents • Multiply with exponents • Divide with exponents • Order of operations II • Simplify expressions • Solve two-step equations • Solve two-step equations with fractions • Solve for slope and intercept using linear function y=mx+b • Point on a line • Linear combinations to solve equations • Substitute equation to solve linear equation • Use comparison method to solve systems of linear equations

*There is planned overlap in skill coverage between grades. All associated prerequisite skills attached to any skill are assessed through diagnostic assessment in planning intervention. We have listed the grade-level expected proficiency. In a few cases, our skills very slightly outpace Common Core standards by design. In these cases, we build proficiency first on the easier skill ranges and introduce the more challenging iteration only following mastery of the easier subskill. In each case, we introduce the challenging iteration to smooth the cross-grade level expectations or to provide greater facility in grade-level understandings that build upon the skill. For example, in grade 3, we emphasize factors to 12 (as opposed to 10) to aid in fraction work that is coming in grade 4. We first build mastery in multiplication 0-9 and then mastery in facts 10-12. In other cases, a below-grade-level skill may be listed and when that happens it provides an entry point to drill down to prerequisite skills that may be interfering with grade-level success (e.g., mixed fraction operations at grade 8). In this way, from any grade level, from screening or from classwide intervention, the child's performance and subsequent assessment can be used to place the child precisely into the right skill for intervention, even if that skill is below grade level.

4.5 Measure development

For all measures, proprietary generator parameters were written to balance problem types within and across generated problem sets (Education Research & Consulting, 2013). For example, if the measure assessed multiplication and division of proper and improper fractions, the number of multiplication problems was the same as the number of division problems, as was the number of proper and improper fractions. The stimulus range (difficulty of items) was specified in the parameter and with as few limits as possible (detailed below). Problems were stratified and then randomly generated according to the written parameters for the first round of testing. The only other limit imposed on the first draft of measures was to balance operation type within and across problem sets (e.g., problems that do and do not require regrouping if both problem types appeared in the problem set) and the range of numerical values included in the problem type.

Given a range of stimuli difficulty on the same measure, difficulty was balanced by defining easier and more challenging problems and then randomly generating from within each problem type without replacement in cycles until the problem set was filled. Multiple iterations were required when building many of the measures. For example, if a measure contained too much variation at round 1, additional limits were imposed. For these measures, problem type was more carefully defined and balanced within and across problem sets or constrained overall. This process is described in greater detail in the following section.

The number of problems included on the assessment page ranged from the mastery criterion for the skill × 1.2 (20% more problems than required to reach mastery) to mastery criterion for the skill × 1.5 (50% more problems than required to reach mastery). If the minimum number was not able to fit onto a single page, a second page was required for the assessment. Most assessments are fit onto a single page and range from eight problems to 48 problems.

4.6 Equivalence of measures (initial development)

To determine measure equivalence, problem sets were generated, and each problem within a problem set was scored for possible digits correct. The digits correct metric comes from the curriculum-based measurement literature (Deno & Mirkin, 1977) and allows for sensitive measurement of child responding. Typically, each digit that appears in the correct place value position to arrive at the correct final answer is counted as a digit correct. Generally, digits correct work is counted for all the work that occurs below the problem but does not include any work that may appear above the problem in composing or decomposing hundreds or tens, for example, when regrouping.

A standard response format was selected for all measures, which reflected the relevant responses in operational steps to arrive at a correct and complete answer. Potential digits correct was the unit of analysis that we used to test the equivalence of generated problem sets. For example, in scoring adding and subtracting fractions with unlike denominators, all digits correct in generating fractions with equivalent denominators, then the digits correct in adding or subtracting the fraction quantity, and digits correct in simplifying the final fraction were counted.

An illustration of how digits correct are computed is provided in Figure 4.3. The first problem type is addition of 2-digit numbers with regrouping. The correct answer is worth 3 possible digits correct. The second problem type is Multiplication of a 2-Digit Number by a 2-Digit Number with Decimal Placement to the Hundredths. The correct answer is worth 12 possible digits correct. The third problem type is Using the Comparison Method to Solve Linear Equations. The correct answer for this problem is worth 25 possible digits correct.

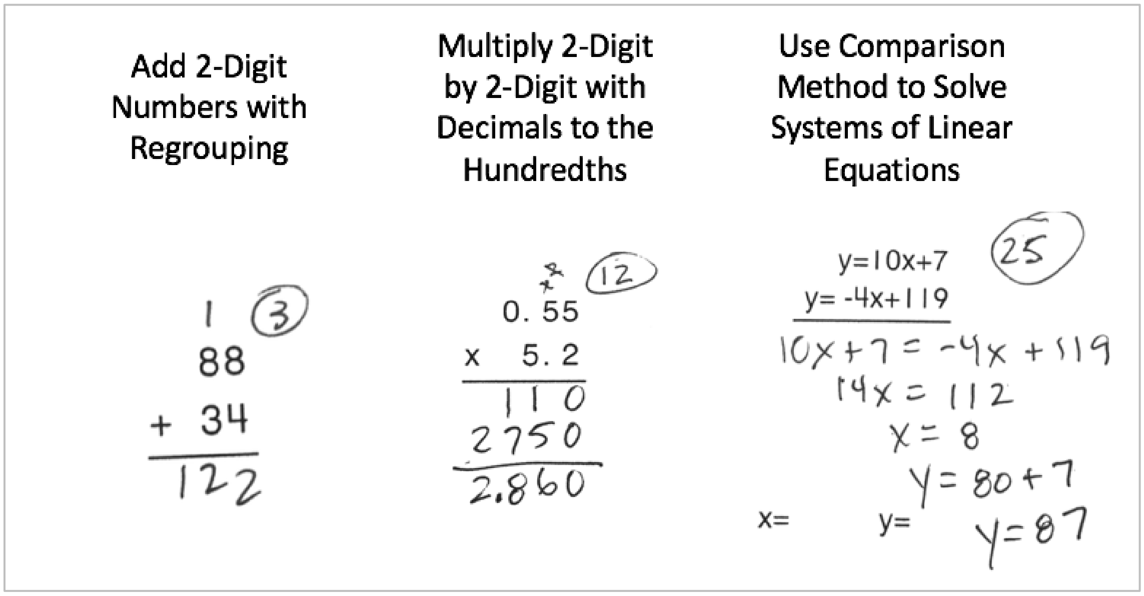


Figure 4.3*.* Answers correct are scored conventionally. Digits correct are scored for each correctly written digit typically below the problem. Digits correct are the metric originally used in curriculum-based measurement.

The number of problems generated depended upon the task difficulty of the measure. If the measure assessed an easier skill (defined as having fewer potential digits correct), then the number of problems generated was greater than the number of problems that were generated and tested for harder skills for which the possible digits correct scores were much higher. Problems generated for equivalence testing ranged from 80 problems to 480 problems per measure.

To date, more than 49,000 problems have been generated, scored for possible digits correct, and tested for equivalence. Problem sets range from eight to 48 problems. Most problem sets contained 30 problems. For each round of testing, 10 problem sets were generated per measure. The mean possible digits correct per problem was computed for each problem set. The standard deviation of possible digits correct across the 10 generated problem sets was computed and was required to be less than 10% of the mean possible digits correct to establish initial equivalence.

During initial development, 38 measures were not tested for equivalence because there was no variation in possible digits correct per problem type. These measures were all single-digit answers and included measures like Sums to 6, Subtraction 0-5, and Number Names. Eighty-three measures met equivalence standards on the first round of testing, with a standard deviation of possible digits correct per problem set that was on average 4% of the mean possible digits correct per problem. Seven measures required revision and a second round of testing. These measures included Mixed Fraction Operations, Multiply Fractions, Convert Improper to Mixed, Solve 2-Step Equations, Solve Equations with Percentages, Convert Fractions to Decimals, and Collect Like Terms. After revision and retesting, the average percent of the mean that the standard deviation represented was 4%. One measure required a third round of revision and retesting. On the third round, it met the equivalence criterion with the standard deviation representing on average 10% of the mean possible digits correct per problem across generated problem sets.

In Figure 4.4, the variation is shown for all measures. In all cases, all measure parameters met our criterion for equivalence indicating that each generated problem set was of equivalent difficulty. Equivalence is important because it allows users to assume that any change in student scores is a reflection of learning (or some other child variable) rather than a change in assessment difficulty between assessment occasions.

VanDerHeyden & Broussard (2019) reported equivalence data for the fall and winter screening measures (84 measures). The standard deviation of the mean potential digits correct per problem in a generated problem set was 4% of the mean digits correct for each specific problem type. In other words, generated problems were equivalent in difficulty according to their potential digits correct.

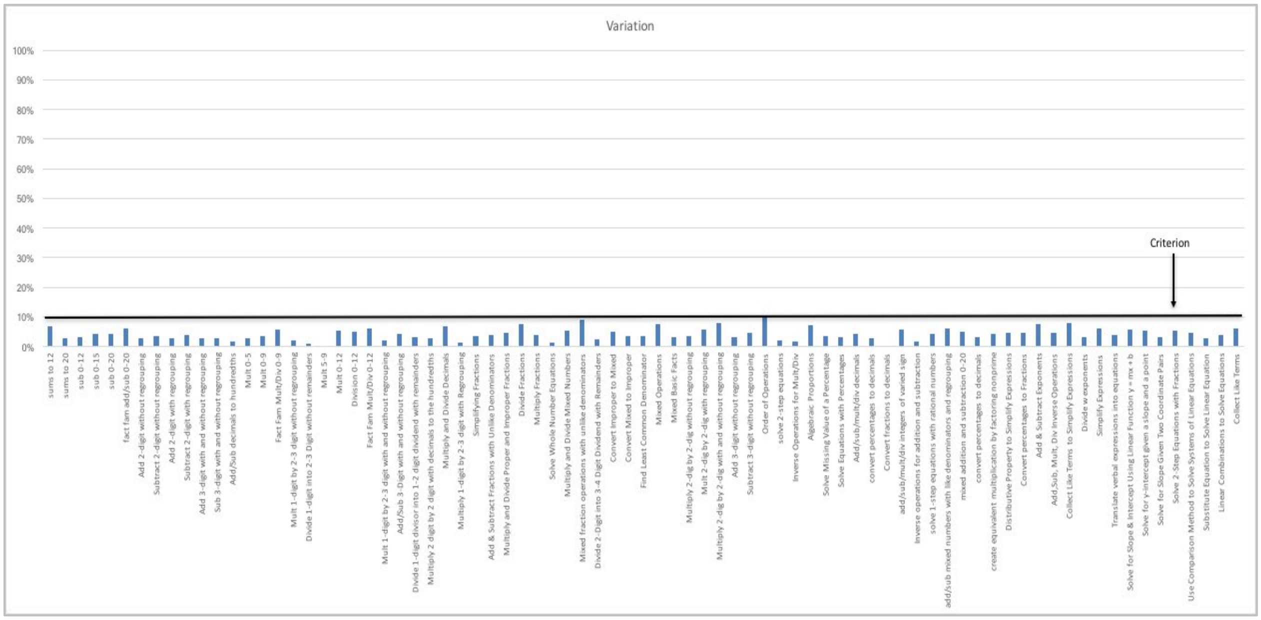


Figure 4.4. The threshold used during testing was that the standard deviation of the mean of generated problem sets for a skill was less than 10% of the mean of the problem sets examined.

4.7 Reliability and validity of scores

Once equivalence was demonstrated, SpringMath measures were tested in research studies with student participants to directly examine reliability. With rigorous research controls, one-week alternate form reliability was tested by administering two generated assessments of the same skill with one week of time between the two assessment occasions. Alternate form reliability ranged from r = 0.77 to r = 0.88 across grades and assessment occasions. Alternate form reliability was slightly higher at the winter occasion than the fall occasion. Mean reliability at fall was r = 0.81 (range, 0.77 - 0.85) and mean reliability at winter was r = 0.85 (range, 0.80 - 0.88) (VanDerHeyden & Broussard, 2019). These data support reliability of scores obtained on these measures.

Delayed alternate form reliability for screening measures in fall and winter across grades

Grade		*n*	One-Week Alternate Form Reliability
Kindergarten	Fall	86	*r* = 0.79 (0.69 - 0.86)
	Winter	79	*r* = 0.80 (0.70 - 0.86)
Grade 1	Fall	79	*r* = 0.85 (0.78 - 0.90)
	Winter	75	*r* = 0.86 (0.78 - 0.91)
Grade 3	Fall	93	*r* = 0.82 (0.74 - 0.88)
	Winter	91	*r* = 0.84 (0.77 - 0.93)
Grade 5	Fall	48	*r* = 0.77 (0.62 - 0.86)
	Winter	45	*r* = 0.87 (0.77 - 0.93)
Grade 7	Fall	41	*r* = 0.80 (0.66 - 0.89)
	Winter	38	*r* = 0.88 (0.78 - 0.94)

Inter-scorer reliability was also examined. A total of 1,564 assessments were scored by two independent scorers. Mean IOA across grade levels was 98% (range, 97%-99%). Thus, interscorer agreement was demonstrated.

Solomon et al. (2022) conducted 17 concurrent generalizability and dependability studies in a partially crossed design investigating the reliability of scores on generated SpringMath measures for students (*N* = 263) in grades K, 1, 3, 5, and 7. This study extended prior research by including novel grade levels and more rigorous math content; using generated rather than static measures; embedding a replication; examining bias by race and sex; and evaluating a simpler scoring method of answers correct compared to digits correct. As desired, most of the variance in scores was accounted for by student. Probe form effects accounted for less than 5% of the variance for 16 of 17 measures and results replicated across days. G coefficients exceeded 0.75 on the first trial for 14 of 17 measures. G studies were repeated by race, sex, and scoring metric. Overall, one to four minutes of assessment were sufficient to meet reliability thresholds, which exceeds prior findings for general outcome measures. No evidence of bias was detected, and the answers correct metric exceeded the reliability of digits correct across measures. Because SpringMath measures are not static, but rather are generated according to specific parameters, demonstrating equivalent and reliable scores on generated measures is a novel contribution to the field.

Bias analyses

Solomon et al. (2022) examined bias and found no evidence of bias in the scores (or differential reliability by subgroups). The SpringMath fall and winter screenings and classwide intervention response data have also been examined for bias and reported on the Tools Chart at NCII in 2020. In these studies a series of binary logistic regression analyses was conducted for subgroups. Scoring below 20th percentile on the year-end test in Arizona was the outcome criterion. Interaction terms were tested for each subgroup and screening scores for fall, winter, and classwide intervention. None of the interaction terms were significant at any grade level for sex, race, free or reduced lunch status, or special education status. These findings replicate all the earlier studies demonstrating screening and intervention is a more equitable basis for determining risk than teacher referral and other forms of assessment (i.e., year-end tests) alone (VanDerHeyden et al., 2003; VanDerHeyden & Witt, 2005).

This more recent research has added to a body of research examining mastery measurement in mathematics in the context of MTSS decision making conducted since 2001 by VanDerHeyden and colleagues. Because these studies were foundational to the development of SpringMath measures, findings will be detailed briefly here.

VanDerHeyden et al. (2001) reported moderate to strong reliability and validity coefficients for scores on the following kindergarten math measures, which are now used in SpringMath.

Reliability and concurrent validity coefficients for kindergarten measures

	Alternate form correlation	**Validity correlation with brigance**
Count & Circle Number	r = .84 n = 47	r = .61 n = 47
Count & Write Number	r = .81 n = 45	r = .52 n = 45
Identify Number & Draw Circles	r = .70 n = 63	r = .44 n = 63

VanDerHeyden et al. (2011) conducted a systematic replication study and again reported moderate to strong reliability and validity coefficients for scores obtained on the kindergarten measures. This study also demonstrated longitudinal correlations in the moderate to strong range.

Reliability and concurrent and predictive validity coefficients for kindergarten measures

	Alternate form correlation	**Concurrent validity correlation with TEMA**	**Predictive validity first grade CBM addition**	**Predictive validity first grade CBM subtraction**
Count & Circle Number	r = .84 n = 43	r = .61 n = 44	r = .55 n = 30	r = .55 n = 30
Count & Write Number	r = .71 n = 45	r = .63 n = 45	r = .71 n = 31	r = .51 n = 31
Identify Number & Draw Circles	r = .77 n = 45	r = .58 n = 45	r = .57 n = 31	r = .54 n = 31
Missing Number	r = .87 n = 43	r = .61 n = 43	r = .56 n = 30	r = .52 n = 30
Quantity Comparison with Dots	r = .82 n = 44	r = .41 n = 44	r = .43 n = 31	r = .43 n = 31

Several researchers have attempted to build and study mastery measures. VanDerHeyden and Burns, for example, have conducted a series of studies examining mastery measurement as a reliable, valid, and useful form of assessment for determining response to intervention in mathematics for students in grades 2-5.

Burns, VanDerHeyden, and Jiban (2006) found that fluency scores were more reliable than accuracy scores with reliability values of r = .64 for grades 2 and 3 and r = .88 for grades 4 and 5. The standard error of the slope across four weeks of progress monitoring was used to calculate the reliability of the slopes for intervention skills with reliabilities of r = .98, .99, .97, and .98 for grades 2 through 5. This study demonstrated that fluency scores on foundation tasks could be used to forecast trials to mastery and stronger slope or Rate of Improvement during intervention on subsequently more challenging and complex tasks, which was an empirical validation of the Instructional Hierarchy and powerful evidence that subskill mastery measurement could be used to drive RTI decisions.

VanDerHeyden & Burns (2008) reported a two-week alternate form reliability of r = .71 for scores on grades 2 and 3 math measures and r = .85 for grades 4 and 5. Decision criteria were tested against the Stanford Achievement Test, ninth edition, and found that in grades 2 and 3 that 34 digits correct per two minutes and 58 digits correct per two minutes in grades 4 and 5 predicted proficiency on the SAT-9, which basically replicated again the criteria set forth by Deno & Mirkin (1977).

VanDerHeyden & Burns (2009) demonstrated, yet again, that early skill proficiency forecasted mastery of more complex, related skills. Empirically derived fluency scores forecasted skill retention, again replicating the functional utility of subskill mastery measures.

4.8 Classification accuracies and decision rules

VanDerHeyden (2013) proposed a model, translated from the medical literature, of threshold decision making, and this model is the basis for classwide math intervention as a second screening gate in SpringMath. VanDerHeyden (2010, 2011, 2013) demonstrated that once prevalence reaches 50%, even very accurate screenings will not function accurately to rule out students as requiring intervention, which is the first purpose of screening.

VanDerHeyden argued for the calculation of post-test probabilities and ruling out students who have a less than 10% chance of failing the year-end test, providing classwide intervention in cases where the probability of failing the year-end test ranged from 11% to 49%, and providing individual intervention to any students with a probability of failing the year-end test greater than 50%.

VanDerHeyden, Codding, and Ryan (2017) applied the threshold model (VanDerHeyden, 2013) to mathematic screening using a variety of measures and found that the subskill mastery measures used in SpringMath outperformed other screening tools and were useful for screening.

VanDerHeyden, Broussard, and Burns (2019) examined the classification agreement values for fall and winter SpringMath measures and response to classwide math intervention as a second screening gate. They found that classwide math intervention effectively lowered the base rate of risk and permitted superior identification of risk. Use of classwide math intervention, thus, was empirically validated as an important active ingredient of SpringMath *screening*, both reducing the number of children who required individual intervention but also demonstrating superior negative post-test probabilities.

Classification agreement between SpringMath measures and scoring below the 20th percentile on the spring composite (grades kindergarten and first) or the year-end state test (grades 3, 5, and 7)

Grade	Gate	*n*	Base Rate	Sensitivity*	Specificity	Positive Post-test Probability	Negative Post-test Probability
Kindergarten	Gate 1: Fall	84	.19	100	24	34	12
	Gate 1: Winter	96	.18	100	32	38	5
	Gate 2: Classwide Intervention	95	.18	65	86	38	7
First	Gate 1: Fall	94	.22	100	15	42	7
	Gate 1: Winter	102	.21	100	21	57	6
	Gate 2: Classwide Intervention	102	.20	85	91	61	4
Third	Gate 1: Fall	86	.20	82	81	33	3
	Gate 1: Winter	96	.20	72	72	35	9
	Gate 2: Classwide Intervention	99	.20	80	80	48	6
Fifth	Gate 1: Fall	88	.20	73	79	38	5
	Gate 1: Winter	94	.20	74	53	32	10
	Gate 2: Classwide Intervention	101	.20	77	86	54	5
Seventh	Gate 1: Fall	210	.20	78	82	43	3
	Gate 1: Winter	215	.20	81	81	47	4

*Sensitivity, specificity, and post-test probability values are percentages.

In addition to research studies, classification agreement data are reported to the National Center for Intensive Intervention to undergo their technical review and rating process. These data were reported during the 2020 call for screening accuracy data and are also reported on the NCII Tool's Chart.

Criterion measures used in the analyses provided to NCII

In grades K and 1, a winter composite (for fall screening) and a spring composite (for winter screening) were the outcome criteria. For grade K, the winter composite was a researcher-constructed measure that reflected the raw score total of four timed measures, timed at one minute each. These measures were Count Objects to 20 and Write Answer, Identify Number and Draw Circles to 20, Quantity Discrimination with Dot Sets to 20, and Missing Number to 20. Thus, the composite score reflected understanding of object-number correspondence, cardinality, and ordinality. To respond correctly, children also had to be facile with identifying and writing numbers. Curriculum-based measurement of understanding of object-number correspondence, cardinality, number identification/naming, and ordinality have been studied extensively by multiple researcher teams (Floyd, Hojnoski, & Key, 2006). Among the first research teams to investigate curriculum-based measures of early numeracy were VanDerHeyden, Witt, Naquin, and Noell (2001), who studied counting objects and writing numbers and identifying numbers and drawing corresponding object sets with alternate form reliability correlations ranging from r = .7 to .84 and concurrent correlation validity evidence of r = .44 to .61 with the Comprehensive Inventory of Basic Skills, Revised (Brigance, 1999). Clarke and Shinn (2004) examined, among other measures, a missing number measure to quantify ordinal understanding among first-graders with excellent findings, reporting 26-week test-retest r = .81 and concurrent validity correlation with the Number Knowledge Test (Okamoto & Case, 1996) of r = .74. In 2011, VanDerHeyden and colleagues developed and tested new measures of early mathematical understanding, but also included the missing number measure, counting objects and writing the number, and identifying the number and drawing circles. These authors reported test-retest correlation values ranging from r = .71 to .87, correlation with Test of Early Mathematical Ability (TEMA; Ginsburg & Baroody, 2003) scores of r = .55 - to .71, and longitudinal correlation values with curriculum-based measures for addition and subtraction at the end of first grade of r = .51 to .55. The quantity comparison measure using dot sets was also examined with test-retest r = .82, concurrent validity with the TEMA r = .49, and predictive validity with year-end first-grade measures of addition and subtraction of r = .43. The spring composite was a researcher-constructed measure that reflected the raw score total of four timed measures, timed at one minute each. These measures were Change Quantity of Dots to Make 10, Missing Number to 20, Addition 0-5 for Kindergarten, and Subtraction 0-5 for Kindergarten. These measures reflected understanding of object-number correspondence to make sets of objects ranging from 1 to 10, giving a starting set of 1-10 (except never matching). This measure required the child to strike out or add to a dot set to create the specified quantity to match a number. This measure required the child to understand the number quantity desired (number identification) and then to add or remove dots to create an equivalent set. The second measure is the Missing Number measure, which assesses ordinality with excellent concurrent and predictive validity for children. The third and fourth measures assessed the child's ability to combine and take quantities to 5 using numbers.

The outcome measure in grades 3, 5, and 7 was the AzMERIT, which is the statewide achievement test in Arizona. The mean scores on the Arizona measure for participants in grades 3, 5, and 7 were in the proficient range. The base rate of nonproficiency in grade 3 was 22% versus 58% nonproficient for the state. The base rate of nonproficiency in grade 5 was 26% versus 60% nonproficient for the state. The base rate of nonproficiency in grade 7 was 32% versus 69% nonproficient for the state. We used a local 20th percentile total math standard score equivalent on the Arizona test as the reference criterion to identify students in need of more intensive intervention. At each grade, the local 20th percentile standard score equivalent was in the nonproficient range, according to the state.

The reference criterion was performing below the 20th percentile (local norm) on the researcher-constructed composite measures in grades K and 1 and Arizona year-end test of mathematics for grades 3, 5 , and 7. In our first submission to NCII (and reported in VanDerHeyden et al., 2019), we used nonparametric ROC analysis in STATA 12 using the fall and winter screening composite scores as the test variable and below the 20th percentile performance on the reference criterion as the reference variable. ROC-generated AUC values and classification agreement values were generated within the ROC analysis. Screening risk was determined in two ways: (1) on universal screening at fall and winter, and (2) response to classwide intervention, which was delivered universally and has been shown to produce learning gains and permit more accurate determination of risk when base rates of risk are high (VanDerHeyden, McLaughlin, Algina, & Snyder, 2012; VanDerHeyden 2013).

Results

Classification agreement analyses were also conducted directly using priori screening decision rules for universal screening at fall and winter and subsequent risk during classwide intervention. Results replicated the ROC-generated classification agreement indices reported here. Universal screening measures were highly sensitive but generated a high number of false-positive errors. Classwide intervention response data (also collected universally) was slightly less sensitive but much more specific than universal screening (e.g., grade 3 sensitivity = .82 and specificity = .80).

In grades K and 1, probability of year-end risk was zero for children who passed the screeners. K and first-grade children who met the risk criterion during classwide intervention had a 62% and 59% probability of year-end risk, which was 2.67 and 2.36 times the base rate of risk respectively. At grade 3, children who passed the screeners had zero probability of year-end risk. Third-graders meeting the risk criterion during classwide intervention had a 33% chance of year-end risk on the Arizona measure, which was 1.5 times the base rate of risk in the sample. Grade 5 students who passed the screener had zero probability of year-end risk but a 39% probability of year-end risk if they met the risk criterion during classwide intervention, which was 2.17 times the base rate of risk for the sample. Grade 7 students who passed the screeners had a 1% chance of year-end risk on the Arizona measure but a 44% chance of year-end risk if they met the risk criterion during classwide intervention, which was 1.63 times the base rate of risk in the sample.

Functionally, universal screening is used to determine the need for classwide intervention. Classwide intervention data can then be used as a second screening gate. For the data reported below (and also reported on the NCII Tools Chart), we recruited new samples (collected in the 2018-19 school year) and conducted new analyses using the previously ROC-identified cut scores. This time, however, for the fall screening at all grades, we used the following screening criterion: fall composite screening score below criterion and/or identified as being at risk during subsequent classwide intervention, which reflects the screening criterion we use in practice to advance students to more intensive instruction (diagnostic assessment and individualized intervention). We used a sample for which all students were exposed to both the screening measures and classwide math intervention. We report the cell values obtained for the classification accuracy metrics and regression-generated ROC AUC values reflecting the combined screening criterion. We also use the spring reference criterion for fall and winter screening accuracy analyses, which we believe to be more rigorous than using the more proximal winter criterion and more reflective of the type of decisions we are making with these data.

Fall screening classification accuracy data

Evidence	Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7
Criterion Measure	Spring Composite Score	Spring Composite Score	AzMERIT	AzMERIT	AzMERIT
Cut points - percentile rank on criterion measure	20	20	20	20	20
Cut points - performance score on criterion measure	17	59	3515	3582	3633
Cut points - corresponding performance score (numeric) on screener measure	12	39	24	35	13
Raw number of true positives	23	47	53	56	33
Raw number of false positives	41	29	48	46	12
Raw number of false negatives	6	5	2	7	7
Raw number of true negatives	77	206	194	241	158
Area under the curve (AUC)	0.80	0.90	0.91	0.86	0.90
AUC estimate’s 95% confidence interval: lower bound	0.70	0.85	0.87	0.81	0.80
AUC estimate’s 95% confidence interval: upper bound	0.90	0.94	0.95	0.90	0.99
Base rate	0.20	0.18	0.19	0.18	0.19
Overall classification rate	0.68	0.88	0.83	0.85	0.91
Sensitivity	0.79	0.90	0.96	0.89	0.83
Specificity	0.65	0.88	0.80	0.84	0.93
False positive rate	0.35	0.12	0.20	0.16	0.07
False negative rate	0.21	0.10	0.04	0.11	0.18
Positive predictive power	0.36	0.62	0.52	0.55	0.73
Negative predictive power	0.93	0.98	0.99	0.97	0.96

Classwide intervention (screening gate two) classification accuracy data

Evidence	Kindergarten	Grade 1	Grade 3	Grade 5	Grade 7
Criterion Measure	Spring Composite Score	Spring Composite Score	AzMERIT	AzMERIT	AzMERIT
Cut points -- percentile rank on criterion measure	20	20	20	20	20
Cut points -- performance score on criterion measure	29	37	3515	3582	3633
Cut points -- corresponding performance score (numeric) on screener measure	21	78	31	31	26
Raw number of true positives	20	17	52	49	34
Raw number of false positives	23	13	48	47	35
Raw number of false negatives	8	4	8	19	8
Raw number of true negatives	101	68	199	254	138
Area under the curve (AUC)	0.85	0.91	0.87	0.85	0.91
AUC estimate’s 95% confidence interval: power bound	0.78	0.86	0.82	0.80	0.87
AUC estimate’s 95% confidence interval: upper bound	0.92	0.97	0.93	0.90	0.95
Base rate	0.18	0.21	0.20	0.18	0.20
Overall classification rate	0.80	0.83	0.82	0.82	0.80
Sensitivity	0.71	0.81	0.87	0.72	0.81
Specificity	0.81	0.84	0.81	0.84	0.80
False positive rate	0.19	0.16	0.19	0.16	0.20
False negative rate	0.29	0.19	0.13	0.28	0.19
Positive predictive power	0.47	0.57	0.52	0.51	0.49
Negative predictive power	0.93	0.94	0.96	0.93	0.95

4.9 Diagnostic accuracy of targets

SpringMath involves a layering of rules that work in concert to identify and provide the level of intensity needed by students to advance learning for the individual student and more broadly for systems. The process begins with screening. Unlike many systems of assessment, SpringMath emphasizes functional benchmark criteria for decision making. In other words, specific answers correct per minute criteria are attached to each skill at each grade level to reflect frustrational, instructional, and mastery level performances. Frustrational performance is performance associated with higher error rates and poor skill retention. Instructional performance is performance reflecting a high probability of learning gains given high-quality skill practice. Mastery-level performance is performance that is likely to be retained over time, associated with faster learning of related more complex skills, and flexible, generalized skill use, including solving novel problem types.

The decision rule targets were initially based upon the best research in academic assessment, including the criteria set forth by Deno and Mirkin (1977), which were taken directly from the field of precision teaching. These criteria were empirically evaluated and validated by Burns, et al. (2006), VanDerHeyden & Burns (2008), and VanDerHeyden & Burns (2009).

Prior screening research has found that classwide math intervention improves screening accuracy and is a necessary mechanism in screening models. MTSS decision models using the criteria built into SpringMath generate decisions that are more accurate, efficient, and equitable than teacher nomination and static screening alone (VanDerHeyden et al., 2003; VanDerHeyden & Witt, 2005; VanDerHeyden et al., 2007). Teachers and other data-interpretation leaders often believe that the consequence of not using classwide intervention is simply a loss of efficiency with too many students being recommended for individual intervention. But, in fact, the wrong children are recommended for individual intervention including both false-positive (children who did not really need individual intervention) and false-negative errors (missing children who really did need individual intervention) when classwide intervention is not used as part of the screening decision.

4.10 Rationale for the necessity of timed assessments

Teachers may view timed assessment and practice as tantamount to rote memorization, but the evidence makes a case that timed assessment is an important component of instructional decision making and that timed practice is a necessary active ingredient of fluency-building intervention. Why do we rely on timed assessment in mathematics? First, it is important to use timed assessment at certain decision points because timed assessment provides superior information than does untimed performance in terms of knowing whether students have attained mastery and whether they are ready for more challenging content. If timed assessment were not necessary to make meaningful instructional decisions, then perhaps it could be avoided altogether. So first, let’s understand why we must have timed assessment.

As part of a randomized control trial of classwide math intervention (VanDerHeyden, McLaughlin, Algina, & Snyder, 2012), fourth- and fifth-graders participated in math screening in the fall (N = 209) and the spring (N = 218). Procedural controls were followed to ensure administration fidelity and interscorer agreement. Each measure was scored for digits correct per 2 minutes for the original study. Raw data were rescored for digits correct per minute and accuracy of responses (computed as the number of correct digits divided by the total number of digits attempted and multiplied by 100%). The data provided for this example come from the fact families measure for multiplication and division facts with numerals 0–12 administered at both time points. In Figure 4.5, for each of 427 administered and scored measures, the digits correct score is plotted against the accuracy score.

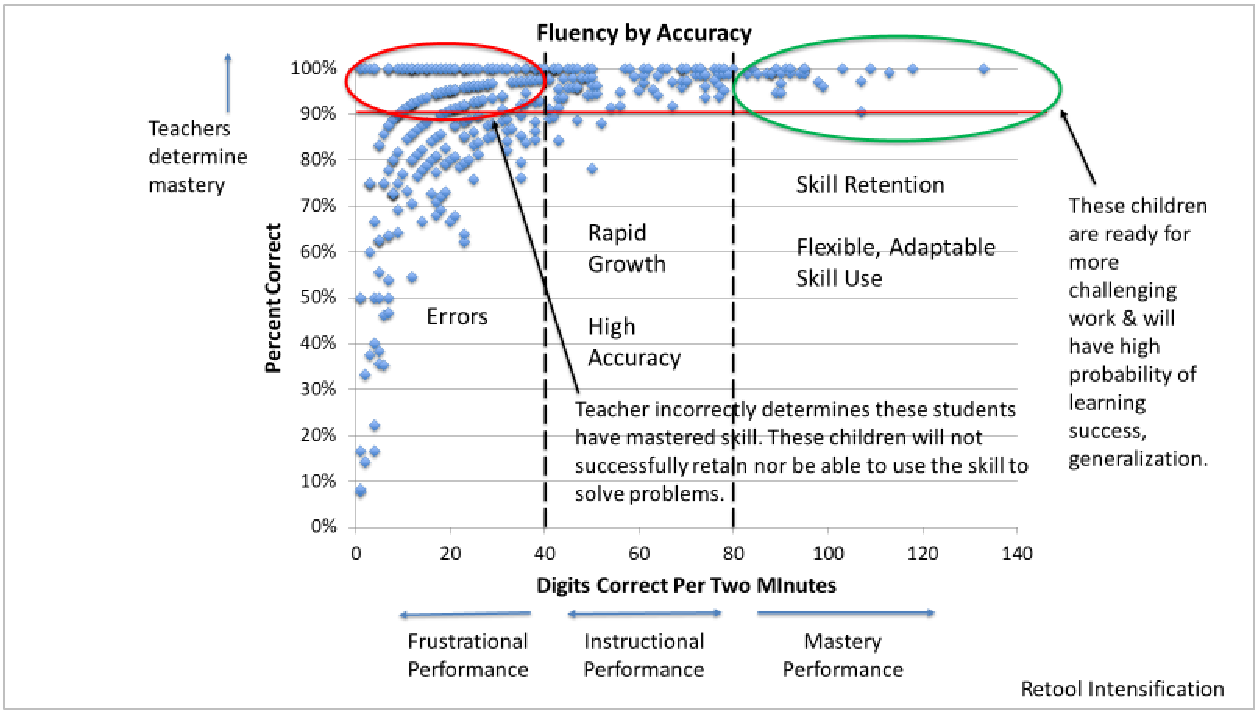


Figure 4.5. Scatterplot of accuracy scores (y-axis) plotted against fluency scores (x-axis) on a math skill.

One pattern that readers should notice right away is the natural tendency of errors to diminish as performance becomes more fluent. This pattern should resonate with readers because with oral reading fluency, students making the most errors while reading are typically those students reading at lower rates. It is a natural pattern of behavior for errors to diminish as speeded performance improves.

The level of accuracy that teachers typically require to consider students proficient might be 90%. That accuracy criterion is reflected by the horizontal line connecting to the *y*-axis at 90%. The score in digits correct per two minutes (fluency score, which is accuracy plus speed, Binder, 1996) has two criteria shown as vertical, dashed bars. The one closest to the *y*-axis is the instructional criterion and the one farther to the right is the mastery criterion (Deno & Mirkin, 1977). Functionally, we know that children who attain the instructional criterion in digits correct per two minutes are likely to be making fewer errors (i.e., they have acquired the skill) and visually we can see that is true in this graph. Students who score at or above the instructional criterion will benefit the most (i.e., grow the most rapidly) given instructional tactics that are designed to build fluency (e.g., increasing opportunities to respond, removing prompts and cues, setting goals, providing rewards, and encouraging self-monitoring of performance gains, and delayed error correction). Students scoring greater than the mastery criterion are students who are ready for generalization opportunities and more challenging problem types. We also know that students who are to the left of the mastery line and especially to the left of the instructional line are highly unlikely to retain the skill in only a few weeks, are highly likely to make errors that compromise understanding, are unlikely to be able to use the skill to solve more complex or novel problems, and will not experience faster learning of more complex related content (Burns, et al., 2006). So here is the punchline: When a teacher uses untimed assessment to judge whether students have mastered important mathematical content and understandings, the teacher will be wrong in all those cases that fall above the 90% criterion but to the left of the instructional line. This is a large number of cases (*n* = 157 errors or 55% of cases in the frustrational range) about whom the teacher would reach an incorrect conclusion and deprive students of additional needed instruction to attain skill mastery.

Accuracy metrics are not sensitive enough for decision making. Once a student reaches 100% accuracy, there is nothing more that the accuracy metric can tell you about proficiency, and yet there is valuable information still to know. If you have two students who score 100% on an addition task, but one student has to draw and count hashmarks while the other student can solve problems immediately or employ a variety of efficient strategies to arrive at the correct answer (e.g., 5 + 4 = 5 + 5 -1), the second student is more proficient and the only way to detect that superior proficiency would be to time the performance. Under timed conditions, the second student would answer more problems correctly than would the first. This truth of assessment (and the limits of accuracy on untimed measures) is exactly why college readiness batteries use timed assessment. It is the timing that separates the 30s from the 36s on the ACT, for example.

References

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163-197.

Brigance, A. (1999). *Comprehensive Inventory of Basic Skills, Revised.* North Billerica, MA: Curriculum As- sociates, Inc.

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review, 35,* 401-418.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*, 234-248.

Codding, R., VanDerHeyden, Martin, R. J., & Perrault, L. (2016). Manipulating treatment dose: Evaluating the frequency of a small group intervention targeting whole number operations. *Learning Disabilities Research & Practice, 31*, 208- 220. doi.org/10.1111/ldrp.121206

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exception Children.

Duhon, G. J., Poncy, B. C., Krawiec, C. F., Davis, R. E., Ellis-Hervey, N., & Skinner, C. H. (2020) Toward a more comprehensive evaluation of interventions: A dose-response curve analysis of an explicit timing intervention. School Psychology Review. doi.org/10.1080/2372966X.2020.1789435

Education of the Handicapped Act, Public Law No. 94–142 (1975).

Education Research & Consulting, Inc. (2013). SpringMath (version 2.2.1) Computer Software. Sourcewell. springmath.org

Floyd, R. G., Hojnoski, R., & Key, J. (2006). Preliminary evidence of the technical adequacy of the preschool numeracy indicators. School Psychology Review, 35, 627-644.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring in mathematics: A review of the literature. *The Journal of Special Education, 41*(2), 121–139.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, *57*, 488-500.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*, 1202-1242. dx.doi.org/10.3102/0034654309334431

Ginsburg, H. P., & Baroody, A. J. (2003). *Tests of early mathematics* (3rd ed.). Austin, TX: Pro-Ed.

Haughton, E. C. (1984). Standards: Refining measurement. *Journal of Precision Teaching, 4*(4), 96-99.

Keller F. S. (1968). "Good-bye, teacher...". *Journal of applied behavior analysis*, *1*(1), 79–89. doi.org/10.1901/jaba.1968.1-79

Kovaleski, J., VanDerHeyden, A. M., Runge, T., Zirkel, P., & Shapiro, E. (2022). *The RtI Approach to Evaluating Learning Disabilities, 2nd Edition*. NY: Guilford.

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., et al. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31,* 659–663.

Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development, 61,* 27-59.

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2012). Fractions: The new fronTier for theories of numerical development. *Trends in Cognitive Sciences*, 1-7.

Skinner, B. F. (1953). Some contributions of an experimental analysis of behavior to psychology as a whole. *American Psychologist*, *8*(2), 69-78.

Solomon, B. G., Poncy, B. C., Battista, C., & Campaña, K. V. (2020). A review of common rates of improvement when implementing whole-number math interventions. *School Psychology*, *35*, 353-362.

Solomon, B., G., VanDerHeyden, A. M., Solomon, E. C., Korzeniewski, E. R., Payne, L. L., Campaña, K. V., & Dillon, C. R. (2022). Mastery Measurement in Mathematics and the Goldilocks Effect. *School Psychology*.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*(8), 795–819.

VanDerHeyden, A. M. (2010). Determining early mathematical risk: Ideas for extending the research. *School Psychology Review, 39*, 196-202.

VanDerHeyden, A. M. (2011). Technical adequacy of RtI decisions. *Exceptional Children, 77*, 335-350. doi.org/10.1177/001440291107700305

VanDerHeyden, A. M. (2013). Universal screening may not be for everyone: Using a threshold model as a smarter way to determine risk. *School Psychology Review, 42,* 402-414.

VanDerHeyden, A. M**.**, & Broussard, C. (2019). Construction and Examination of Math Subskill Mastery Measures. *Assessment for Effective Intervention*. doi.org/10.1177/1534508419883947

VanDerHeyden, A. M., Broussard, C., & Burns, M. K. (2019). Classification agreement for gated screening in mathematics: Subskill mastery measurement and classwide intervention. *Assessment for Effective Intervention.* doi.org/10.1177/1534508419882484

VanDerHeyden, A. M., Broussard, C., Snyder, P., George, J., LaFleur, S. M., Williams, C. (2011). Measurement of kindergartners' understanding of early mathematical concepts. *School Psychology Review, 40,* 296-305.

VanDerHeyden, A. M., & Burns, M. K. (2008). Examination of the utility of various measures of mathematics proficiency. *Assessment for Effective Intervention, 33,* 215-224. doi.org/10.1177/1534508407313482

VanDerHeyden, A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education, 18,* 71-91. doi.org/10.1007/s10864-009-9081-x

VanDerHeyden, A. M., Burns, M. K., PelTier, C., & Codding, R. S. (2022). The Science of Math – The Importance of Mastery Measures and the Quest for a General Outcome Measure. *Communique, 51* (1).

VanDerHeyden, A. M., Codding, R., Martin, R. (2017). Relative value of common screening measures in mathematics. *School Psychology Review, 46,* 65-87. doi.org/10.17105/SPR46-1.65-87

VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. *American Education Research Journal, 49,* 1251-1284. doi.org/10.3102/0002831212462736

VanDerHeyden, A. M., & PelTier, C. (in submission). Best practices in the science of math. *Best Practices in School Psychology, 7th Edition.* Bethesda, MD: National Association of School Psychologists.

VanDerHeyden, A. M. & Witt, J. C. (2005). Quantifying the context of assessment: Capturing the effect of base rates on teacher referral and a problem-solving model of identification. *School Psychology Review, 34,* 161-183.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. A (2007). Multi-Year Evaluation of the Effects of a Response to Intervention (RTI) Model on Identification of Children for Special Education. *Journal of School Psychology, 45,* 225-256. doi.org/10.1016/j.jsp.2006.11.004

VanDerHeyden, A. M., Witt, J. C., Naquin, G., Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30,* 363-382.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32,* 204-227.

CHAPTER 5

Interventions in SpringMath

5.1 How SpringMath builds interventions

Intervention protocols in SpringMath target both procedural and conceptual knowledge. Protocols include a procedural knowledge-building script and practice activities that provide the student explicit instruction in how to use algorithms to solve problems. Explicitly teaching children the procedures that can be followed to solve problems in math is especially critical for students who struggle in mathematics (Gersten et al., 2009). Building procedural knowledge with teacher support to ensure that errors are trapped and corrected improves the child's understanding of how the numbers work and improves the child's confidence in attempting to solve problems. The second part of the intervention includes scripted activities designed to build conceptual understanding. This part of the intervention addresses why certain procedures can reliably be used to solve problems.

The scripted interventions (think of these like recipes) provided in SpringMath give the teacher a model for clear and accurate instruction of mathematics skills. The scripts were designed to follow best-available evidence in presenting new mathematical concepts to learners. For example, the scripts help teachers avoid inadvertently creating misunderstanding by using confusing language (e.g., referring to taking a 10 as borrowing, which obscures the decomposition purpose of the lesson; referring to decimals as "points," which obscures the place value understanding purpose of the lesson; using the word "reduce," which implies that the quantity changes when simplifying a fraction) (Karp, Bush, & Dougherty, 2014). SpringMath provides specific language scripted in bold-face print that the teacher can use to effectively convey the correct conceptual understandings.

Each intervention packet has two parts, one focused on procedural knowledge and one focused on building conceptual understanding. Procedural knowledge is targeted using one of the research-based standard intervention formats (e.g., cover copy compare, guided practice, response cards, timed trials) customized to the task and difficulty level appropriate for the student. Each step of the intervention is spelled out like a recipe for the teacher, and daily materials are provided to conduct the intervention each day in the packet.

The second part of the packet includes a script to build conceptual understanding. In the math K-12 math education literature, there is no clarity around what conceptual understanding means — how we define it, how we measure whether we've attained it. SpringMath conceptual understanding-building activities are designed to:

- Unpack the algorithm and proof it out for students, using the mathematical knowledge the student already has
- Demonstrate more than one way to solve problems using skills the student has already mastered, thus connecting the concept to combining, taking, multiplying, or dividing
- Provide practice discriminating quantities within the context of the skill being taught
- Provide practice converting the quantities within the problems to solve in different ways
- Provide practice solving for unknowns

SpringMath defines mastery of mathematics skills as knowing *that* a certain algorithm works *and* knowing *why* the algorithm works (Wu, 2011). Fluent use of algorithms to solve problems is both benefitted by, and in turn, benefits conceptual understanding (National Research Council, 2001; Siegler, Fazio, Bailey, & Zhou, 2012). Wu has described standard algorithms as an elegant form of shorthand that allows students to more easily and more efficiently use mathematical laws to solve problems. Wu (2011) argues,

> ...the resistance that some math educators (and therefore teachers) have to explicitly teaching children the standard algorithms may arise from not knowing the coherent structure that underlies these algorithms: The essence of all four standard algorithms is the reduction of any whole number computation to the computation of single-digit numbers. (p. 9)

Thus, mastery in mathematics implies both conceptual understanding and fluent procedural computation. SpringMath views effective fluency-building instruction as a vaccination against mathematics failure. It benefits current skill development and prevents future deficits by easing the cognitive burden for students as new content is introduced. Research by Burns, VanDerHeyden, and Jiban (2006) replicates others' research findings that students who attain fluent skill performance make fewer errors in solving math problems, are more likely to retain what they have learned, and are more likely to experience faster learning of related but more complex skills. In math, all new learning depends upon past learning. When the past learning is fluent for the student, the student can think about the new content in more expansive ways. There are many analogies that can be used here, for example, the idea of a musician practicing scales. In order to create new music and play new music, the musician must master the scales. Subskill fluency helps the student think about and master solving problems in multiple ways, explain why a given solution works, and explain how the concept is connected to other mathematical concepts and skills. This is distinct from a halting, incomplete explanation or one that simply repeats the algorithm. Algorithms that are not understood will be forgotten (Ma, 1999).

SpringMath emphasizes fluency-building intervention for the classwide intervention, building upon and supplementing the acquisition instruction that has occurred in the classroom for below-grade and current-grade-level content. The skill sequences for classwide intervention are static and increase in difficulty such that success on the early skills positions students for success on subsequent skills. For individual intervention, SpringMath drills down to verify mastery of prerequisite skills to determine initial skill placement and type of intervention needed. Individual intervention is more intensive instruction that is individualized based upon the student's measured mastery of prerequisite understandings. All individual intervention protocols (acquisition and fluency building) include scripted activities to build procedural knowledge and conceptual understanding. Intervention success criteria reflect skill mastery.

Once intervention is initiated, intervention protocols systematically target development of three or more of the seven major areas of understanding for mathematical proficiency for all skills. The philosophy of SpringMath is not to shy away from challenging problems. The progression of concrete-representational-abstract content is apparent in the intervention scripts within and across grades as children gain understanding. However, SpringMath is grounded in the philosophy that children build mathematical proficiency best not by constraining their work to only simple problems that can be visualized and solved using mental math, but rather by providing challenging content and support for students to master the procedural problem solving and conceptual understanding to understand why certain procedures work no matter how challenging the problem. Supporting students to gain conceptual understanding within progressively more challenging content gives students a more robust mathematical skill set and the confidence to use what they know about combining, taking, multiplying, or dividing to solve; to convert quantities within a problem to allow for a different solution path; to use what they know about setting up problems to solve for an unknown; and to anticipate the resulting quantities as they are problem solving so that they may trap their own errors and adjust their problem solving as they go.

5.2 Intervention skill targets

The goal of SpringMath is to ensure children have mastered the skills needed to thrive in algebra by grade 8. SpringMath has about 145 assessments that map onto one or more of the seven major areas of mathematical proficiency from kindergarten through eighth grade. A full list of skills can be found on the SpringMath website at springmath.org (sourcewell.co/SpringMathAssessments). We think of these as “tool” skills that equip the child for mastery of mathematical thinking and problem solving (Johnson & Street, 2013). Tool skills are those skills that, if mastered, allow a child to rapidly learn new content or understandings using the math thinking skills that they have developed.

SpringMath assesses and provides intervention for grade-level skills that are pertinent to students’ mathematical development in seven major areas, (a) Combining Rational Numbers and Variables, (b) Taking Rational Numbers and Variables, (c) Multiplicative Reasoning, (d) Proportional Reasoning, (e) Quantity Comparison, Ordinal Position, and Place Value, (f) Solving for Unknowns, and (g) Creating Equivalent Quantities. These seven areas cut across multiple operations and provide the foundation from which a child can question, speculate, reason, solve, and explain real-world problems in mathematics.

Combining Rational Numbers and Variables (Adding)
Understanding how to combine quantities given different quantity representations (e.g., dots, whole numbers, variables, negative numbers) and being able to anticipate the approximate resulting quantity.

Taking Rational Numbers and Variables (Subtracting)
Understanding how to remove or take quantities given different quantity representations and being able to anticipate the approximate resulting quantity.

Quantity Comparison, Ordinal Position, and Place Value
Understanding quantity and ordinal position given different quantity representations.

Multiplicative Reasoning
Understanding how to combine multiple quantity sets given different quantity representations. Includes understanding of distributive property.

Proportional Reasoning
Understanding how to combine and factor with quantities using a base unit that is less than 1. Includes collecting terms and simplifying expressions.

Solving for Unknowns
Understanding how to construct equations and use properties of operations to solve for unknown quantities.

Creating Equivalent Quantities
Flexible conversion of quantities using different quantity representations.

All children in grades K-2 are assessed using multiple measures in all of these areas of understanding except multiplicative and proportional reasoning. Children in grades 3-6 are assessed using multiple measures in all areas. Students in grades 7-8 are assessed in all areas except quantity comparison, ordinal position, and place value. When a child performs in the risk range on a measure associated with one or more areas of understanding, the child participates in a diagnostic assessment process to identify the prerequisite understandings that need to be established to permit the child to master grade-level expectations. What this means is that the assessment process is designed, if needed, to go all the way back to the root skill, which may represent below-grade-level content. This is how the areas of understanding create coherence among the measures such that they are not just 145 disparate and disconnected skills.

In the space below, we will define each area of understanding emphasized within SpringMath assessments and interventions.

Combining Rational Numbers and Variables (Adding)

The ability to combine quantities is a tool skill that allows the child to sum quantities and later to identify like terms and condense and simplify terms within equations to create equivalent quantities and to solve for unknown quantities. With younger students, scripted, game-like activities are provided to give children ample experience comparing and changing quantities by combining. Manipulatives, symbols, number lines, and number grids are used to support the child's development of this understanding. Children are guided to create equivalent quantities using manipulatives, number lines, and mathematical expressions. For example, first-graders are directed to, "Show me three ways to make 5" and are guided to use what they know about combining quantities to make 5 (counting objects, counting on a number line, drawing hash marks, writing mathematical expressions like 1 + 1 + 1 + 1 + 1). Games like "war" are used to support students' development of quantity understanding for sums. From kindergarten up, children are guided to solve problems involving "How many more?" to connect combining and taking as inverse operations that are useful in solving real-world problems. Combining quantities using place value understanding is developed with scripted activities and experience making tens and hundreds, using mental math to create and solve easier problems (creating equivalent quantities) and using expanded notation to solve addition. Combining rational numbers and variables also helps the student master understanding of the Identity, Commutative, and Associative properties.

Taking Rational Numbers and Variables (Subtracting)

The ability to take quantities is a tool skill that allows the child to subtract and later condense and simplify terms within equations. With younger students, scripted game-like activities are provided to give children experience comparing quantities and changing quantities by subtracting. Manipulatives, symbols, number lines, and number grids are used to support the child's development of this understanding. Subtraction is taught as solving for an unknown addend. Understanding of subtracting as being the inverse of adding is specifically cultivated from kindergarten forward, using developmentally appropriate activities. For example, young children are guided to anticipate quantities based on combining and taking, to use taking to "undo" combining, and to fluidly use the conceptual relationship of addition and subtraction to solve fact family problems and to create equivalent expressions using addition and subtraction. As children subtract, this deepens their understanding of place value. Children are given guided support to take tens and hundreds, to convert challenging subtraction problems to easier problems using what they understand about adding and subtracting, and to decompose tens and hundreds to take quantities. Games like "war" are used to support students' development of quantity understanding and their ability to anticipate resulting quantities when subtracting (e.g., when comparing 20 – 1 and 10 – 1, a child can rapidly identify that the first difference will be a greater quantity because the starting quantity is greater and the same quantity is being taken). In the later grades, children are guided to take quantities with more complex number expressions (e.g., subtracting with decimals, exponents, and negative numbers), and to identify and take like terms to simplify and solve equations.

Multiplicative Reasoning

Multiplicative reasoning gives the student the ability to work with greater quantities more efficiently. Beginning with simple multiplication, which is taught as repeated addition, the child can learn the value of finding greater quantities more efficiently. Before multiplication is introduced, children learn to add and subtract using place value understandings, and this work is foundational to understanding base ten principles and the rate of composition when moving from right to left in place value positions. The use of place value properties in multiplication expands the potential for working with greater quantities. Factoring provides the foundation for working with proportions in different base units. In the higher grades, working with exponents and using multiplicative reasoning to distribute terms allows for more complex problem solving. Mastery of multiplication facts through factors of 12 is expected (and taught) in grade 3 in SpringMath. This expectation represents a slight departure from some state frameworks and the Common Core standards for which children are only expected to master facts 0-10 by grade 3. SpringMath requires 0-12 because multiplicative reasoning is the basis for proportional understanding, and mastery of the "11" and "12" facts makes available additional known factors that can be used to create equivalent proportion quantities or to manipulate proportions, which is a key part of grade 4-6 math instruction. In SpringMath, Division 0-12 is also taught at grade 3, and division is conceptually taught as finding an unknown factor. Students are supported to understand the conceptual relationship of multiplying and dividing through solving fact family problems, creating equivalent expressions using multiplication and division, solving multi-digit multiplication problems using place value understandings (i.e., the rate of composition or decomposition is multiplicative moving right to left and divided moving left to right), working with inverse operations, and solving one-step equations. Simple distributing and collecting of terms are introduced in grade 6 following mastery of multiplication facts, multi-digit multiplication, multiplication with decimals, and identifying common factors and multiples. Students are provided guided support to convert quantities using multiplicative reasoning (e.g., decimals, percentages). Mastery of multiplicative reasoning establishes the child's understanding of the Identify, Commutative, and Distributive properties of multiplication. These are the laws that allow for more complex problem solving, but also allow students to understand why certain algorithms work. For example, when children can work through a proof using what they know about division of whole numbers and the Identify Property of multiplication to see why the invert and multiply tactic in dividing fractions works, it becomes a useful algorithm rather than a trick that will soon be forgotten. It also provides a foundation for solving more complex equations that require inverse operations with fractions.

Proportional Reasoning

Beginning in earnest in grade 4 and continuing through grade 6, understanding of fractions becomes the next critical milestone that must be achieved to open the door to higher-level mathematics understanding. Siegler et al. (2012) found,

> "Poor fraction knowledge in elementary school predicts low mathematics achievement and algebra knowledge in high school, even after controlling for general cognitive abilities, knowledge of whole number arithmetic, and family education and income. High school algebra

teachers recognize this relation; they rank students' fraction knowledge as among the largest impediments to success in their course." (p. 6)

One common feature of successful fraction knowledge interventions is helping students understand the quantity that fractions represent, how these quantities map onto a number line, and how these quantities are subject to the same mathematical laws that govern whole number operations when those operations are applied to fractions (Fuchs et al., 2013; Wu, 2009). The introduction of fractions in mathematics is a critical window that can separate children destined for long-term mathematics success from children who will experience poor math outcomes. Typical children (meaning students who are not at risk) require sustained and supported experience working with fractions to master working with base units that differ from "1." Place value understanding and locating quantities on number lines are emphasized from grade 1 forward with whole numbers so that children can become facile thinking in base units other than one before they begin working with fractions. When children are taught multi-digit addition and subtraction using just the standard algorithm, they will tend to think of the entire unit as ones. Explicit instruction is required to assist the learner to think of quantities in hundreds, tens, and ones by helping the child take apart and combine quantities to convert, to make problems simpler, and to find solutions. Connecting quantities to the number line and helping the child become confident and accustomed to representing equivalent quantities in multiple ways will help avoid the "cliff of conceptual misunderstanding" that occurs so often when fractions are introduced, especially for struggling learners.

Fraction quantities and operations are very confusing to struggling students because suddenly concepts they learned to be true with whole numbers are no longer correct with fractions. For example, numbers lower in the counting sequence could represent greater quantities for the first time in a child's math learning when working with fractions. When dividing fraction quantities, it's possible to get a quotient that is greater than the dividend. Similarly, when multiplying with fractions, it's possible for the product to be lesser than the factors. Fractions have been described as turning what students have learned with whole numbers "upside down" (VanDerHeyden & Allsopp, 2014). Graphics, number line demonstrations, and scripted lessons for conceptual understanding are used to demystify fraction quantities, to make explicit how fraction quantities can be identified and manipulated (for example, when creating fractions with common denominators), and what to expect when conducting operations with fractions. Understanding of fractions is explicitly connected to understanding of whole number quantities on a number line, and children are explicitly taught that for any fraction (e.g., 3/n), the base unit is 1/n and the quantity can be mapped and counted at 1/n + 1/n + 1/n. Fractions are taught as a special case of division, and this lesson is repeated when children work with order of operations and simplifying expressions and holding a division operation as a fraction quantity might make the problem much easier to solve.

Proportional understanding is broader than "just" fractions, however. Proportional reasoning is connected to multiplicative reasoning with division being taught as finding an unknown factor, which is the basis of factoring and which will be critical to understand how and why certain operations with fractions work (e.g., finding a least common denominator, dividing with fractions). Proportional reasoning begins in earnest in grade 3 with division concepts and fraction work, continues into grade 4 with more complex division operations and more fraction work. When children work with multi-digit operations and especially decimals, they master the rate of decomposition in moving from left to right in place value positions. In grade 6, students master distributive property of expression, collection of like terms, order of operations where division operations might be represented as a fraction quantity, and finding percent of a whole number.

In grades 7 and 8, students use proportions to solve for unknown quantities (e.g., solving algebraic proportions, working with percentage problems, solving complex fractions), and solving linear equations. Students are provided guided support to connect solving for proportions to practical problems for rate and ratio. Flexibility with proportions is explicitly developed. For example, specific scripted activities guide students to determine when to use a decimal or a fraction to solve a particular problem, and practice is provided representing division as a fraction when applying order of operations problems based on problem demands.

Quantity Comparison, Ordinal Position, and Place Value

The ability to rapidly compare quantities assists the student to anticipate correct answers when solving problems. When a child has a general anticipation of the range within which an answer should fall, the child is able to trap his or her own calculation errors and be more confident solving applied problems. Quantity comparison assessments are used in grades K-6, including the number quantities and symbols that students will commonly encounter at each grade level in solving problems. For intervention, number quantity understandings are supported with the use of manipulatives when appropriate, place value strategies, and number line demonstrations and activities. When working with multi-digit operations, understanding of place value is emphasized as a means to rapidly understand and compare quantities with specific assessments and interventions using expanded notation, practice “taking” and “making” tens and hundreds in the early grades beginning with concrete (e.g., counting object sets with first-graders) and progressing to representational (composing and decomposing higher-value units to add and subtract) to abstract (working with decimal quantities, converting decimal quantities to fractions, determining when to use a fraction or a decimal). Place value understanding in the older grades is supported with multiplicative reasoning with exponents and in converting between decimal units, mixed numbers, and percentages. Fluency with integer quantities is taught starting at grade 6.

Solving for Unknowns

All math problems can be thought of as solving for an unknown variable (quantity, operation). We can imagine very simple operations such as 3 + 5 = ? where the sum is the “unknown.” In SpringMath we use the four operations and delete one of the numbers in the problem so the child must conceptually understand the relationship between the operations to solve for the missing number. This understanding paves the way for understanding how and why inverse operations work in the later grades. In the middle grades, finding shared factors to solve fraction operations, solving one- and two-step equations, and solving linear equations are more complex skills that provide the child practice solving for unknowns. Eventually, this type of mathematical reasoning allows the child to engage in complex and practical problem solving. Thus, solving for unknowns is a tool skill for mathematical reasoning and problem solving. Across grades, beginning with kindergarten, specific assessments and interventions assist the child to solve for an unknown. In kindergarten, a student is asked to change existing dot quantities to specific dot quantities ranging from 1 to 10 by adding or removing dots. In grades 1-4, fact families are used to facilitate student proficiency and confidence in solving for an unknown. Applied problems include solving for remaining time and units of measurement so children are well acquainted with the practical value of this skill. In the middle to upper grades, solving for unknowns is reflected in solving equations, creating equivalent expressions, distributing and collecting terms, simplifying expressions, and solving systems of linear equations.

Creating Equivalent Quantities

Creating equivalent quantities is another skill that unlocks the door for students to turn difficult problems into easier problems. This tool skill will serve the student for the rest of his or her formal schooling career, but also in life. For example, when adults make change, they commonly use strategies (sometimes referred to as “mental math”) to convert one quantity to near-dollar or 10-cent value and then adjust the answer by the remaining ones (e.g., a $21 item is reduced by $13 -- $21.00 - $10 = $11 minus $3 to reach $8). When a child encounters a challenging problem, the ability to create equivalent quantities helps the child use what he or she knows to solve the problem. For example, if the child is asked to give the decimal equivalent for 5/8, the child can anticipate that it will be greater than 0.5 because 5/8 is greater than 4/8, or 1/2. In fact, using what the child understands about fractions and combining quantities, the child can solve as 4/8 + 1/8 and then can reason the answer will be 0.5 + 1/8. To figure out the decimal equivalent of 1/8, the child can reason that 2/8 is the same as 1/4, which is a “known” or “easy” decimal quantity of .25. Because 1/8 is half of 2/8, then the child can reason that 1/8 is equal to 1/2 of .25 or .125. So, the child knows that 5/8 is equivalent to 0.5 + .125, or .625. Reasoning to the solution in this way is easier than dividing 5 by 8. Across grades, the ability to create equivalent quantities is emphasized using grade-level skills and scripted activities. Within interventions, conceptual understanding activities provide practice creating equivalent quantities. Number quantity understandings are supported with the use of manipulatives when appropriate, place value strategies (e.g., expanded notation), and number line demonstrations and activities. Grade-appropriate assessment and practice are provided converting quantities in all grades, beginning with drawing circles to match a given number quantity in kindergarten to simplifying expressions in grade 8, for example.

5.3 Intervention features

Beyond the skills targeted, SpringMath interventions were engineered to include instructional features that produce strong learning gains. We conducted an audit of intervention protocols in 2020. Acquisition protocols and fluency-building protocols were randomly selected for 46 of 143 skills (32%) and were audited for certain features, some of which are summarized below. Protocols came from grades 1-6. All interventions provided scripted instruction for conceptual understanding and procedural instruction, interleaved daily; multiple representations of mathematical concepts and/or equations; explicit and systematic instruction, including modeling, verbalization of thought processes, guided practice, corrective feedback, and frequent cumulative review. In addition, interventions sampled contained the following features in scripted form with all practice problems provided in the weekly intervention packet. Practice problems are generated using our assessment generator to provide technically equivalent (similarly difficult) problems that differ each time the intervention protocol is generated by SpringMath.

Intervention features

Intervention features	Percentage of interventions using this feature
Explicit proofing of algorithm	100%
More than one way to solve	100%
Connection to previous understanding	100%
Connection to future understanding	100%
A model of correct responding	100%
Immediate corrective feedback	100%
Practice discriminating quantity	98%
Converting quantities	100%
Constructing and/or solving for an unknown	96%
Create an equivalent quantity	100%
Fill in the missing number	95%
Detect, repair, explain errors	13% in conceptual understanding activities; 100% in procedural skill building
Conceptual practice materials provided?	100%
A game	57% (mostly for fluency-building protocols)
Manipulatives	28%
Graphic representation	85%
Asks child to articulate a rule or pattern, estimate quantity based on understanding of operations and starting quantities, justify solution	100%
Word problem	51% (used in fluency-building protocols)
Solve a more challenging problem type	78%
Goals for improvement with reward system	100%

SpringMath interventions are designed to include dimensions of explicit instruction

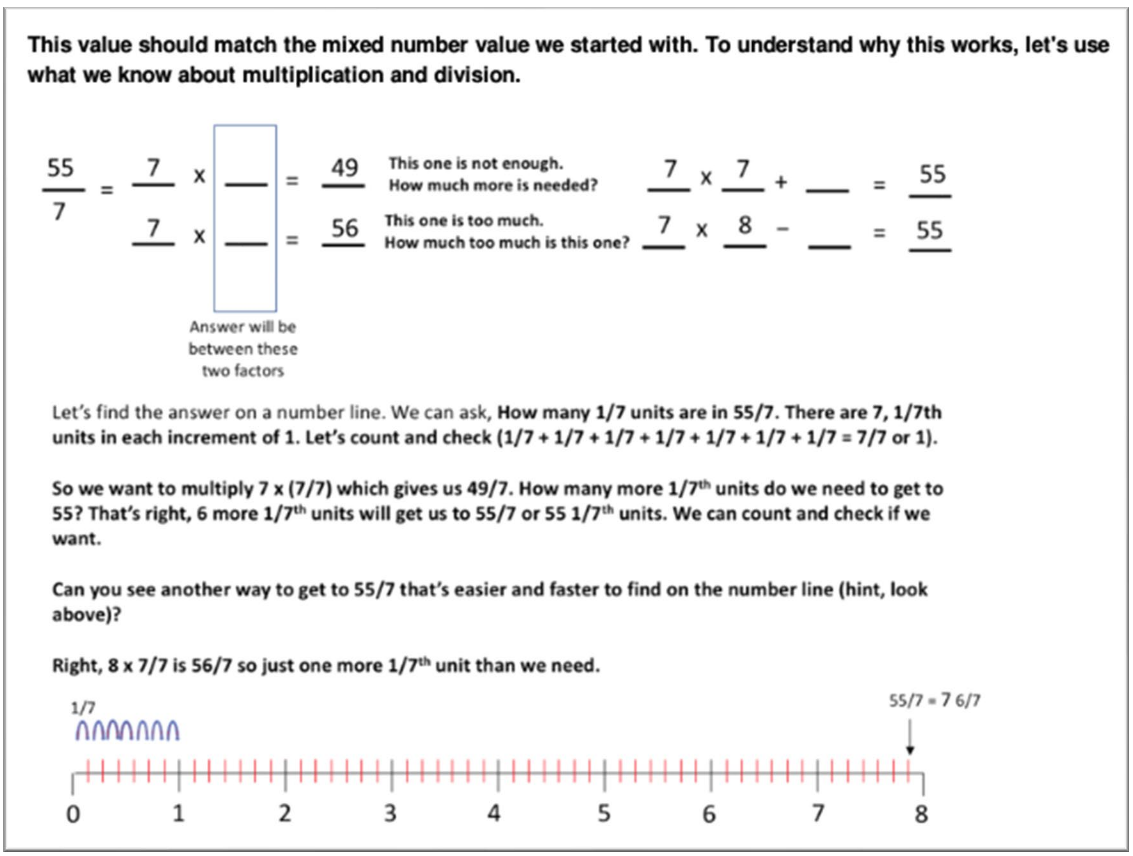
SpringMath interventions were also designed to include the following elements of explicit instruction during procedural and conceptual skill building.

SpringMath provides directions in clear, direct language

Intervention scripts are provided in printable weekly packets. Teacher language is provided in boldface print, step by step, to explain how to solve a particular problem. An error correction script also is provided.

SpringMath models efficient solution strategies

We model (and script for the teacher) correct problem solving prior to assessment and during the procedural skill-building portion of daily intervention. Procedural interventions include cover/copy/compare and guided practice intervention types with well-controlled novel practice problems generated within each protocol pack for the week. We also provide scripted and modeled activities to build conceptual understanding in the conceptual understanding part of the daily intervention. For example, the teacher follows this script as part of the Guided Practice Intervention for converting improper fractions to mixed numbers shown in Figure 5.1.



This value should match the mixed number value we started with. To understand why this works, let's use what we know about multiplication and division.

Let's find the answer on a number line. We can ask, **How many 1/7 units are in 55/7. There are 7, 1/7th units in each increment of 1. Let's count and check (1/7 + 1/7 + 1/7 + 1/7 + 1/7 + 1/7 + 1/7 = 7/7 or 1).**

So we want to multiply 7 x (7/7) which gives us 49/7. How many more 1/7th units do we need to get to 55? That's right, 6 more 1/7th units will get us to 55/7 or 55 1/7th units. We can count and check if we want.

Can you see another way to get to 55/7 that's easier and faster to find on the number line (hint, look above)?

Right, 8 x 7/7 is 56/7 so just one more 1/7th unit than we need.

Figure 5.1. Teacher script to guide the student to develop understanding of converting improper fractions to mixed numbers.

SpringMath ensures that students have adequate background knowledge and skills

Diagnostic assessment is conducted to verify mastery of prerequisite skills. If needed, prerequisite skills are targeted with fluency-building or acquisition instruction. All interventions explicitly connect the targeted skill to skills and understandings that the student already has mastered (e.g., multiplication is taught as repeated addition and mapped onto a number line to make the quantity and operation discriminable).

SpringMath gradually fades support for correct execution of strategies

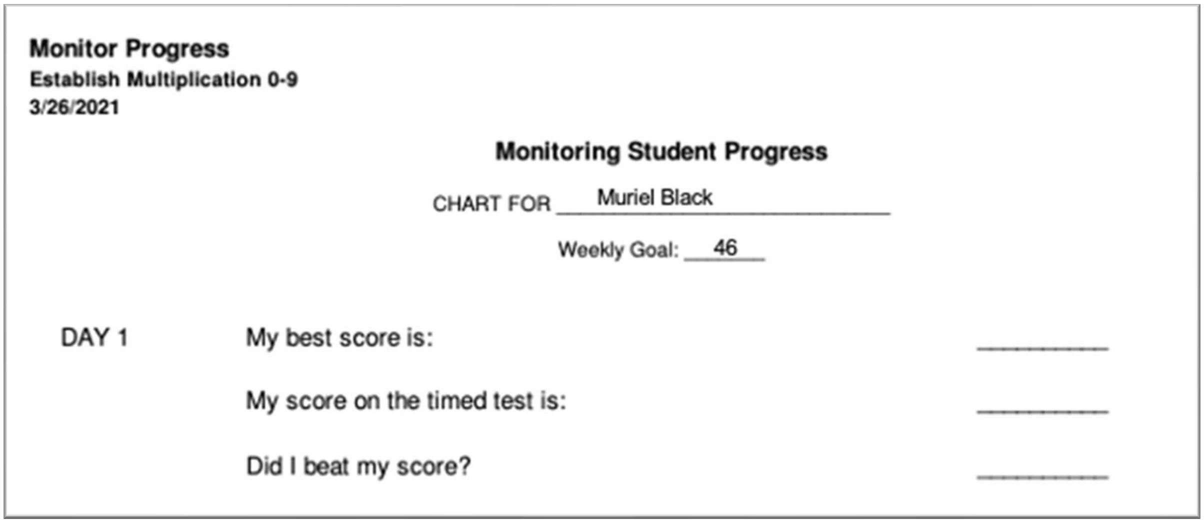
The teacher models correct responding, provide guided practice with immediate corrective feedback, and then provide an interval of independent practice with delayed error correction in the procedural skill portion of the lesson. The conceptual understanding portion of the lesson is scripted and includes visual supports (number lines), practice creating equivalent quantities, and solving for missing or unknown quantities using modeling and guided practice that requires the student to think aloud, make estimations, and describe observed patterns. Task difficulty is advanced in increments based on student learning gains. For example, if sums to 20 is the terminal skill, sums to 6 and sums to 12 are taught to mastery before introducing sums to 20.

SpringMath provides adequate practice opportunities

Practice problems are generated to include problems reflecting a particular skill and difficulty level. We generate five days of technically equivalent practice problems representing 120% of the mastery criterion for the practice interval plus the follow-up progress-monitoring assessment in each weekly intervention packet. All conceptual understanding activities are scripted. Students are typically asked to complete number line activities, solve word problems, complete area models, fill in missing values, create equivalent expressions or quantities, state whether a quantity is greater or lesser, and detect and repair errors. Word problems and games are provided in the fluency-building intervention protocols and also can be accessed in Support.

SpringMath includes behavioral support features

Diagnostic assessment is used to verify mastery of prerequisite understandings and determine the need for acquisition support versus fluency-building support such that a student is placed into the correctly aligned intervention skill (task difficulty) and receives the aligned instructional support according to their needs. This makes the learning experience productive, engaging, and rewarding to the student. Weekly progress monitoring and data-driven intervention adjustments ensure that the intervention content and tactic remain well aligned with the student's needs. Intervention scripts provide the goal to the student ("Remember your score from last time. Your goal today is to beat your score! Remember, your brain is like a muscle. You just worked your math muscle. Now let's see how much stronger you are getting!"). A self-monitoring chart is completed by each student each day during individual intervention, and game-like self-monitoring charts are available in Support for tracking gains during classwide intervention (see Figure 5.2).


Monitor Progress
Establish Multiplication 0-9
3/26/2021

Monitoring Student Progress

CHART FOR Muriel Black

Weekly Goal: 46

DAY 1	My best score is:	______
	My score on the timed test is:	______
	Did I beat my score?	______

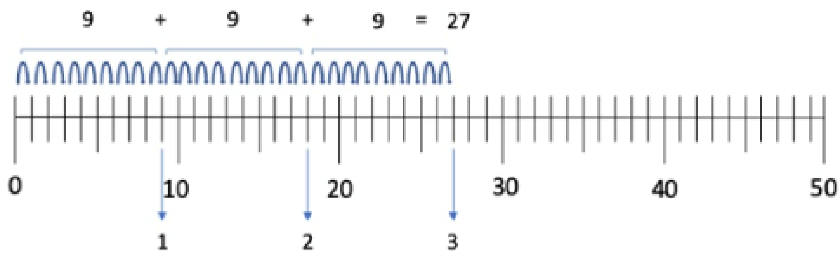
Figure 5.2. An example of one of the progress-monitoring charts used by students to set goals, self-monitor, and celebrate progress.

Daily improvement goals are set, and the student earns rewards each day for performance improvements. In addition, weekly progress monitoring is charted and rewarded with celebrations, small privileges, or tangible rewards for weekly gains. The coach dashboard tracks intervention use and effects throughout the school. Where implementation integrity (dosage and effect) is weak, SpringMath alerts the coach to visit the classroom for in-class coaching support to facilitate correct intervention to promote student learning gains and student motivation for continued growth.

SpringMath teaches for transfer of learning or generalization

We explicitly teach for transfer using multiple representations, practice creating equivalent quantities within the context of the skill being taught, and embedding a high dosage of opportunities to respond to build fluency and facilitate generalization. For example, all intervention scripts show multiple ways to solve problems using mathematical understanding the student has already mastered and provide multiple opportunities for the child to practice solving with immediate corrective feedback until the student is accurate. During acquisition, providing multiple representations helps students connect understanding of new skills to already-mastered understanding/skills, which builds confidence, reduces task difficulty, and facilitates acquisition. Providing multiple representations reduces task difficulty/complexity, improves student engagement, and increases likelihood of correct responding when new skills are introduced, but also cultivates robust learning (i.e., learning that is more likely to be generalized to new stimulus conditions) because the task/stimulus similarities for which the learned solution can be successful are made explicit such that the student can readily recognize the stimulus conditions under which learned skill can be useful. For example, as shown in the script in Figure 5.3, whole number multiplication is introduced as repeated addition, mapped on a number line, and solved by counting on a grid.

Let's find the solution to 3 x 9 on a number line. We can think of 3 x 9 as 3 sets of 9. Let's count out each set and then find the product or answer of 3 x 9.



9 + 9 + 9 = 27

How many sets of 9 are in 27?

3 x _______ = 27

We can also solve 3 x 9 using an area model.

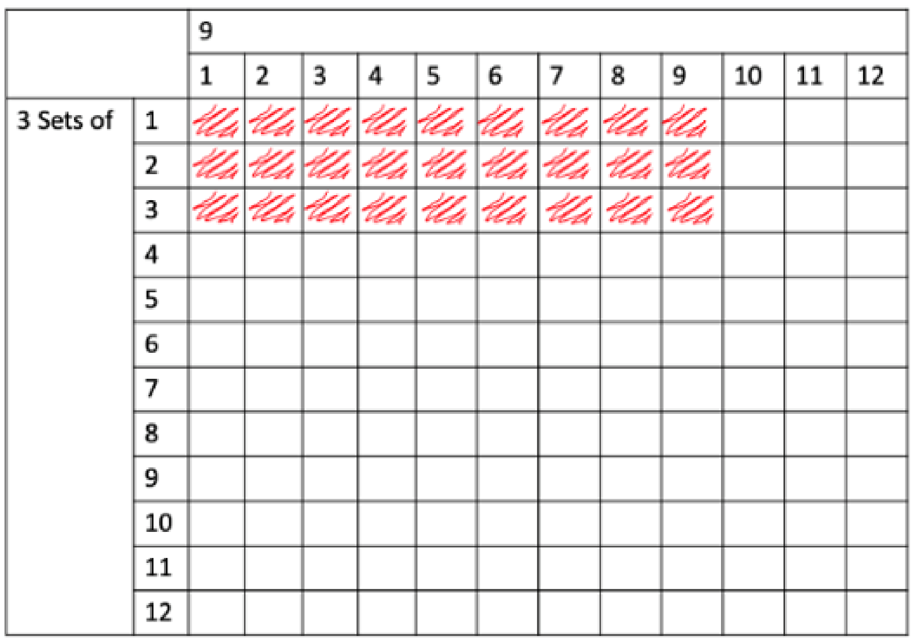


Figure 5.3. This scripted lesson introduces whole-number multiplication as repeated addition, maps the operation on a number line, and solves by counting on a grid.

Interventions also provide a model and guided practice to create equivalent expressions using the skill or operation. Creating equivalent expressions helps students make challenging problems easier to solve. In the example below, assisting the student to represent a multiplication problem as repeated addition provides the student another way to solve a multiplication problem for which the student may not readily know the answer. Similarly finding an “easy known” and counting up or down to the correct answer is another solution path that the student can use when encountering unknown problems. Finally, solving for an unknown via the “fill-in the missing number” activity provides practice in how the student can solve for an unknown (which is the general goal of math — to find solutions to real-world problems). In effect, these activities are guided practice to connect skills and content in mathematics (i.e., uncover the coherent nature of mathematics) and transfer the learned skill. Figure 5.4 show specific activities designed to teach for transfer for multiplication.

	Write as Addition	Find the Double or Near Easy	Fill-In Missing Number
6 x 7 =	6 + 6 + 6 + 6 + 6 + 6 + 6 = 42 Or 7 + 7 + 7 + 7+ 7 + 7 = 42	6 x 6 + 6 = 42 Or 7 x 7 – 7 = 42	6 x 4 + ___ = 42
8 x 9 =			8 x 10 - ___ = 72
7 x 8 =			7 x 4 x ___ = 56
9 x 6 =			3 x 3 x 3 x ___ = 54
7 x 4 =			7 x 2 x ___ = 28

Figure 5.4. This activity facilitates skill transfer in multiplication by creating equivalent quantities using addition, finding a near-double or near-easy fact, and filling in the missing number to create an equivalent quantity.

Finally, SpringMath interventions build fluency using games to embed high dosages of opportunities to respond (see Figure 5.5). High dosages of opportunities to respond at the instructional level of skill difficulty are one of the most effective ways to promote skill fluency. Fluent skills are resistant to forgetting and more likely to be used or modified to solve novel related problems (i.e., more likely to generalize). Thus, building fluency is a powerful way to build robust skills that can endure across time and be useful in learning new, related content.

Play a game with the student. You have several options here. Using the two bingo cards below, you can play several games. You can play against the student, using the problems from the day's worksheet to find and cover products on each card. You can simply use a pencil to cross products out and the winner is the first person to cross out all of the product on the grid. You can play the game such that the first person to get four products in a row either vertically or horizontally is the winner. You can have two students play against each other. Finally, a really fun variation is to make a secret mark in a few places on the grid using a Crayola "changeable" marker; the students color in each square when the square contains the correct product for the multiplication problem you call out. When the secret code is revealed, the student wins the game.

30	35	40	54	63
36	54	64	42	72
56	49	56	81	63
81	48	42	54	36
25	49	45	64	72

Figure 5.5. This bingo game is used to provide a high dosage of opportunities to respond practicing multiplication facts in a group format or one on one. Nearly all fluency-building interventions in SpringMath include at least one game to build fluency. All materials needed to play the game are contained in the protocol and/or in Support.

SpringMath interventions include practices recommended by the Institute for Education Sciences

The IES Practice Guide is another useful frame through which to examine the instructional features in SpringMath. The IES practice guide (Fuchs et al., 2021) identified six instructional strategies that were determined to have strong experimental evidence to support their use. The six strategies that should be used during instruction and intervention with students in math during the elementary years are provided in the table below, along with a description of how SpringMath addresses each.

IES Practice Guide (2020) recommended actions and SpringMath

	Recommended action	SpringMath features
1.	Systematic Instruction: Provide systematic instruction during intervention to develop student understanding of mathematical ideas.	Instruction progresses through specific skill sequences based on measured student performance and instructional tactic is aligned with student proficiency. Tasks and tactics are adjusted weekly based on progress-monitoring data. All instructional protocols use explicit instruction as explained in the last section and interleave conceptual understanding activities with procedural skill-building activities each day. Daily practice materials are provided using problem sets that are technically equivalent and designed to prevent misunderstanding. The minimum dosage for intervention is specified and demonstrates a very low incremental cost-benefit ratio (lower is better).
2.	Mathematical Language: Teach clear and concise mathematical language and support students' use of the language to help students effectively communicate their understanding of mathematical concepts and procedures.	All interventions are scripted to ensure clarity of mathematical language used by the teacher. These scripts avoid confusing and imprecise language (e.g., "borrowing," "reducing") and provide a model of effective math communication. All intervention protocols ask students to explain their thinking aloud and provide prompts for students to identify patterns and rules.
3.	Representations: Use a well-chosen set of concrete and semi-concrete representations to support students' learning of mathematical concepts and procedures.	In all intervention protocols, multiple quantity representations are provided. For example, when working with sums, manipulatives, number cards, dots or hashmarks, grids, and number lines are used. With more advanced skills, representations make use of using known operations to learn new operations (for example, multiplication is taught as repeated addition, mapped on a number line, area models, and eventually expanded notation). All interventions provided scripted activities to create equivalent quantities. Flexible quantity conversion is directly measured and cultivated via intervention with multiple quantity representations.
4.	Number Lines: Use the number line to facilitate the learning of mathematical concepts and procedures, build understanding of grade-level material, and prepare students for advanced mathematics.	Number lines are used extensively to map quantities, identify equivalent quantities, and map operations. Students label number lines, place quantities on number lines, draw operations on number lines, and choose the number line that will allow them to most easily map quantity. Number lines are used for whole number operations, fractions, decimals, percentages, and integers.
5.	Word Problems: Provide deliberate instruction on word problems to deepen students' mathematical understanding and support their capacity to apply mathematical ideas.	Word problems are provided for nearly all skills in SpringMath. These are solved with teacher support and the student is asked to "think aloud" while solving each problem.
6.	Timed Activities: Regularly include timed activities as one way to build fluency in mathematics.	SpringMath uses timed assessments and timed activities daily to build procedural skill fluency. Problem sets are generated according to specific parameters such that new problems are provided daily in the student material packets that are of the same difficulty level from day to day. Problem sets have also been designed to advance in known increments based on student learning gains. Thus, all timed practice is provided at the student's instructional level, which promotes strong student engagement, motivation to respond, and learning gains.

References

Burns, M. K., VanDerHeyden, A. M., & Jiban, C. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review, 35,* 401-418.

Fuchs, L.S., Newman-Gonchar, R., Schumacher, R., Dougherty, B., Bucka, N., Karp, K.S., Woodward,

J., Clarke, B., Jordan, N. C., Gersten, R., Jayanthi, M., Keating, B., and Morgan, S. (2021). Assisting

Students Struggling with Mathematics: Intervention in the Elementary Grades (WWC 2021006).

Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of

Education Sciences, U.S. Department of Education. Retrieved from whatworks.ed.gov

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., Jordan, N. C., Siegler, R., Gersten, R., Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105*, 683-700. dx.doi.org/10.1037/a0032446

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*, 1202-1242. dx.doi.org/10.3102/0034654309334431

Johnson, K., & Street, E. M. (2013). Response to intervention and precision teaching: Creating synergy in the classroom. NY, NY: Guilford.

Karp, K. S., Bush, S. B., & Dougherty, B. J. (2014). 13 rules that expire. *Teaching Children Mathematics, 21*, 18-25.

Ma, L. (1999). Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States. NY, NY: Routledge.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2012). Fractions: The new fronTier for theories of numerical development. *Trends in Cognitive Sciences*, 1-7.

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2012). Fractions: The new fronTier for theories of numerical development. *Trends in Cognitive Sciences*, 1-7.

VanDerHeyden, A., & Allsopp, D. (2014). *Evidence-based practices for mathematics* (Document No. IC-6). Retrieved from University of Florida, Collaboration for Effective Page 52 of 72 Educator, Development, Accountability, and Reform Center website: ceedar.education.ufl.edu/tools/innovation-configuration

Wu, H. H. (2009). *Understanding numbers in elementary school mathematics.* Providence, Rhode Island: American Mathematical Society.

Wu, H. H. (2011). Phoenix rising: Bringing the common core state mathematics standards to life. *American Educator*, Fall 2011, 3-13.

CHAPTER 6

SpringMath Effects

In Chapter 4, research data were provided on the assessments used for screening, diagnostic decisions, and progress monitoring in SpringMath. In this chapter, we will summarize the key findings regarding effects of SpringMath intervention on student learning. As a reminder, SpringMath functions as an integrated whole — the assessments, decision rules, interventions, and implementation support work in tandem to promote better math learning for all students.

SpringMath begins with a transparent logic model with scientific references attached to each claim. The basic theory of change, as detailed in Chapter 3, is that improved intervention content and tactic alignment and provision of all needed materials will improve intervention efficacy and correct intervention use, which will, in turn, cause more effective instruction, more efficient allocation of resources, and improved decision accuracy. That will, in turn, improve mathematics achievement. We expect to see improved math performance on proximal (sensitive, direct skill) and distal (more comprehensive, delayed measures such as year-end tests) measures. We expect to see closure of opportunity gaps (or comparable results for all subgroups in terms of proficiency and risk reduction). We expect that incremental cost effectiveness ratios are very strong and lower than most other models of math improvement.

6.1 Intervention effects

VanDerHeyden and Burns (2005) reported that CBM scores were higher following classwide math intervention (effect sizes ranged from 0.49 - 0.97) and year-end scores on the SAT-9 were stronger following intervention compared to the preceding year before intervention was introduced (effect sizes ranged from 0.29 to 0.45). These findings were promising but were not experimental. Subsequent studies have examined classwide intervention in a multiple baseline design and in two randomized, controlled trials.

A study using multiple baseline design across schools at grades 3 and 5 provided experimental evidence of classwide math intervention within MTSS (see Figure 6.1). In addition to improvements in proportionality and accuracy of students identified for special education (VanDerHeyden et al., 2007), gains in the percent of students meeting the year-end proficiency criterion on the state test were observed. Grade 3 saw an increase of 10% of students meeting the proficiency criterion, and grade 5 saw an increase of 20% with both grades approaching the 100% ceiling.

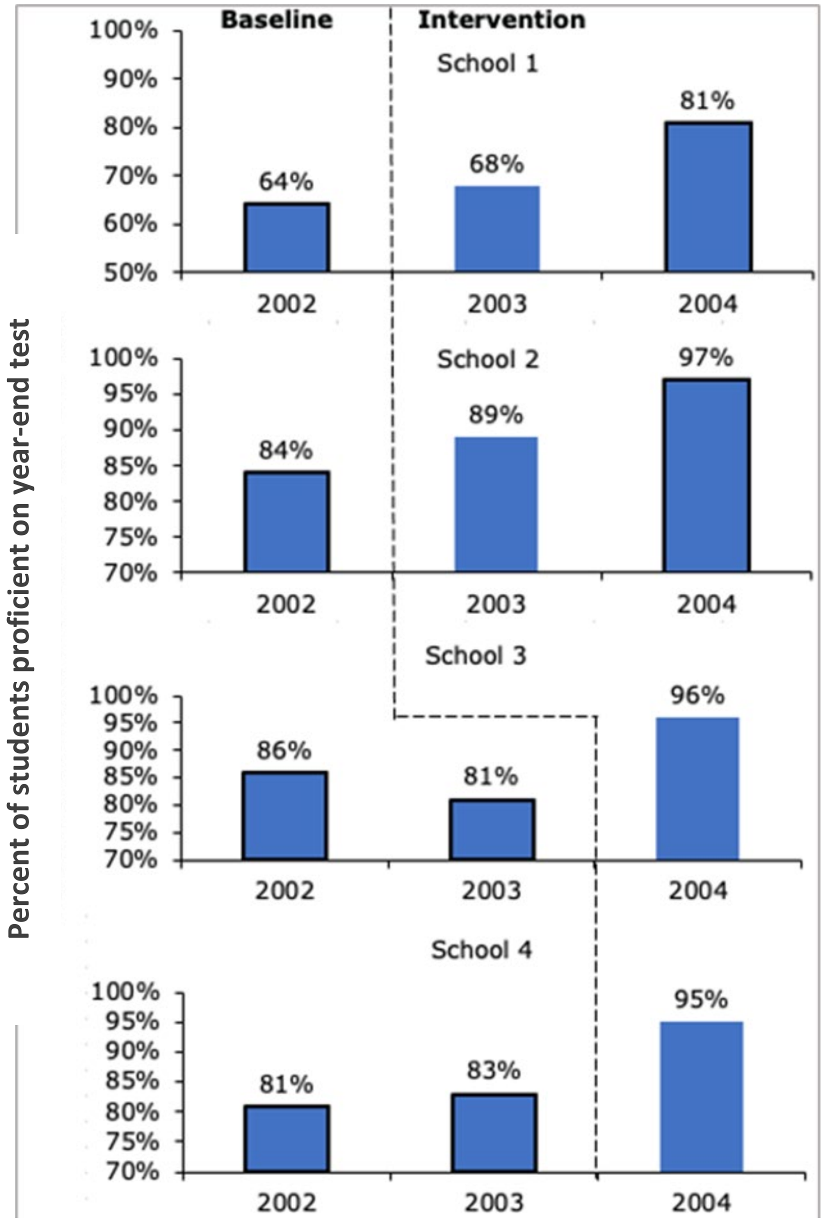


Figure 6.1. Math MTSS began in two elementary schools in 2003 and then was implemented in two additional schools in 2004. Values are the percentage of students proficient on the year-end state test in grade 5. All schools demonstrated an increase in percent proficient when math MTSS (with classwide math intervention) was introduced. This study provided experimental evidence of the intervention effect but provided only one replication of the effect (schools 3 and 4 replicated the effect seen in schools 1 and 2).

Overall effects on student achievement

A districtwide randomized, controlled trial with fourth- and fifth-grade students found strong gains on CBMs and moderate to strong gains on the year-end test scores at grade 4 (VanDerHeyden et al., 2012). Gains were stronger for students who had greater risk at baseline. Specifically, for fourth-graders who scored 1 SD below the mean on the preceding year's state math test had a treatment effect size of 0.41, and students who scored 2 SDs below the mean on the preceding year's math test had a treatment effect size of .66 on overall test scores for math. On the number and operation subscale, effect sizes were 0.29 (overall), 0.65 (for students 1 SD below mean on preceding year's test), and 1.00 (for students 2 SDs below mean on preceding year's test). These effects are all considered moderate to very strong.

Effects were also observed on all CBMs with students in the control groups and treatment groups scoring similarly at baseline, but with treatment effects ranging from 0.52 to 0.78 following intervention. Implementation integrity accounted for treatment outcomes in the treatment groups, which provides additional evidence of the intervention being the active ingredient in the improvements seen in the treatment groups.

Summary of effect sizes obtained in randomized controlled trials with SpringMath

Subdomain	Outcome measures	Effect size
Math Computation Fact Families for Addition and Subtraction 0–20 3-Digit Addition and Subtraction With and Without Regrouping Fact Families for Multiplication and Division 0–12	SpringMath measures from grades 2–4	0.49
Math Computation	Monitoring Basic Skills Progress — Computation	-0.22
Math Concepts Math Application	Monitoring Basic Skills Progress — Concepts and Applications	0.34
Math Computation	3-digit addition & subtraction with & without regrouping (grade 4)	0.56
Math Computation	Multiplication 0-12 (grade 4)	1.11*
Math Computation	Fact Families Multiplication & Division 0-12 (grade 4)	0.9*
Math Computation	Fact Families Multiplication & Division 0-12 (grade 5)	0.63
Math Computation	Multiply 2 by 3 digits without regrouping (grade 5)	0.66
Fractions	Simplify Fractions (grade 5)	0.98*
Comprehensive	Mississippi Curriculum Test, 2nd Ed. (grade 4)	0.79*
Comprehensive	Mississippi Curriculum Test, 2nd Ed. (grade 5)	-0.05

Table 6.1

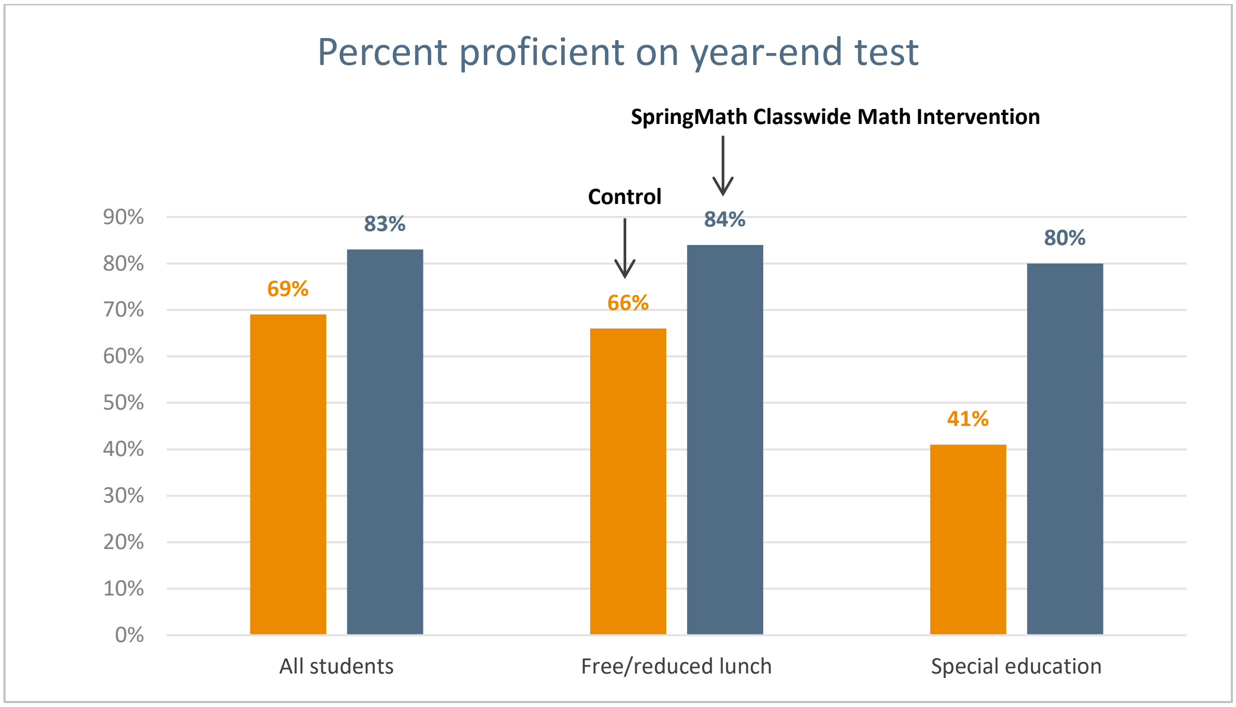


Figure 6.2. Percent proficient on the state year-end test for classes randomly assigned to business-as-usual control (orange) versus classes randomly assigned to classwide math intervention (blue). Strong effects were found for classwide math intervention for all students in grade 4 and for subgroups of students (e.g., students receiving free or reduced lunch and students receiving special education services).

Risk reduction effects

In a secondary analysis of the RCT data from the 2012 study, VanDerHeyden and Codding (2015) examined the intervention effects on risk reduction and equity in the fourth-grade sample.

They found very strong risk reduction for all students and especially pronounced risk reduction where risk was elevated at baseline. Specifically, they found that for every seven students who participated in classwide intervention, one of those students was prevented from failing the year-end test of math via classwide math intervention. For students who scored below the 25th percentile on the preceding year-end test, the number needed to treat was two, meaning for every two students who scored below the 25th percentile on the preceding year-end test and received classwide math intervention in the current year, one of those students was prevented from failing the current-year's test via the classwide math intervention.

Risk reduction effects with SpringMath classwide math intervention in a randomized controlled trial

	Absolute risk reduction	Number needed to treat
All students	15%	7
Students receiving free or reduced lunch	18%	6
Students receiving special education services	39%	3
Students scoring nonproficient on the preceding year's test	44%	2

Closure of opportunity gaps in achievement

Additionally, strong equity effects were also found, favoring intervention as shown in Figure 6.3. Achievement was disproportionate by race at baseline. In the intervention classes, achievement was proportionate by race following intervention. In the control classes, achievement remained disproportionate by race, with Black students performing much lower than White students. Specifically, in the SpringMath intervention classes, 85% of White students and 77% of Black students were proficient post-intervention. In the control classrooms where the intervention did not occur, 82% of White students were proficient, whereas only 57% of Black students were proficient.

Importantly, because race was comparably disproportionate in both control and intervention classrooms before intervention, this study provided experimental evidence that intervention produces equitable achievement. The following table summarizes effect sizes that have been reported in experimental (randomized controlled trials) studies of SpringMath on various outcome measures.

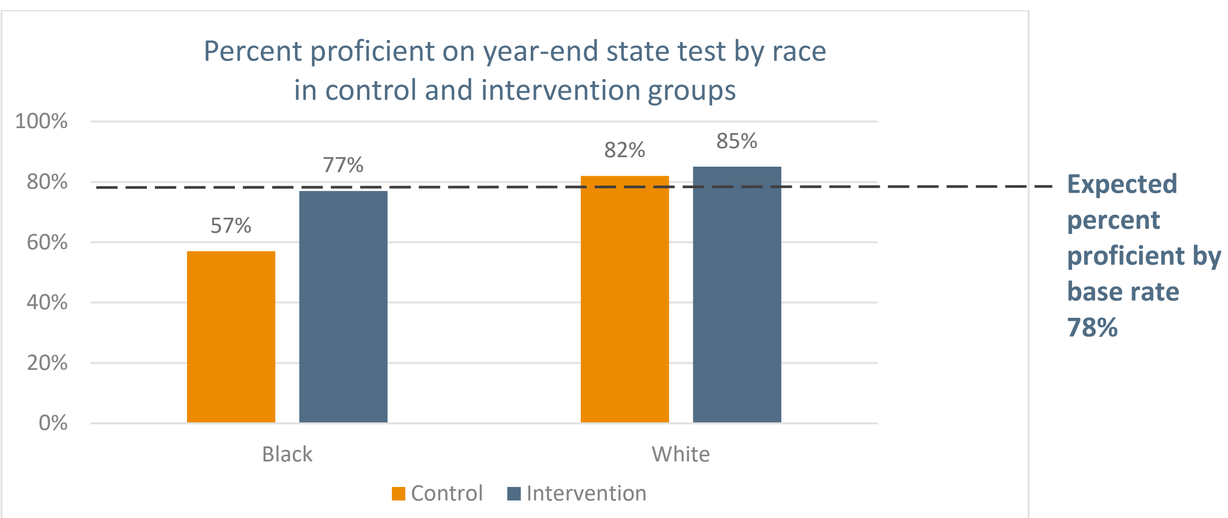


Figure 6.3. Classes were randomly assigned to classwide intervention or business-as-usual control conditions. Before intervention, Black students had a much lower rate of percent proficient on the state year-end test. Following intervention, this disproportionate achievement persisted in the control group, but the gap closed in the classwide math intervention group. Because all students received screening three times per year, this study supports the conclusion that classwide intervention is an effective mechanism to close opportunity gaps. Similar gap-closure effects were seen for students receiving special education services and students receiving free or reduced lunch.

6.2 Dosage research

Codding et al. (2016) conducted a dosage study using SpringMath classwide intervention. A total of 101 students who needed supplemental math instruction in grades 2, 3, and 4 were randomly assigned to one of four conditions: control, intervention delivered once per week in a 48-minute session, intervention delivered twice per week in two, 24-minute sessions, and intervention delivered four times per week in 12-minute sessions. Groups were equivalent on baseline measures. All students were monitored weekly on a variety of measures over four weeks of intervention. Intervention sessions and procedures were controlled and delivered by researchers to ensure that the overall number of minutes, learning trials, rate of praise, task materials, and other variables were identical in every way except for the way in which the dosages were administered. In other words, all students in all the intervention groups got the same number of minutes of intervention in a week and the intervention features were the same. Figure 6.4 shows effects on a computational progress-monitoring measure. The proximal computation progress-monitoring measures showed that the once-per-week intervention produced growth comparable to no intervention at all (same as control). The twice-per-week group showed better growth than the once-per-week group. The most frequent, shortest duration intervention (four days per week, 12 minutes per day) showed the strongest growth.

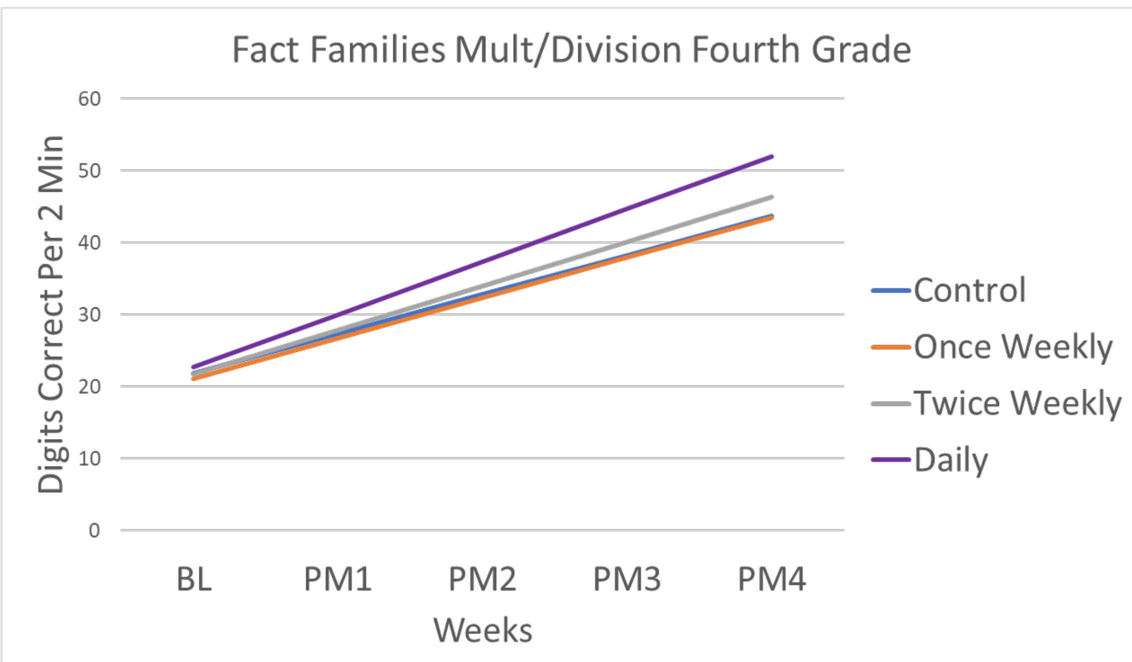


Figure 6.4. Progress-monitoring data are shown for a proximal computation skill across differing dosage groups with 101 students in a randomized controlled trial. All intervention students received the same overall amount of intervention, but it was parsed differently between groups. The group that got intervention in one, 48-minute session per week showed growth that was no different from the control group, who got no intervention at all. The 24-minute, twice-per-week group showed slightly better growth than the 48-minute and control groups. The 12-minute, four-times-per-week group showed superior growth. This finding is consistent with other current dosage research in MTSS and is the basis for recommending shorter duration, more frequent intervention sessions in SpringMath.

6.3 Cost analyses

Education research has begun to recognize that cost is an important feature of evidence-based intervention. Given two equally effective interventions, educators should consider the cost of intervention implementation and choose the lowest cost, most sustainable option. Costs in interventions reflect more than just the dollars required to buy the materials or the program. Costs include personnel time spent in training and implementation, instructional minutes required to use the tool, special equipment, and of course, the costs of the tool or materials. Such costs can be substantial and are often not fully considered by decision teams. Failing to consider the costs of a tool may threaten successful adoption when schools realize they do not have the resources for correct or sustained implementation.

Similarly, many tools are touted as “free,” and what this claim actually means is that there is no cost to the system to use the materials. Yet, implementation is never free for any tool. All tools require preparation, training, and time to use in addition to possible consumable material and equipment purchases. Second, building in-house tools and materials can ultimately compromise a successful implementation effort. For example, when systems are encouraged to build their own free math measures, this advice is actually really costly. Building math measures is technically difficult and requires a program of research to ensure that the measures function correctly and meet conventional assessment standards (e.g., reliability, validity). Building one’s own measures is also very time consuming. So now you are asking school employees to engage in technically difficult and time-consuming work to save the cost of buying commercial measures. In time and results, such advice is likely a higher cost than buying a set of technically validated measures would have been. In fact, it’s likely that the effort would be abandoned altogether (high up-front costs and complexity are known obstacles to adoption and implementation of evidence-based interventions). Incremental Cost Effectiveness Ratios (ICERs) are the most rigorous way to evaluate costs and benefits in education. ICERs are reported as the cost in dollars for one standard deviation of gain in the intervention group; thus, lower ICERs are better, indicating greater efficiency. Because the metric is a ratio, there are two ways to attain a lower ICER: lower the cost to implement and/or improve the effect given implementation. Readers should question cost-benefit analyses that are computed in the absence of a randomized, controlled trial.

Barrett and VanDerHeyden (2020) conducted a rigorous cost-benefit analysis of classwide math intervention implemented within SpringMath from a randomized, controlled trial. In this study the authors reported ICERs for classwide math intervention in SpringMath from a randomized, controlled trial. ICERs were computed three ways across a number of outcome measures, including year-end state test scores and more proximal outcomes like CBM screening scores. Specifically, they calculated costs in increasing degrees of comprehensiveness. First, they calculated the costs needed to implement, including the training, ongoing coaching and support, including salary costs for an RTI coach and school principal and program costs and materials. Second, they calculated the costs, adding the personnel costs required for teachers, RTI coach, and school principal to attend the trainings. Third, they calculated the costs, adding teacher salary costs for the time spent delivering the intervention each day. Among fourth-graders, incremental cost-effectiveness ratios ranged from \$74.01 to \$362.39, depending on which ingredients were included for CBM measures, and from \$57.73 to \$862.24, depending on which ingredients were included and level of prior risk for state assessment outcomes. Among fifth-graders, incremental cost-effectiveness ratios ranged from \$96.95 to \$540.21, depending on which ingredients were included for CBM measures. Across all students receiving treatment, it cost \$296.10 to prevent one student from failing the state assessment. The program was most cost effective for students who were already at high risk for poor educational outcomes because of the smaller number needed to treat among those subpopulations. Specifically, the cost was \$126.90 to prevent failure on the state assessment for one fourth-grader receiving special education services or for one student who scored below the 25th percentile on the prior year's state assessment. SpringMath ICERs were very strong and superior to changing curricula in math as shown in Figure 6.5.

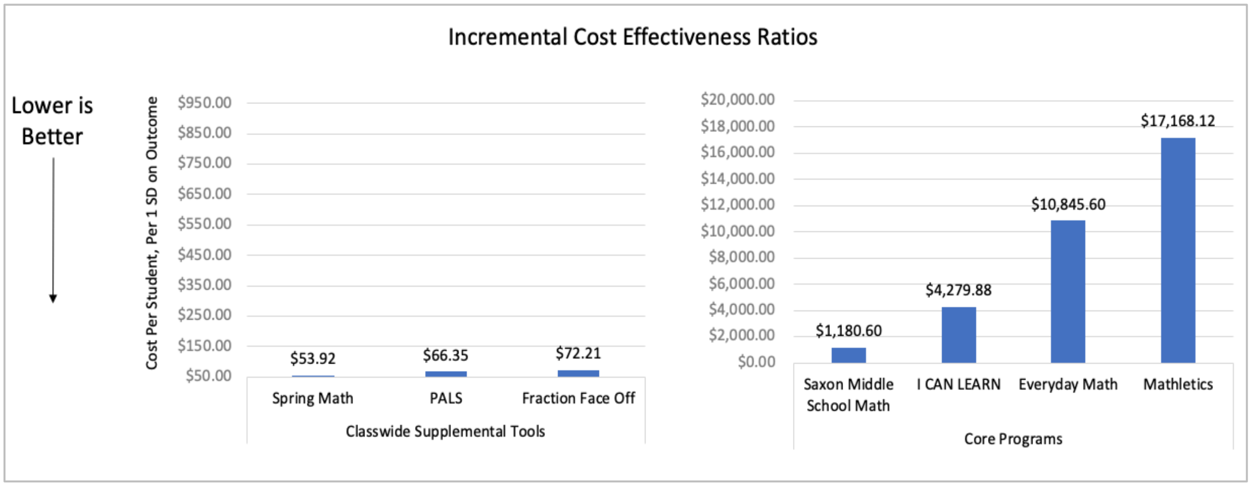


Figure 6.5. Incremental Cost Effectiveness Ratios (ICERs) are the cost in dollars (including costs broadly) needed to attain one standard deviation of gain on the outcome measure(s) in a randomized, controlled trial. On the left, researcher-constructed tools are shown that use classwide intervention. On the right panel (and a much higher scale) are the ICERs for core programs used in many schools. This graph shows the superior ICERs with supplemental intervention tools like SpringMath.

References

Barrett, C. A., & VanDerHeyden, A. M. (2020). A cost-effectiveness analysis of classwide math intervention. *Journal of School Psychology, 80,* 54-65. doi.org/10.1016/j.jsp.2020.04.002

Codding, R., VanDerHeyden, Martin, R. J., & Perrault, L. (2016). Manipulating Treatment Dose: Evaluating the Frequency of a Small Group Intervention Targeting Whole Number Operations. *Learning Disabilities Research & Practice, 31*, 208-220.

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. Assessment for Effective Intervention, 30, 15-31. doi.org/10.1177/073724770503000302

VanDerHeyden, A. M. & Codding, R. (2015). Practical effects of classwide mathematics intervention. *School Psychology Review, 44,* 169-190. doi: dx.doi.org/10.17105/spr-13-0087.1

VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. *American Education Research Journal*, 49, 1251-1284. doi.org/10.3102/0002831212462736

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. A (2007). Multi-Year Evaluation of the Effects of a Response to Intervention (RTI) Model on Identification of Children for Special Education. *Journal of School Psychology, 45*, 225-256. dx.doi.org/10.1016/j.jsp.2006.11.004

Implementation Support

Implementation support is the linchpin of any effective MTSS model (VanDerHeyden & Tilly, 2008). Logically, it is the natural human condition to wax and wane in taking the specific actions needed to accomplish goals in all areas of life. Poor implementation is the natural course of events and should be expected rather than bemoaned. Readers can most likely recall a time that they decided to undertake a nutrition or exercise program, change their budget, learn to play a musical instrument, or some other such goal. Even efforts that begin with great enthusiasm attenuate over time due to fatigue, distraction, extinction, or competition. Academic intervention is no different. It is the natural tendency of systems to drift in their implementation even when they are well committed to and well resourced for the effort. This problem of implementation is even more pronounced in systems that are not adequately resourced and skilled and in which no one is paying attention to implementation.

7.1 Implementation science as a pillar of math MTSS

Signs of poor implementation might arrive very late, after a sustained period of intervention failure. The leaders in the school or district may be the last to know, and before they have even had a chance to respond, teachers have lost faith in the effort and moved on. Worse, leaders often make two critical errors when an intervention effort is not being well implemented. First, they allow the failing effort to die slowly over a long period of time while hoping that somehow the effort will mature its way into better results. The second critical error that leaders often make is to try to solve an implementation problem by adding to the innovation as shown in Figure 7.1. When leaders add components to an innovation that is not working, complexity increases, and increased complexity is associated with a lower probability of correct implementation. Thus, a terrible cycle of implementation failure is predictable.

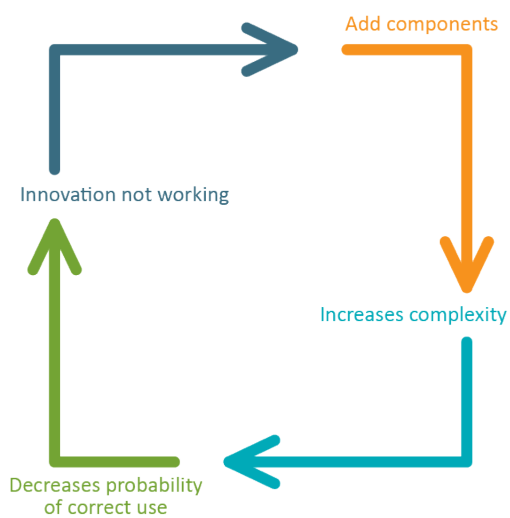


Figure 7.1. An unfortunately common response when an innovation is not working is to add to the innovation, which increases complexity. Increased complexity is associated with lower probability of correct use, and lack of use leads to lack of effects.

Sometimes, leaders are influenced by vendors who advise that a sustained period of implementation is required (sometimes years) before gains can be detected. That advice is an implementation red flag. No one should be expected to continue implementation on faith alone without seeing some gains right away. Leaders must position and equip themselves to the be the first to know about implementation and its effects. It is easy to imagine that after teachers have experienced a period of low-yield implementation, asking them to continue implementing to "see if things improve" is frustrating and demoralizing. Human behavior was not meant to be sustained in the absence of reinforcement (Skinner, 1953). Stokes and Baer (1977) wrote about the problem of a "train and hope" strategy for the use of newly learned skills. They argued for more systematic cultivation of what they called "generalization." Generalization occurs when people use or adapt a skill that they have recently learned or been trained to do in novel environments or in slightly different ways.

Intervention implementation can be thought of as generalized performance because our goal is for teachers to implement an intervention in classroom environments that are always shifting. Teachers need to be facile with implementing under different conditions (stimulus generalization) and adapting their implementation (response generalization) to improve results. In other words, teachers have to be agile implementers. Adaptive leadership is necessary to cultivate agile implementers (Heifetz, 1994). Heifetz defined technical leadership as fully engaged, responsive management to implement known tactics and reliably return results. He argued that adaptive leadership, in contrast, is necessary to deliver results when the tactics and sequence of tactics are not known in advance and implementation is more complex, which is certainly the case with MTSS. Determining which actions to take depends upon the ongoing problem-solving process. Different schools will need different MTSS actions, though the problem-solving structures are stable and can be taught and mastered. In characterizing adaptive leadership, the National Implementation Research Network stated,

> "When systems undergo change, the natural tendency of those in the system is to look to those in authority to minimize the tension of change and regain stability. However, when change is the goal, formal authority can get in the way of leadership because it is designed to maintain systems, not to help people overcome their natural tendencies to maintain the status quo. When organizations and systems are being changed on purpose, adaptive leadership is needed to manage the change process" (National Implementation Research Network).

Adaptive leaders promote active problem solving, build trust, and continually refine implementation in increments as the path forward becomes more obvious. Adaptive leadership is possible only when leaders know where intervention is being implemented well and where it is not, in real time. Leaders may try one approach but then make an adjustment when that approach does not bear fruit, and this requires specific adaptive leadership skills so that process does not feel chaotic to implementers. We said in the first chapter that changing organizational behavior is a human enterprise. At the crux of this effort is trust. Teachers have to trust that the leader can effectively lead an effort, be a good steward of resources, and guide the effort to benefit the students in their classrooms.

Schein (2013) defines trust as "believing that the other person will acknowledge me, not take advantage of me, not embarrass or humiliate me, tell me the truth, not cheat me, but rather work on my behalf and support the goals we have agreed to" (p. 79). If the leader does not systematically build trust during implementation, teachers may short change the implementation effort, hedge their bets by using alternative strategies, or simply avoid investing their time in the implementation. Trust cannot develop in situations where leaders do not understand in a very granular and technical way what the key implementation behaviors must be. Most readers can probably recall one or two examples during their careers where a teacher leader did not have the technical know-how to train and support the implementation of an instructional technique. Such implementations will fail because such leaders can give only vague and nonspecific guidance. Leaders must have some facility with the tactics so that anticipated barriers to implementation can be identified, teachers can receive up-front training but also ongoing coaching support to “do” the effort in their classrooms, and leaders can be effective allies toward attaining the desired outcomes.

Rethinking the flow of information can facilitate MTSS implementation. In the organizational behavior change literature, Daniels (2013) has written about information flow as traditionally being a top-down process where information flows from the executive to the manager to the supervisor and then to the front-line worker as shown in Figure 7.2. Daniels argues instead for a process he calls “reverse behavioral engineering” and has demonstrated strong improvements on key metrics in industries targeting outcomes like improved safety measures, for example. Daniels suggests that when the leaders begin with an operational understanding of the front-line implementation, and that information flows up to the supervisor, to the manager, and finally to the executive, that there will be better alignment between front-line implementation needs, implementation support, and ultimately implementation effects.

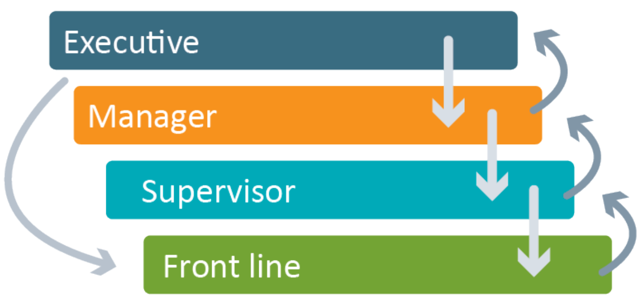


Figure 7.2. Daniels’ model of “reverse behavioral engineering” reverses the typical flow of information to systems to begin with the front line and flow up to the executive to inform leadership decisions about how to support effective front-line implementation.

Daniels (2013) has written about reverse behavioral engineering in schools, stating,

> “If teacher accountability is to create successful students, what is the accountability of the other staff? It is actually quite simple. Every staff member’s charge is to help teachers be successful. The only reason any education staff exists at the school, county, state, or federal level is to help teachers educate children more effectively. Right away you can see that accountability in these jobs should be primarily for valuable behaviors that have a direct link or connection to increased student learning” p. 82.

It is exactly this advice from behavioral science that guided our own placement of the teacher in the center of SpringMath’s theory of change described in Chapter 3. MTSS can be an effective way to reorganize a system’s goals and contingencies for high-quality instruction and construct the MTSS implementation around that. We suggest teacher-student instructional interactions be the center of the universe in any MTSS effort (see Figure 7.3).

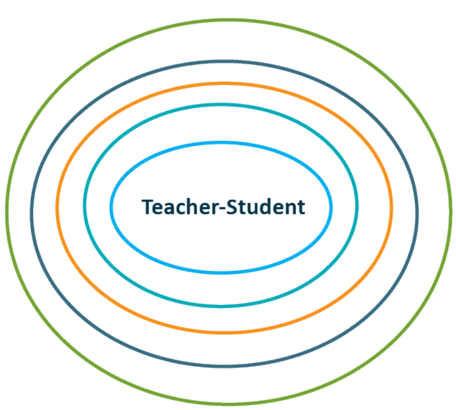


Figure 7.3. If the purpose of schooling is to improve learning and that learning begins with the relationship between the student and teacher, every other structure and support can be evaluated for whether it supports what happens between the teacher and the student.

In MTSS, student learning is the fuel that drives MTSS. Instructional design cannot be separated from its result (i.e., learning). It is the learning, the quality of the learning, the rate of learning, and the robustness of the learning that must guide the instruction that follows. On outcomes, Bushell & Baer (1994) described two worlds in education.

> "One is the world in which we choose everything we want to teach from the universe of measurable goals, and then find the teaching techniques to maximize those measures. In the other world, we choose only *how* we will teach, as if there were an a priori correct way; whatever we get by teaching that way then must be what we want or should want. It is especially easy to believe that, when what we want is not measurable — or at least not measured. Perhaps the second world is so consistently preferred over the first because the first is doubly embarrassing: It requires us to decide what we want to teach, and it is always embarrassing to discover that in some small part we don't know and in some small part we don't agree. And it exposes us to a series of surprises as we discover what techniques teach the goals better than others" (p. 266).

Academic screening and year-end testing data are a great basis for identifying learning needs and understanding one's baseline. Outcome goals should be about growth. Measures should be sensitive to detect small increments of growth, and these gains should be used to promote continued and better implementation. There is a saying in mountain biking that, "the bike goes where the eyes go." The same is true for MTSS implementation. That which is attended to and measured is that which will receive implementation focus from teachers.

In any implementation effort, teachers bring a certain skill set, and the environment presents opportunities to enable and reward implementation and also to stall, punish, and extinguish implementation. Cultivating implementation means improving teacher skill and capacity to implement and cultivating an environment that prioritizes and rewards implementation (see Figure 7.4).

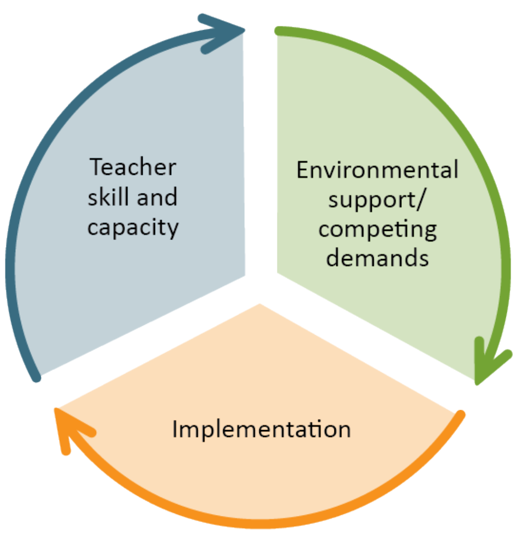


Figure 7.4. Teacher skill and capacity to implement and environmental conditions that either reinforce or punish implementation can be systematically cultivated to favor correct implementation through behavioral shaping at the system, teacher, and student level.

When a system prioritizes intervention implementation, one should see that structural decisions made at the school are aligned with that goal. Far too often, teachers are expected to undertake a new initiative and no one has attended to whether sufficient instructional minutes are available for the effort and if not, what adjustments will be made to ensure that such time is made available. This lack of alignment between a structure like the school's schedule and the system's stated intervention goal creates a culture where adults focus entirely on the lack of a resource like time when the team could tactically troubleshoot and move forward (Burns & Gibbons, 2012).

Effective implementation leaders identify when behaviors that compete with effective implementation are unwittingly being reinforced or when the implementation itself is being punished. For example, teachers may be expected to deliver classwide intervention for 15 minutes each day, but special school programs are frequently scheduled to occur during that time period. Or perhaps classwide intervention occurs just before lunch and the teacher receives negative feedback from the principal when the teacher arrives to the cafeteria several minutes after the lunch period has begun so the teacher begins to cut the classwide intervention short to allow more time to get students to the cafeteria. Schools are ripe with examples of competing contingencies that undermine intervention implementation: not allowing teachers to make copies of practice materials because of copy budgets, removing teachers from the classroom to participate in professional development activities, creating schedules that are untenable and generally guided by shared enrichment teachers and spaces rather than what would be ideal for instruction. When implementation is the behavior we wish to cultivate, then we must measure implementation directly and then think of environmental antecedents and consequences that can be adjusted to establish, accelerate, and sustain implementation. Table 7.1 summarizes barriers and rewards for implementation. Rewards will vary among teachers. For example, some teachers may appreciate recognition whereas other teachers may find public recognition uncomfortable. Effective leaders learn to identify the rewards that matter to teachers and attempt to give access to those reinforcing consequences when implementation is going well.

Barriers and rewards for implementation

Barriers/Punishers for Implementation	Rewards for Implementation
Lack of skill	Child learning gains
Lack of materials	Positive appreciation from admin
Lack of time to implement	Recognition among peers
Lack of follow-up or progress monitoring (no knowledge if working)	Positive appreciation from family
Complex and resource intensive	Change of role or opportunities for new work
↑ Minimize, remove, attenuate	↑ Facilitate immediacy, frequency, quality. Make less predictable.

Table 7.1.

SpringMath is designed specifically to optimize the antecedent conditions associated with optimal implementation, to measure implementation directly along with gains in student learning, and to cue effective collaboration, troubleshooting, and in-class coaching for improved implementation. Before we explain the implementation science built into SpringMath, we will summarize key strategies for creating an MTSS-implementation-friendly environment.

7.2 Create an MTSS-implementation-friendly environment

We said in Chapter 1 that you cannot wait for core instructional problems to change without intervention, especially in mathematics. Core curricula in math are notoriously weak (Slavin & Lake, 2008), so waiting for curricula changes to improve learning is a costly decision in dollars (Barrett & VanDerHeyden, 2020) and especially in learning opportunities for students (Koon & Davis, 2019). The structures of MTSS, on the other hand, can be used to deploy intervention, measure improvements, and make better decisions. Toward that end, we have some specific implementation advice.

Optimize the host environment

Core instruction can be improved in tandem with the implementation of screening, progress monitoring, and classwide intervention. In-tandem improvement is essential because the assessment structures of MTSS provide the best data from which to evaluate learning improvements, set meaningful targets for improvement, and evaluate the core program over time. Classwide intervention offers immediate, protective benefit to students who are receiving core instruction that is not working well. Table 7.2 provides a checklist of core features that can be examined and targeted during MTSS implementation.

Core features to verify and adjust during MTSS

Curriculum factor	Evidence
Curriculum is aligned with state standards.	There is substantial overlap between skills and content taught in each subject and the specific items described in the state content standards at each grade level.
Curriculum provides coordinated instructional sequences.	There is a logical flow from basic to more advanced skills in a sequence that corresponds to research literature (e.g., sequence of letter sounds to be taught; sequence of math computational skills to be presented). An instructional calendar specifies time points by which certain skills ought to be mastered by all students. Student assessment data are available to know whether most students are meeting benchmarks.
Curriculum facilitates explicit teaching.	Explicit teaching techniques that address the skills and content for that subject and grade are described in the curriculum materials. Teaching techniques include specific procedures for teacher instruction, including teaching scripts and routines. Skills to be taught are not merely embedded in more generic teaching guides but are explicitly stated and connected to particular lessons designed to establish those skills for students.
Grouping practices maximize student engaged time.	Strategies for creating flexible instructional groups during core instruction that match instruction to student skills are described. Clear focus is given to instructing students at their instructional levels.
Instruction moves from initial scaffolding to transfer and generalization.	The curriculum includes ample strategies for providing instructional supports (scaffolding) during skill acquisition as well as specific strategies for application of skills in real-life contexts that allow for generalization and transfer of learning. Adequate instructional time is provided for fluency building to ensure students reach mastery for essential skills. Strategies are provided to promote retention of learned skills. Strategies are provided to support students who do not master skills at expected time points.
Implementation with high integrity leads to positive outcomes for students.	Procedures for assessing the integrity of curriculum delivery are part of the curriculum and are used on a regular basis (e.g., peer coaching, administrative observation). There is evidence that the core curriculum produces proficiency for at least 80% of students.

Table 7.2. From Kovaleski, VanDerHeyden, and Shapiro (2012). Table reprinted with permission from Guilford.

Select tools that are aligned with science

Selecting tools that are consistent with the science of how students learn and have sufficient evidence to indicate their efficacy in classrooms is the foundation of stewardship. Understanding the science of instruction allows decision makers to evaluate potential tools with regard to their demonstrated alignment with the science of instruction. In this text, we detailed how SpringMath is aligned with MTSS (Chapter 2), technically adequate assessment (Chapter 4), and the science of instruction (Chapter 5). Decision makers can also consult the tools charts provided by the National Center for Intensive Intervention (NCII) (intensiveintervention.org), which provide external quality ratings of vendor-supplied evidence. NCII also provides Intervention Taxonomy briefs, which summarize effects of various interventions, content coverage, and specific inclusion and dosage of evidence-based instructional tactics within specific tools.

Verify that the selected tool is a good fit with your needs and resources

Selecting the right tool means that systems must have a good grasp of what their specific needs are and their available resources to implement. If a tool requires resources that your system does not have, it is not the right tool for your system. Teachers liking a specific tool should not be a precondition for the tool's adoption, but tending to the implementation to minimize effort and intrusion and to maximize learning gains should be a priority so that teacher acceptability of the tool can improve over time. If teachers require specific training to use a tool, the implementation plan should be built to ensure that teachers are trained and equipped for the initial implementation. Ongoing implementation monitoring and support are always necessary to cultivate sustained intervention use.

Don't pilot, scale

It seems that systems have determined that pilot programs are always necessary in an adoption and that the purpose of the pilot is to "see if the innovation can work" before resources are invested in the innovation at scale. The unintended consequence of seemingly logical advice is that many systems find themselves in a series of poorly run mini experiments to try to select an intervention program that works. Such an approach invites underinvesting in the innovation because implementers logically view the innovation as possibly temporary and experimental. Most "pilots" are poorly run and no one really attends to the data, so the final result can become evaluated according to how people feel about the tool or innovation, reported informally. How teachers feel certainly matters, but adults' feelings should not be the basis for the selection and implementation of instructional strategies in schools. Instead, efforts should be selected that are known to work in well-controlled research and field trials. Fortunately, excellent resources are available to vet tools. For example, readers can consult the Academic Interventions Tools Chart at the National Center for Intensive Intervention (intensiveintervention.org). Importantly, most interventions fail because they simply were not implemented well, so adult effort needs to be allocated to supporting the intervention (VanDerHeyden & Tilly, 2008).

We recommend choosing an intervention with a proven track record that seems affordable and to suit the needs of your system; then, direct most of the effort to scaling the innovation in your system. This subtle shift in approach changes everything. Scaling, rather than piloting, conveys that leaders have done their due diligence to select a tool that can work, that is well-aligned with the needs of the school, and have been good stewards of resource investments in the school. Scaling tells teachers that the innovation is expected to work. The implementation effort becomes about making the innovation work in their system. Scaling implies that the effort may start smaller and expand over time as skill and capacity to implement improve. The pace at which a system scales has to be determined locally in concert with available resources to support implementation and ideally follows a timeline that is not set in advance, but rather advances or accelerates implementation in increments as the system demonstrates readiness. In our experience, failing to scale rapidly enough is as common as scaling too rapidly. The key is to attend to the available resources for the effort and gains attained via implementation. As soon as gains are observed, implementers should advance implementation. Critically, leaders are wise to keep the charge front and center to teachers that implementation of the selected tool is the key to driving improvements.

In-house, ongoing program evaluation will change your life

Selecting a tool that has been shown to work is an important first step, but it is no substitute for in-house program evaluation. Systems over focus on selection of the tool and under focus on implementation and program evaluation. Ideally, systems implement like detectives with ongoing program evaluation that tells them how well the tool is being implemented and with what results in real time. Such ongoing program evaluation is the basis for adjusting program implementation to bring about better results. Centering the conversations on student outcome data allows data teams to make smart decisions about how to improve results in real time.

Detect implementation failures when they are still small and rapidly respond

Weekly progress-monitoring data can be used to sensitively indicate where growth is not occurring as expected. Lack of growth is the canary in the coalmine that should immediately prompt implementers to examine use of the tool directly. Is the tool being used at the full recommended dose? What is the quality of the implementation? What barriers could be interfering with correct and complete tool use? What adjustments might improve the implementation and results? Responding weekly allows poor implementation behaviors to be diagnosed and corrected so that implementers do not develop entrenched implementation habits that do not work. Correcting implementation errors early also builds and sustains motivation to use a tool because the student growth can be quickly established or re-established following troubleshooting or coaching to repair the implementation error. It's useful to remember that "train and hope" is not a strategy. Ongoing monitoring, rapid detection, and specific troubleshooting and coaching are necessary to get the implementation on track and sustain the effort. Implementation can be monitored via student learning. Where gains are not observed in student learning is where the coach should direct in-class support (Gilbertson et al., 2008).

Because implementation integrity is such a pernicious threat to most intervention work in schools, there is a large body of literature that provides excellent tactical guidance on how to deliver optimal implementation support (see Figure 7.5). For example, based on implementation integrity research, one can dismantle or avoid known barriers to implementation by minimizing intervention complexity, minimizing the number of adults needed to conduct intervention, minimizing the time required to implement an intervention, and providing all materials needed to conduct the intervention. Implementation integrity research shows that even given very strong antecedent supports for intervention implementation, most interventions will still not be implemented as planned (Wickstrom et al., 1998; Witt, et al., 1997). But one tactic has been found to dramatically improve implementation integrity -- performance feedback (Mortenson & Witt, 1997; Noell et al., 2005). Performance feedback involves a trained coach keeping track of intervention use and conducting consultation with the teacher to troubleshoot and improve intervention use.

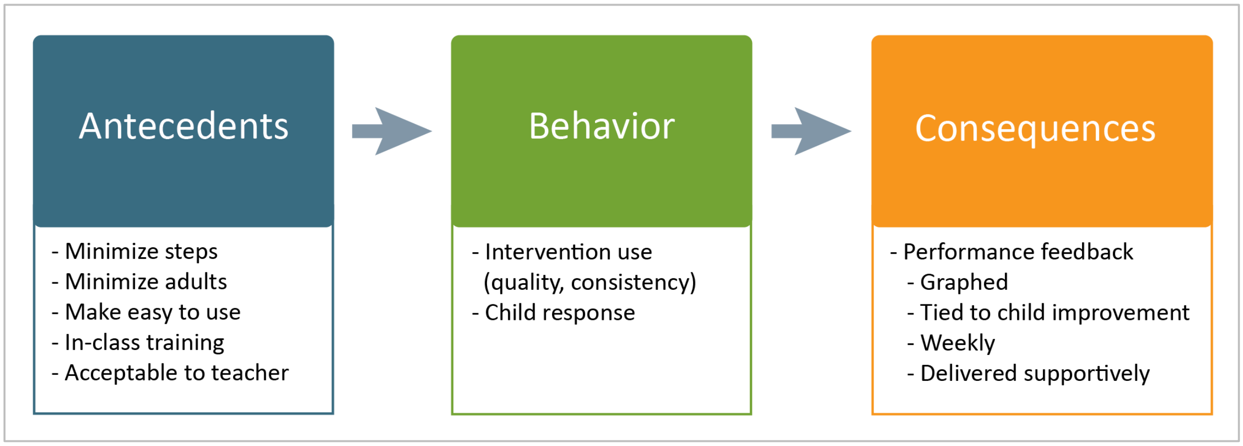


Figure 7.5. Implementation and student learning gains are the targeted behavior. Antecedent and consequent events that are known to promote implementation and student learning gains can be planned and adjusted to maximize correct implementation.

In SpringMath, antecedent implementation supports include time-efficient, easy-to-administer assessments, automated data interpretation, time-efficient mechanisms of intervention delivery (e.g., classwide intervention), scripted protocols for intervention, and provision of all needed materials to take the action that is recommended by SpringMath at each step. We provide onboarding training so that the school is ready to implement. SpringMath was built to include specific antecedent supports but also to enable consequent supports for implementation because consequent supports are essential to maximize intervention integrity. SpringMath provides consequent support for implementation by automating growth charts in the teacher and coach dashboards so that teams can see the results of their instructional efforts immediately.

Specific antecedent and consequence supports for implementation built into SpringMath

Antecedent supports for implementation	Improve efficiency to minimize number needed	• Use of screening rules to minimize over identification of students for more intensive instruction. • Use of implementation support in classes using classwide intervention to avoid over identification of students for more intensive instruction.
	Minimize complexity	• Assessments are brief, with standard instructions printed on each. Scoring is answers correct and answer keys are provided. • Assessment data interpretation is automated. • Intervention protocols are scripted. • All materials needed to implement intervention are provided.
	Improve teacher skill to implement	• Training is embedded in SpringMath with brief "just-in-time" videos explaining how to complete each recommended step. • Support portal contains materials, training videos, and guidance on specific questions that might arise. • Onboarding training is required and positions teams to implement. • Virtual and on-site training support is available at an added cost.
Consequence supports for implementation	Automated data interpretation recommending next action	• Following each score entry, performance is summarized, and the next action is recommended.
	Frequent, sensitive feedback on learning gains	• The coach dashboard shows gains at the school, grade, class, and student level. These data are harvested and updated with assessment that occurs as part of the intervention. • The teacher dashboard shows gains at the class and student level. These data are harvested and updated with assessment that occurs as part of the intervention.
	Guidance to enable performance feedback	• Sensitive implementation metrics are tracked and summarized in the coach dashboard so coaches can understand what actions are occurring and with what results. • Recommendations to provide in-class coaching or consultation with the teacher are suggested for classes and students who are not showing expected progress. • Coach and teacher dashboards can be used to guide data team meetings.
	Program evaluation	• Program evaluation is automated and provides effects by dosage of SpringMath to enable more effective implementation.

Table 7.3.

Next, we will detail how the coach dashboard is designed to assist coaches to know where teachers require in-class troubleshooting and coaching support for better implementation. We will then detail how the support portal is organized for implementers to provide resources to improve implementation. Proprietary metrics (Education Research & Consulting, 2013) drive recommendations in the coach dashboard to provide in-class coaching support where it is needed to improve intervention effects schoolwide (Joyce, & Showers, 1981; Fixsen & Blasé, 1993; VanDerHeyden et al., 2012). These metrics are sensitive and reflect dosage of intervention and results across grade levels. They are provided via the coach dashboard, which allows school leaders to know where intervention is being used, with what results, and to see the results in real time. To provide consistent metrics and facilitate data team discussions, a parallel version of this dashboard is provided for teachers for their own classrooms. In other words, the coach can view all teachers' information and each individual teacher can view the same information but only specific to his or her own classroom.

7.3 How to use the coach dashboard to drive implementation

Coaches can simplify and organize their MTSS work by thinking of certain milestones for implementation. The first milestone is to ensure complete screening in the school. This action creates a database that will be highly useful to tracking growth over time and providing feedback to teachers about the learning gains in their classrooms. The second milestone is initiating classwide interventions where they have been recommended. The third milestone is ensuring classwide math intervention growth for most students. The fourth milestone is beginning individual interventions. The fifth and final annual milestone is the annual program evaluation process to plan the intervention effort for the following year.

When a coach thinks of MTSS implementation as a series of milestone actions to accomplish, the coach can create implementation chains in the classroom. For example, in SpringMath, we make the first behavior easy for the teacher. During onboarding training, we assist teachers to log in and access their already-loaded rosters. The consequence of completing the first action becomes the antecedent for the next action, which is screening. All screening materials are provided with easy-to-administer instructions printed at the top of each screening. A short video tutorial explains how to give the screening. The coach dashboard summarizes which teachers have and have not completed screening so the MTSS effort can focus on making sure all teachers complete the screening as that is our second milestone behavior. When the screening is complete, the consequence is live summary reports of student performance and the recommended action, which in the example below is classwide intervention. Classwide intervention thus becomes our next milestone behavior. The consequence of classwide intervention implementation is the live summary reports of student learning gains and recommendations of students for individual intervention.

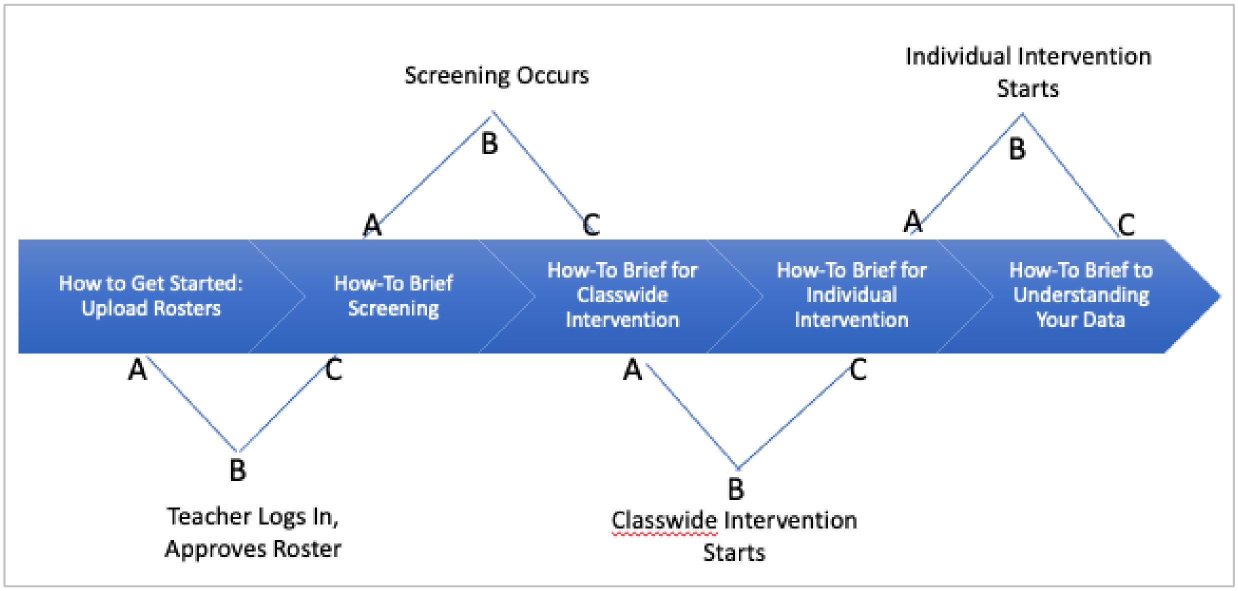


Figure 7.6. Organizing MTSS implementation as a series of milestone behaviors allows implementers to create implementation chains where the consequence for the last action is also the antecedent for the next action.

We recommend that coaches log in daily to their SpringMath account and view their dashboard as their "command central" to characterize the quality of implementation of key actions and make their weekly to-do lists for troubleshooting and supporting teachers. In this section, we will detail how to use the coach dashboard to its fullest advantage. The coach can check to see if any new interventions have been recommended this week and make a note to check in with those teachers. Next, the coach can use the following process to plan important in-class coaching support for the week. The coach dashboard will recommend visits where they are needed, but the coach can also work through the following to know where to go and why to support intervention implementation.

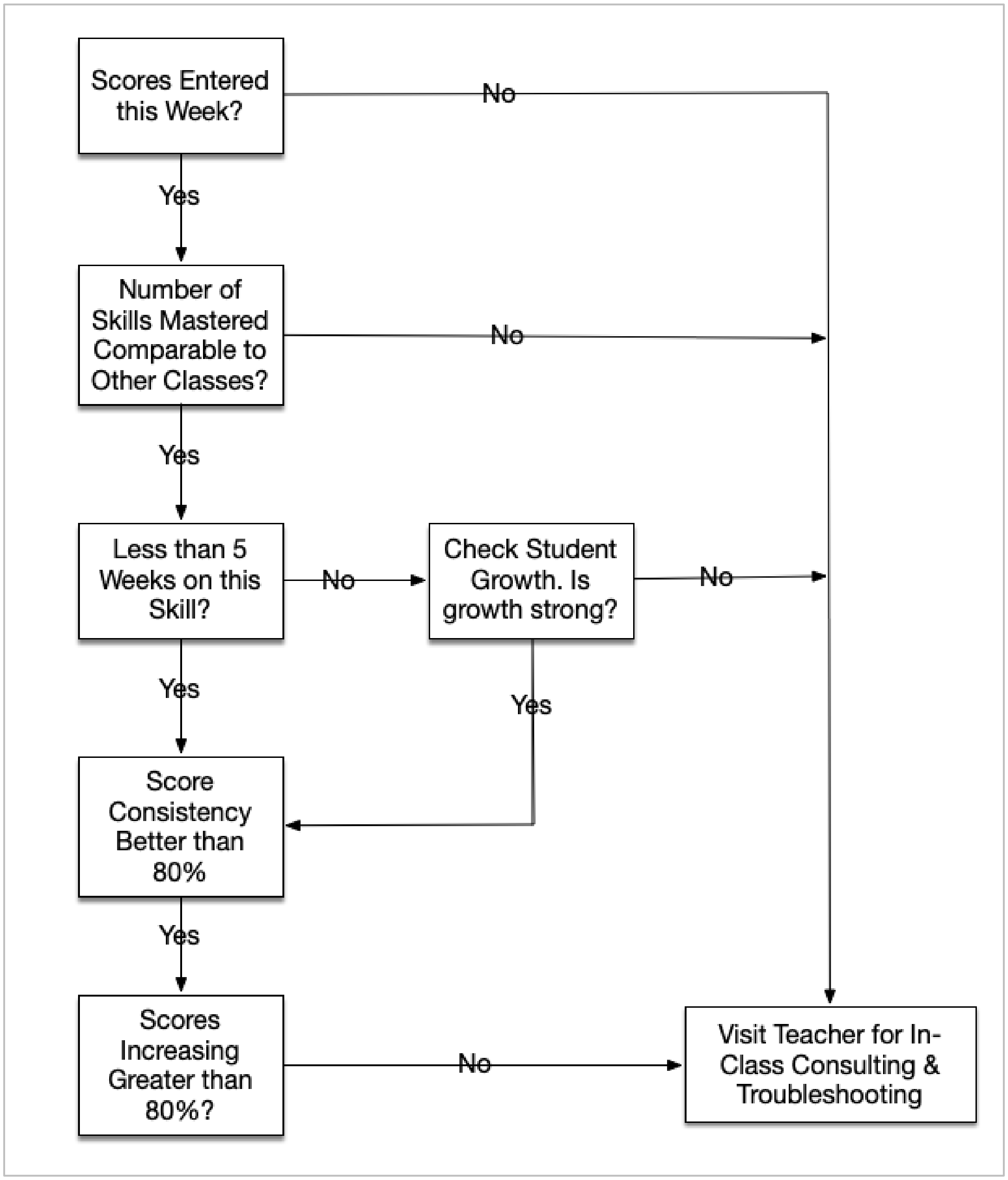


Figure 7.7. Coach dashboard process for recommending in-class coaching support.

School overview dashboard

The coach dashboard opens on the School Overview page. The School Overview page is divided into three sections. The first section details screening progress for the current season, which is shown in Figure 7.8. The purpose of this summary report is to enable the coach to see where screening has and has not been completed in the school so that the coach can encourage complete collection of screening data, which will be the starting basis for MTSS and program evaluation. There is flexibility on the exact timing of screening within a screening window. SpringMath will alert you that the window is open and report who has and has not completed screening. The coach should pay attention to whether the screening is complete by the agreed-upon date for screening for the school. The coach should be asking, "Have all teachers in all grades completed the current screening? If not, why?"

To determine which teachers are missing data, the coach can click on the incomplete circle and see which teachers have and have not completed their screenings. For teachers who have not completed the screening, the coach can check in with the teacher and ask if the teacher is able to print the materials and conduct the brief screening. If the teacher has never screened before, there will be a brief pop-up video in the teacher's dashboard that overviews how to conduct the screening. Sometimes teachers have completed screening but not entered the screening scores. The coach can recommend to the teacher that he or she enter the screening scores, even if some students have missing scores.

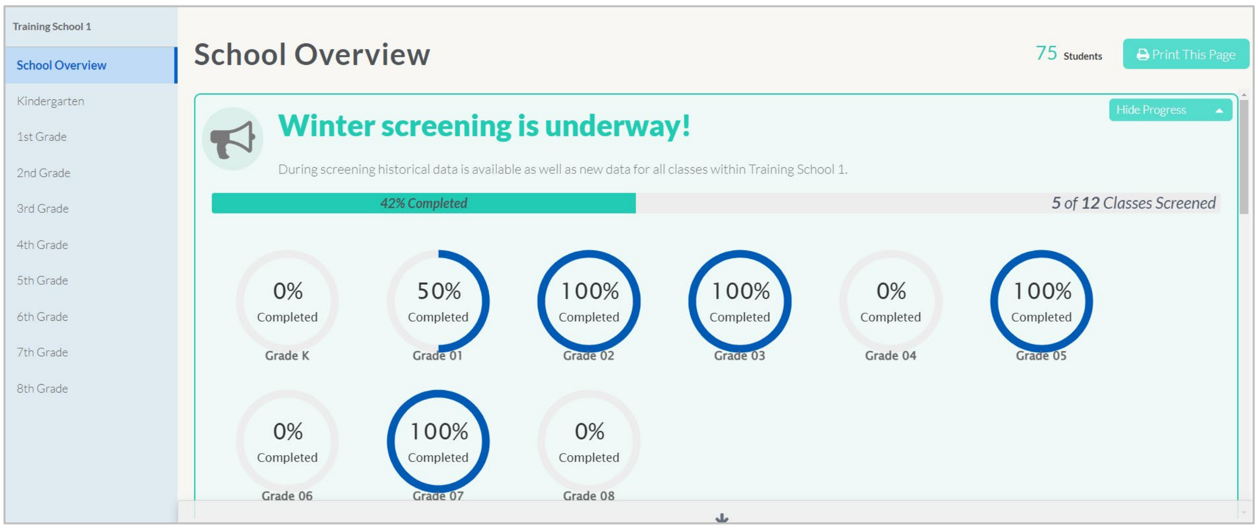


Figure 7.8. Summary of screening completion. Grades 2, 3, 5, and 7 have completed all screening for winter. No classes in grades K, 4, 6, and 8 have completed the winter screening. Half of the teachers in grade 1 have completed screening.

Once all of the grades have completed screening, coaches can hide this section of the dashboard by clicking the Hide Progress button in the upper righthand corner. The next order of business for the coach is to view where intervention has been recommended, to determine if the intervention has begun, and to view the effects of intervention in real time.

Figure 7.9 shows the coach dashboard landing page. From here, the coach can access all the information needed to promote effective implementation, prepare for data team meetings, and conduct data team meetings in person or remotely. In fact, conducting a consultation without the benefit of actual implementation data is an implementation pitfall to be avoided. When a teacher says, "my students have made no progress," the coach must think like a detective to verify that lack of progress is a problem, then consider all the possible reasons that lack of progress may be occurring. The coach dashboard is designed to help coaches troubleshoot intervention and improve overall effects in the school.

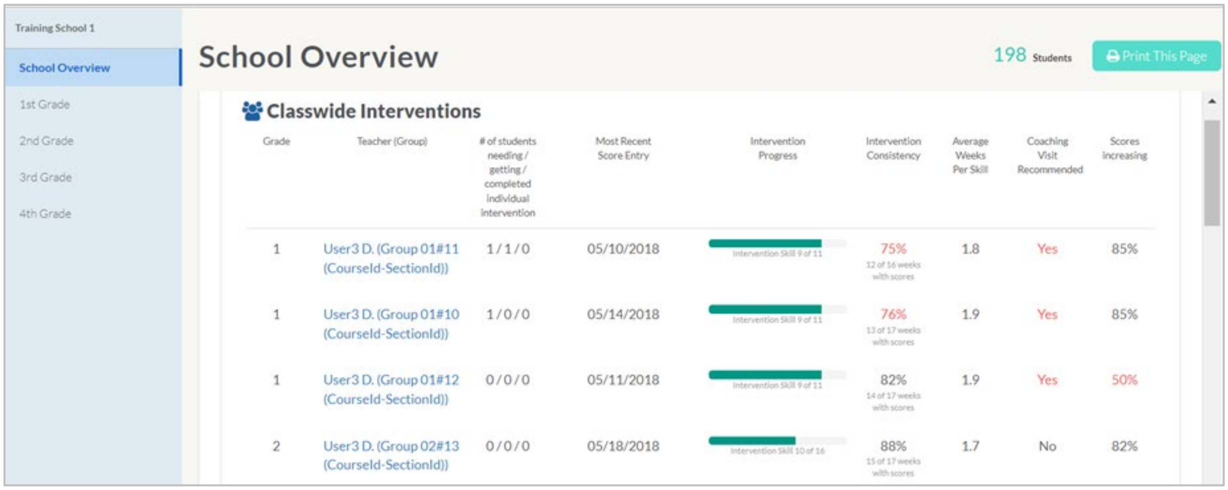
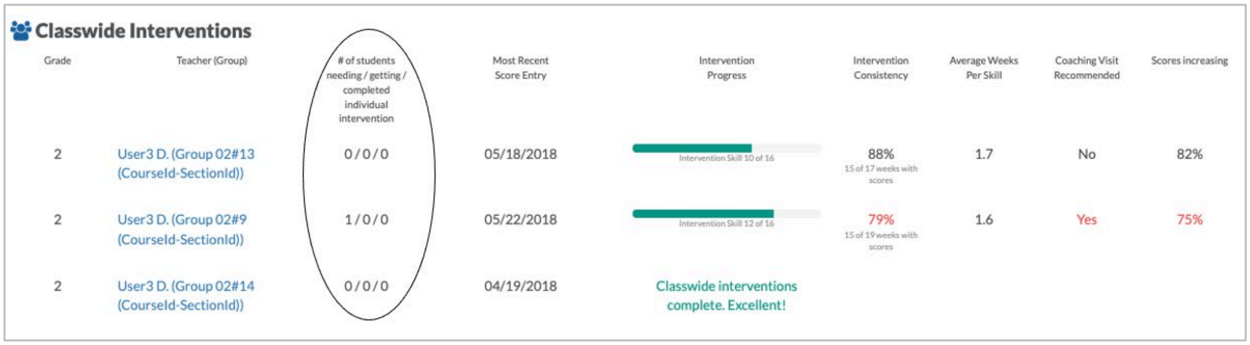


Figure 7.9. The School Overview tab in the coach dashboard provides all information to know where intervention is occurring, with what results, and where in-class support might be needed for improved implementation.

The second part of the School Overview page of the coach dashboard summarizes the progress of classwide math intervention in the school (see Figure 7.10). As an orientation to this page, the first column indicates the grade level of a given class. Coaches mostly want to compare progress in number of skills mastered (the green progress bars) between classes at the same grade level, but it is useful to identify if some grades are lagging behind other grades as that should signal to the coach that support should be allocated to lagging grades. When an entire grade seems to be lagging in progress, coaches can choose to spend time observing core instruction, attending grade-level data team meetings, and checking curriculum materials and resources for that grade level. The second column contains the teacher name and course section. A coach can click on the teacher name and the teacher's dashboard will open up, creating an easy way to collaborate with the teacher on his or her specific progress with intervention.

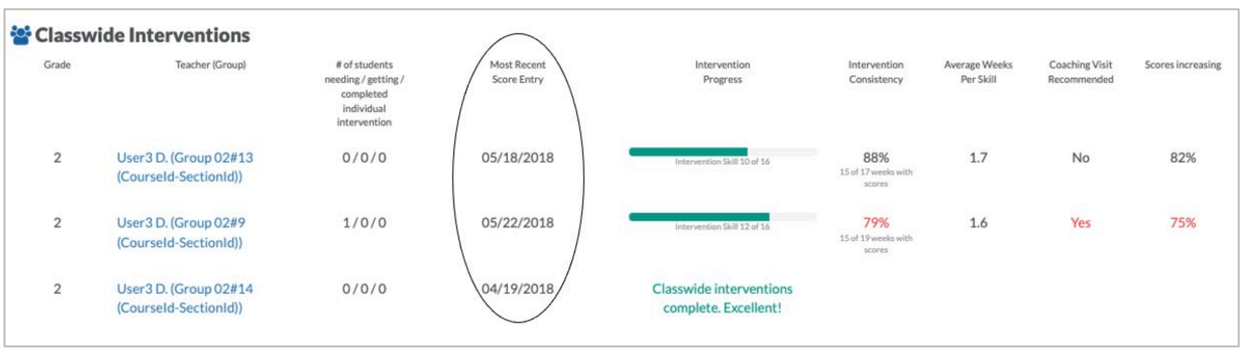
The next two columns tell the coach whether individual interventions have been recommended in a given class grouping, whether those interventions have commenced, and whether any of the interventions have been completed. The following column is the most recent score entry. Coaches want to scan these two columns to know that scores have been entered for the preceding week if applicable. The coach will want to check in with any teacher who did not record a weekly score because this is the first sign that intervention is off track. There is only one exception to this rule: When a class has completed classwide intervention, the most recent score entry will be the day that the class completed intervention and we would not expect additional scores as was the case for the class in the bottom row of Figure 7.10 and Figure 7.11. In this example, we can see both classes are currently engaged in the intervention and entered a score for last week.



Classwide Interventions

Grade	Teacher (Group)	# of students needing / getting / completed individual intervention	Most Recent Score Entry	Intervention Progress	Intervention Consistency	Average Weeks Per Skill	Coaching Visit Recommended	Scores increasing
2	User3 D. (Group 02#13 (CourseId-SectionId))	0/0/0	05/18/2018	Intervention Skill 10 of 16	88% 15 of 17 weeks with scores	1.7	No	82%
2	User3 D. (Group 02#9 (CourseId-SectionId))	1/0/0	05/22/2018	Intervention Skill 12 of 16	79% 15 of 19 weeks with scores	1.6	Yes	75%
2	User3 D. (Group 02#14 (CourseId-SectionId))	0/0/0	04/19/2018	Classwide interventions complete. Excellent!				

Figure 7.10. The coach can see the number of individual interventions that have been recommended, the number that have been started, and the number that have been completed.

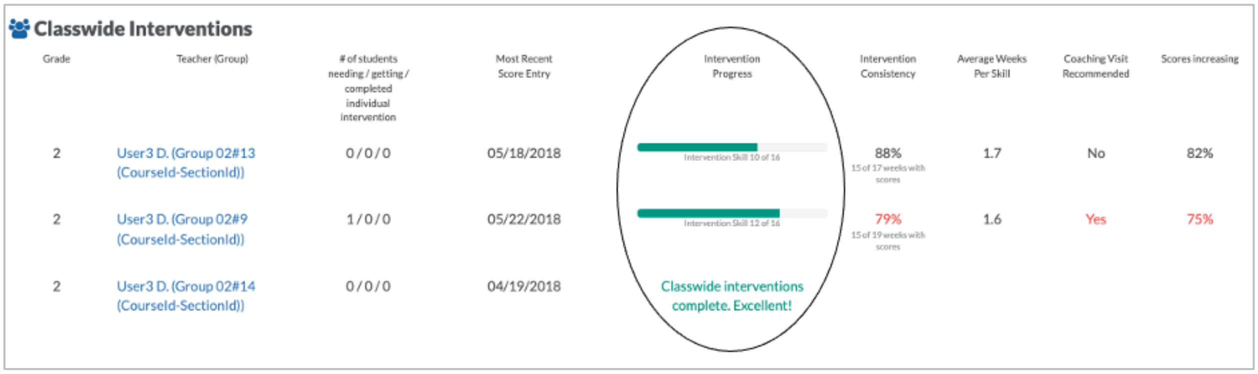


Classwide Interventions

Grade	Teacher (Group)	# of students needing / getting / completed individual intervention	Most Recent Score Entry	Intervention Progress	Intervention Consistency	Average Weeks Per Skill	Coaching Visit Recommended	Scores increasing
2	User3 D. (Group 02#13 (CourseId-SectionId))	0/0/0	05/18/2018	Intervention Skill 10 of 16	88% 15 of 17 weeks with scores	1.7	No	82%
2	User3 D. (Group 02#9 (CourseId-SectionId))	1/0/0	05/22/2018	Intervention Skill 12 of 16	79% 15 of 19 weeks with scores	1.6	Yes	75%
2	User3 D. (Group 02#14 (CourseId-SectionId))	0/0/0	04/19/2018	Classwide interventions complete. Excellent!				

Figure 7.11. The coach can view the date of the most recent score entry, which is an effective way to note where teachers are not engaged in implementation and might need a coach visit.

Next, the coach should examine intervention progress to determine where in-class coaching will improve intervention effects. All remaining columns of information will help the coach know where troubleshooting support is needed and what implementation targets might be fruitful. In Figure 7.12, the Intervention Progress column is circled. The coach can see the relative rate of skill mastery (or trials to criterion) data across classes, getting the same intervention in the same context with the same degree of integrity. The expectation is that classes should roughly track along at a similar pace of skill mastery under those conditions. Where a class is lagging, the coach dashboard could recommend an in-class visit to improve the intervention effects. In this example, the coach dashboard is showing the progress for all three grade 2 classes. The first class has mastered 10 of 16 skills. The second class has mastered 12 of 16 skills, and the third class has mastered all 16 skills. The coach might be thinking that class one needs a visit, but the dashboard will help you dig a little deeper before investing in an in-class visit that might not be needed.



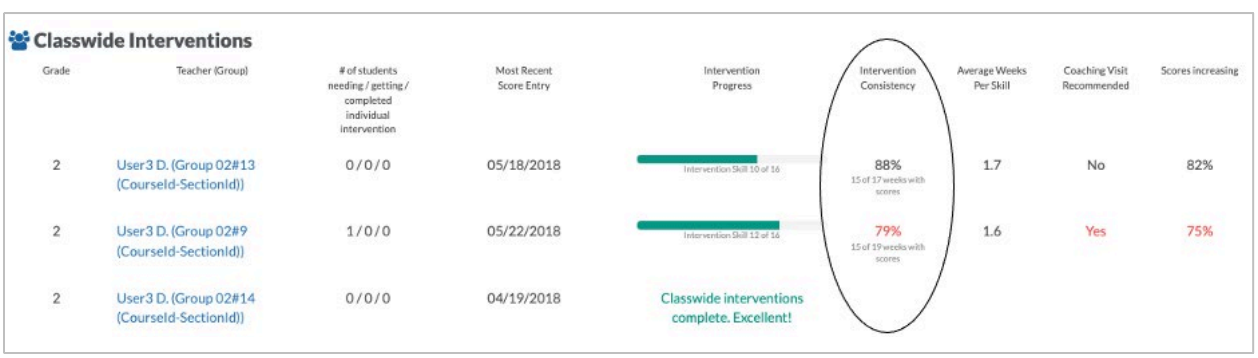
Classwide Interventions

Grade	Teacher (Group)	# of students needing / getting / completed individual intervention	Most Recent Score Entry	Intervention Progress	Intervention Consistency	Average Weeks Per Skill	Coaching Visit Recommended	Scores increasing
2	User3 D. (Group 02#13 (CourseId-SectionId))	0 / 0 / 0	05/18/2018	Intervention Skill 10 of 16	88% 15 of 17 weeks with scores	1.7	No	82%
2	User3 D. (Group 02#9 (CourseId-SectionId))	1 / 0 / 0	05/22/2018	Intervention Skill 12 of 16	79% 15 of 19 weeks with scores	1.6	Yes	75%
2	User3 D. (Group 02#14 (CourseId-SectionId))	0 / 0 / 0	04/19/2018	Classwide interventions complete. Excellent!				

Figure 7.12. This column shows the number of skills mastered across classes within a given grade. In this example, one class in grade 2 has finished classwide intervention, one class has mastered 10 of 16 skills, and the other class has mastered 12 of 16 skills. In this case, the class that has mastered only 10 skills is making good progress and does not need an in-class coaching visit, but the class that has mastered 12 skills needs a coach visit because Intervention Consistency and Scores Increasing are just below the recommended thresholds.

The next column is Intervention Consistency (see Figure 7.13). In this second-grade example, we can see that the first class has 88% Intervention Consistency and the second class has 79% Intervention Consistency. This value reflects intervention use because the intervention cannot be used without entering a score. Thus, when the class is recommended for classwide intervention and the teacher clicks the button to initiate classwide intervention, SpringMath begins to look for at least one score per seven-day cycle (with a little grace period to accommodate a weekend here and there).

To obtain the intervention materials, a score is required; thus, when a score has not been entered we can assume that the intervention was not used correctly. Intervention Consistency can be affected by school calendars. There is an automatic (under-the-hood) adjustment to correct for conventional holidays when school would be closed and we would not expect scores to be entered. Intervention Consistency can be negatively impacted by school calendars that take available classwide math intervention time for special programs and field trips. Intervention Consistency can also be affected by teacher absences when the substitute teacher does not know how to conduct the session. Finally, Intervention Consistency can be affected by school closures related to weather, for example. These effects on Intervention Consistency are permitted because they rightfully reflect a loss of instructional opportunity that is extremely important for adults to attend to and be intentional about.



Classwide Interventions

Grade	Teacher (Group)	# of students needing / getting / completed individual intervention	Most Recent Score Entry	Intervention Progress	Intervention Consistency	Average Weeks Per Skill	Coaching Visit Recommended	Scores increasing
2	User3 D. (Group 02#13 (CourseId-SectionId))	0 / 0 / 0	05/18/2018	Intervention Skill 10 of 16	88% 15 of 17 weeks with scores	1.7	No	82%
2	User3 D. (Group 02#9 (CourseId-SectionId))	1 / 0 / 0	05/22/2018	Intervention Skill 12 of 16	79% 15 of 19 weeks with scores	1.6	Yes	75%
2	User3 D. (Group 02#14 (CourseId-SectionId))	0 / 0 / 0	04/19/2018	Classwide interventions complete. Excellent!				

Figure 7.13. Intervention Consistency is the number of weeks with scores entered. Because the teacher must enter a score to get the next week's intervention materials, consistent score entry is a proxy for implementation.

When the value is shown in red font, that is intended to call the coach's attention to a concerning value. In the example shown in Figure 7.13, 79% is considered low for Intervention Consistency. But again, the next column can help the coach understand what might be happening here. The next column (Figure 7.14) is Average Weeks Per Skill. This value is the average number of weeks that it has taken a given class to reach mastery on skills during classwide intervention. Again, the expectation is that skill mastery should track similarly between classes within a given grade level. Here we can see that the two classes still conducting classwide math intervention are mastering skills at roughly the same pace. When taken in concert, these three columns help us understand that the second class started the intervention earlier than the first class, and, therefore, has mastered two additional skills compared to the first class. The Intervention Consistency score is therefore most likely just below the desired threshold because the teacher did not a record a score or two early in the intervention course. The coach can always check that assumption by clicking on the teacher's name and looking directly at gains made during classwide intervention for this class over the last few weeks.

The goal is that classes will master a given skill in no greater than four weeks; thus, the coach dashboard will recommend an in-class coaching visit for any class that has been stuck on a skill for four weeks. The purpose of this in-class coach visit is to improve intervention implementation to accelerate growth. The teacher classwide intervention page indicates roughly where classes should be at the mid-point of the school year in the classwide intervention skill sequence.

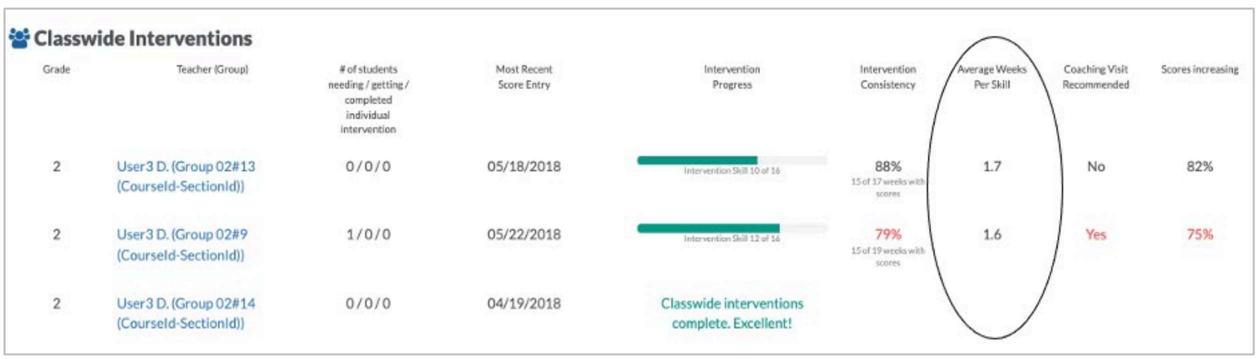


Figure 7.14. Average weeks that the class is requiring to reach mastery on skills are shown here. Given a limited number of weeks in the school year and a given number of skills to cover (16 in this example), this metric allows the coach to have a sense of how many skills a class is likely to master. Ideally, classes work to keep average weeks per skill to about two weeks or less.

The next two columns are Coaching Visit Recommended and Scores Increasing (see Figures 7.15 and 7.16). Coaching Visit Recommended is SpringMath's interpretation of whether an in-class coaching visit is recommended. Scores Increasing is an indication of the quality of the intervention. Specifically, Scores Increasing computes the percent of students in a given class for whom this week's score was at least one point greater than last week's score on the same skill. Technically, this metric reflects the stimulus control of the intervention, or the extent to which the intervention is producing growth for most students, which is the goal of the intervention. Fast growth in the face of a very low Scores Increasing value would signify that classes are rushing through the intervention and failing to get most students to mastery.

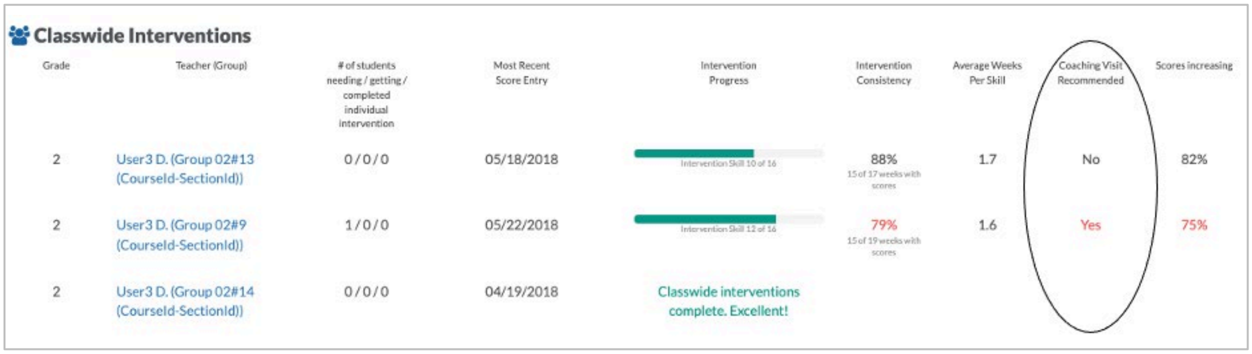


Figure 7.15. Coaching Visit Recommended interprets the progress and integrity metrics and recommends an in-class coaching visit as needed.

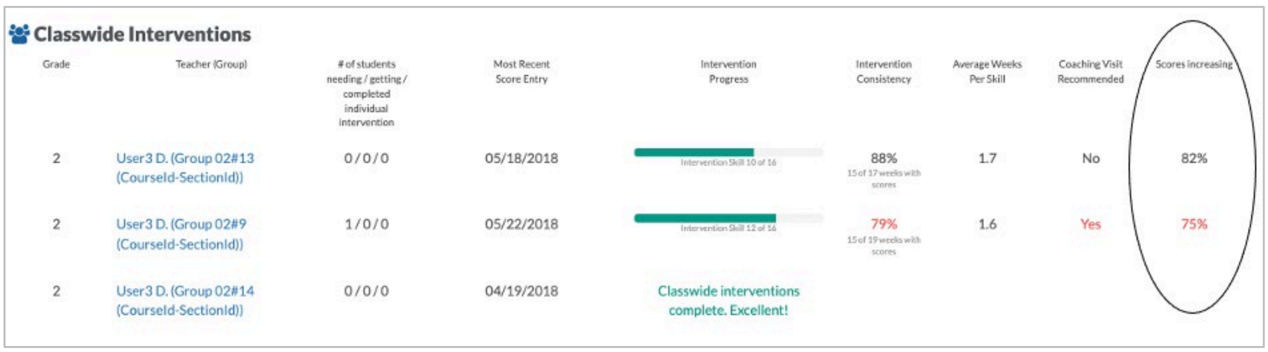


Figure 7.16. Scores Increasing is an indication of intervention quality. It reflects the percent of students for whom this week's score is greater than last week's score on the same skill. Ideally, 100% of students show growth from week to week.

The last section of the school overview coach dashboard is called Screening Results and an example is shown in Figure 7.17.

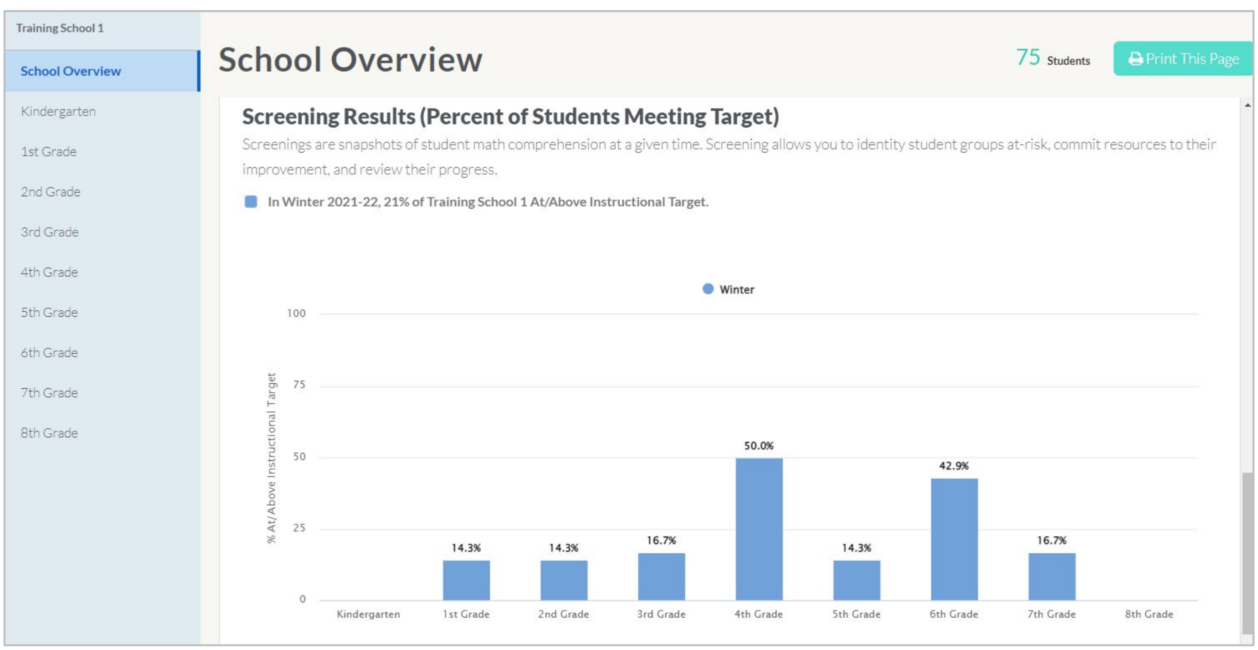


Figure 7.17. At the bottom of the School Overview page in the coach dashboard is the percent of students scoring in the instructional range or higher on all of the most recent screening assessments. Because screening assessments increase in difficulty during the course of the year, this is a rigorous and difficult metric to move.

The Screening Results summarize the percent of students scoring at or above the instructional target for each grade level on all the current screening assessments. It is amazing how often teachers and even coaches and principals in a school do not have a clear and current understanding of the percentage of students who are proficient on the year-end test in reading and mathematics. We often ask this question when we are working with specific schools, and schools that can instantly answer are the rare exception. Similarly, one of the pitfalls of MTSS is collecting but then not using or acting on screening data.

The Screening Results summary reflects a high bar for proficiency at each grade level in SpringMath because the screening assessments themselves are rigorous and increase in rigor with each screening occasion (from fall to winter to spring), and requiring a student to be above the instructional target on all screening measures to be counted is also rigorous. These graphs are something to pay attention to as a school implements SpringMath, especially across consecutive years. The coach can click on a grade level on the left, and the same graph will be shown by teacher. This way a coach can determine if any given class grouping sticks out. If a given class sticks out like a sore thumb, the coach can click into that teacher's dashboard to view the teacher's screening results, and most importantly, that teacher's Growth tab to see whether the teacher is moving students in the right direction. Before the Screening Results move, the "baby indicators" must move in a positive direction. The baby indicators include consistent use of classwide math intervention and gains in the Growth tab. If gains are not apparent in the Growth tab, then that is a teacher who is most in need of coach support.

Additionally, data teams and leaders can consider or reconsider resource allocation, examination of placement practices, and more frequent examination of progress-monitoring data for any class groupings that are showing risk. It's always useful to remember that learning is the most predictable outcome of effective instruction and so weak gains are just a signal to collaborate, troubleshoot, and try to change that equation (Gilbertson, et al., 2007). SpringMath can be a springboard to core improvements and classwide math intervention is a great starting point to nudge up the intensity of core instruction in math. When gains are not happening as desired, teams can ask questions about the time allocated to core instruction, instructional calendars, teacher use of evidence-based instructional tactics along with the math curriculum to move learning in the right direction.

Grade-level dashboards

After scanning the School Overview page and getting a sense of where you will want to be spending your time as a coach in the coming week, it is extremely helpful to click into each grade-level dashboard on the lefthand menu (Figure 7.18).

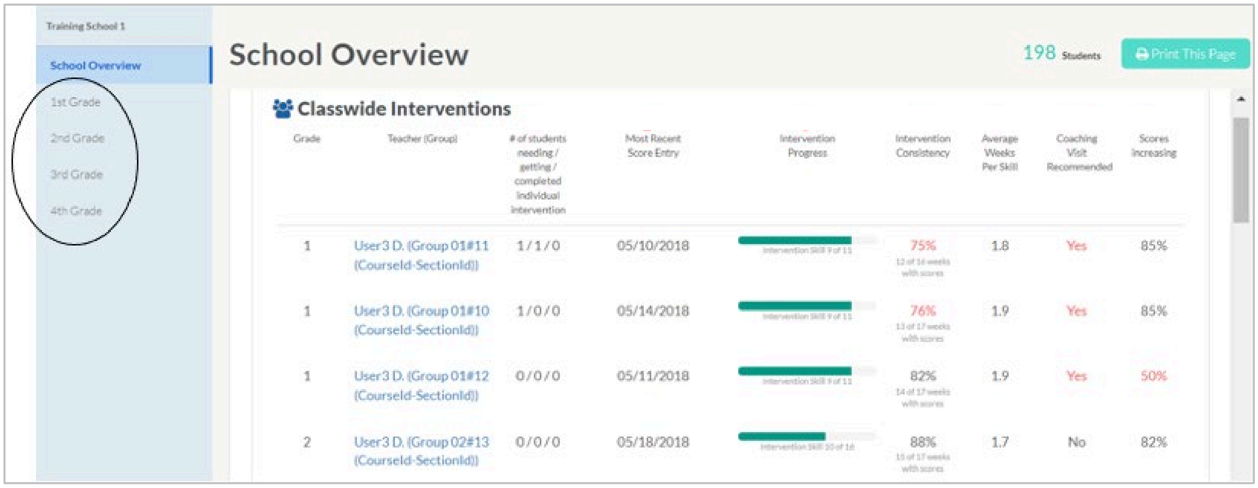


Figure 7.18. The coach can click on a grade level to open a dashboard.

Grade-level dashboards resemble the School Overview dashboard with a few notable differences.

The first section of the grade-level dashboard shows screening completion by teacher. Thus, if a grade has incomplete screening data, the coach can click on the grade level to see which teachers have and have not completed screenings. The incomplete screening is shown with the red "X" next to the teacher's name in Figure 7.19. Once all classes have completed screening the green Hide Progress box can be used to hide this section of the dashboard.

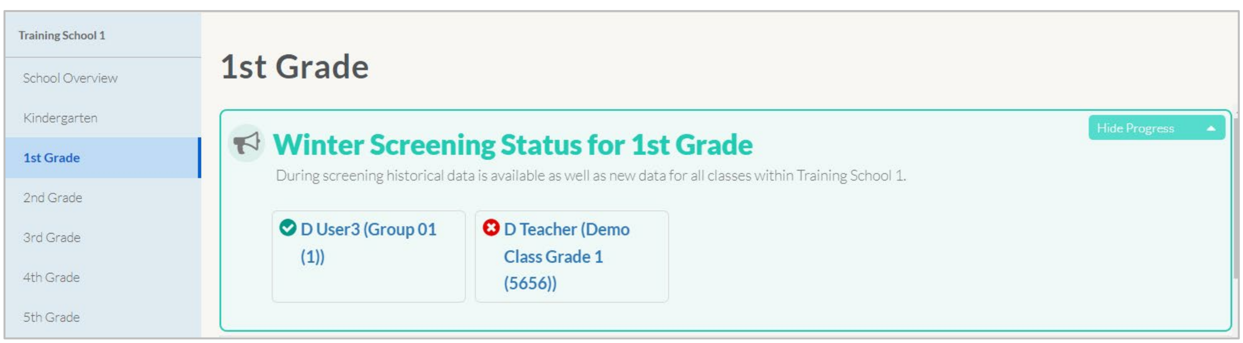


Figure 7.19. On a grade-level dashboard, coaches can see which teachers have completed screening and which have not.

The next section provides the Summary Notes for each grade (see Figure 7.20).

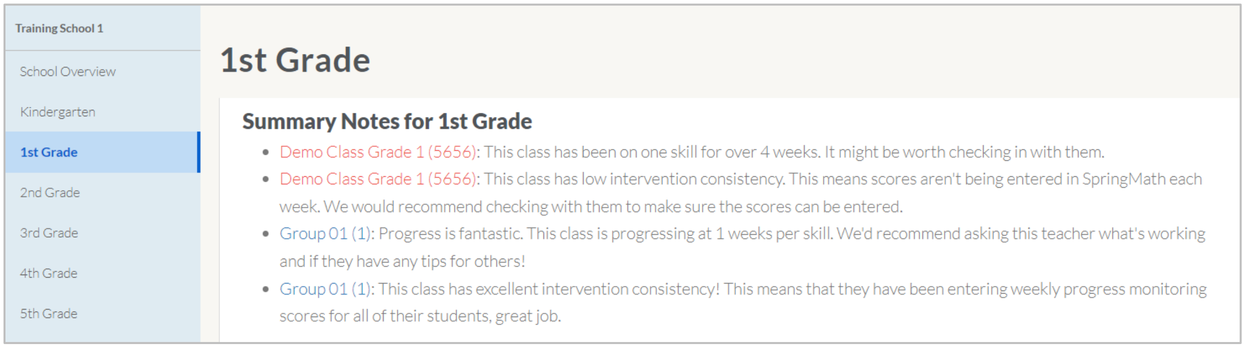


Figure 7.20. Here coaches can read the recommended coaching actions for the week for a given grade level. The coach can treat this list as a to-do list for the week

Summary Notes appear only when classwide intervention progress-monitoring data have been entered and SpringMath has detected a trend in the results. These notes provide SpringMath's recommendations for key coach actions to take given the current data. These notes may be especially helpful for coaches who are not yet facile with interpreting the metrics on the School Overview page and again on the grade-level dashboard pertaining to implementation progress and results.

Below the Summary Notes is a summary of classwide intervention progress, very similar to what is reported on the School Overview tab, but with some additional utility, which we will now explain (Figure 7.21).

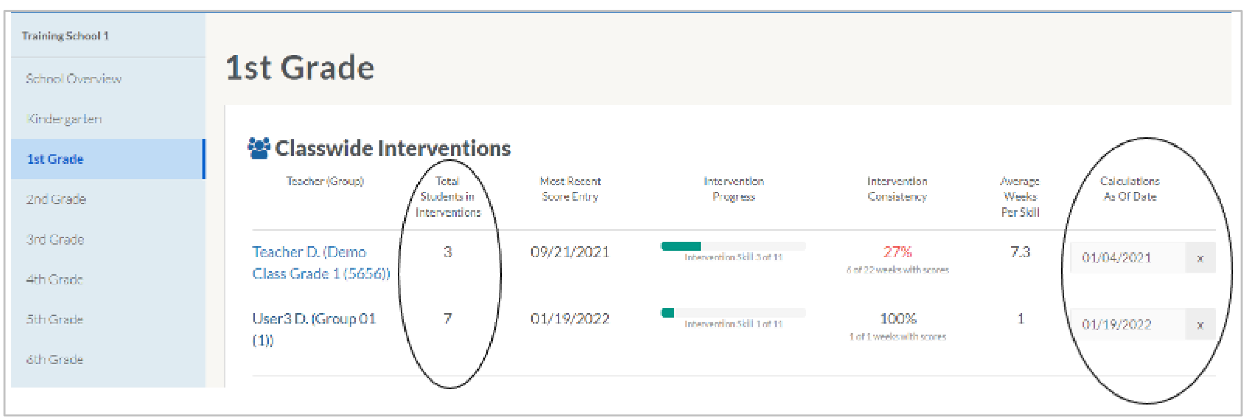


Figure 7.21. Total Students in Interventions tells the coach how many students are on a class roster that is receiving classwide intervention. One common implementation error that coaches can be vigilant about is teachers excluding some students from classwide intervention or not entering all students' scores on the final day of intervention for a given skill.

Each teacher in the grade level who is using classwide intervention will be shown here. Next to each student's name, the coach can see the number of students enrolled in classwide intervention. This is an important feature to determine whether all students in a class are being provided access to classwide intervention, which is the recommendation of SpringMath. Total Students in Intervention can also signify whether you have a group of students who represent a special grouping of students for intervention. Such groups often require coach attention to make sure that they attain the growth they need to drive more robust learning gains in math.

As recommended on the School Overview page, it's useful to verify that teachers have entered their most recent weeks' scores, then to examine Intervention Progress, Intervention Consistency, and Average Weeks Per Skill as previously described. Here it is especially easy to notice which classes are lagging in growth relative to other classes at the same grade level so that the coach can initiate a troubleshooting conversation. Finally, the Calculations As-Of-Date allows the coach to change the date from which Intervention Consistency and Average Weeks Per Skill are calculated. When a coach works with a teacher to improve consistency and quality of implementation, the coach can then reset the calculation date to the current date to reflect consistency going forward.

The next section provides notes for any students who are receiving individual intervention (Figure 7.22). From each note, the coach can click on the student name and SpringMath will open that student's individual intervention data. It is easy to return to the coach dashboard by clicking "Return to Overview" and then clicking on the grade level on the left. These notes are SpringMath's recommendations of key actions for the coach to take to support individual intervention progress that week in each grade.

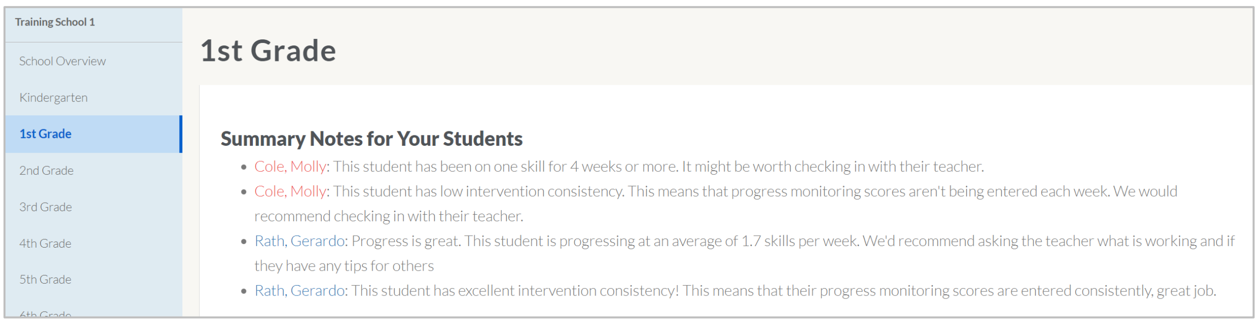


Figure 7.22. Individual interventions are summarized the same way and the coach can view which students are getting individual intervention in classes at a given grade level along with specific recommendations for coaching actions.

Below the summary notes for individual interventions, the coach can view progress for individual interventions (Figure 7.23).

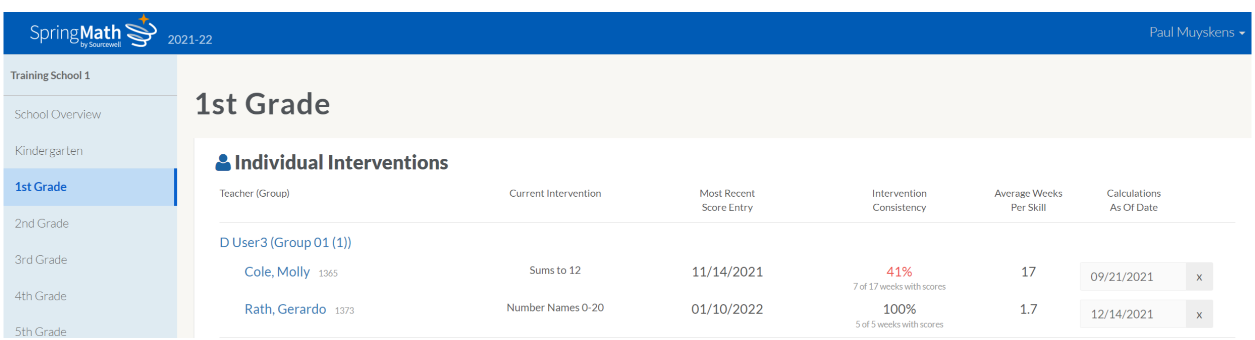


Figure 7.23. Under Individual Interventions, coaches can see what skill a student is working on, the most recent score entry date, intervention consistency, and the average weeks per skill. From this view, the coach can open the student's data directly by clicking on the teacher or student's name. In this example, the coach should visit Molly's teacher to troubleshoot this intervention, which is not being well implemented. Following a coach visit, the coach can reset the integrity metrics that occasioned in the in-class coaching visit.

Students are organized under each teacher's name. The coach can click on the teacher's name or the student's name to drop into the teacher's dashboard. Current Intervention is the skill that is currently being targeted via intervention. Next is the Most Recent Score Entry. Then Intervention Consistency and Average Weeks Per Skill are reported, followed by the ability to change the date from which the implementation metrics are calculated (e.g., following a coach visit).

The final (and best) part of the grade-level coach dashboard are the growth summaries. These are, in effect, aggregates from each teacher's data reported under Growth in the teacher dashboards. This chart provides the average percent of students who are at or above the instructional target on related screening measures in consecutive seasons and on the same classwide intervention skill that corresponds to the first screening measure. For example, in Figure 7.24, no students scored above the instructional target on the Sums to 6 measure at fall screening in this first grade. Following classwide intervention on this skill, 90% of the first-graders scored above the instructional target. These gains showed some generalization to the winter screening skill, which was Sums to 12. These graphs provide clear information about the progress made on the key skills for each grade level, and when gains are detected here, those gains signify that more distal progress indicators of math learning success will also show improvement.

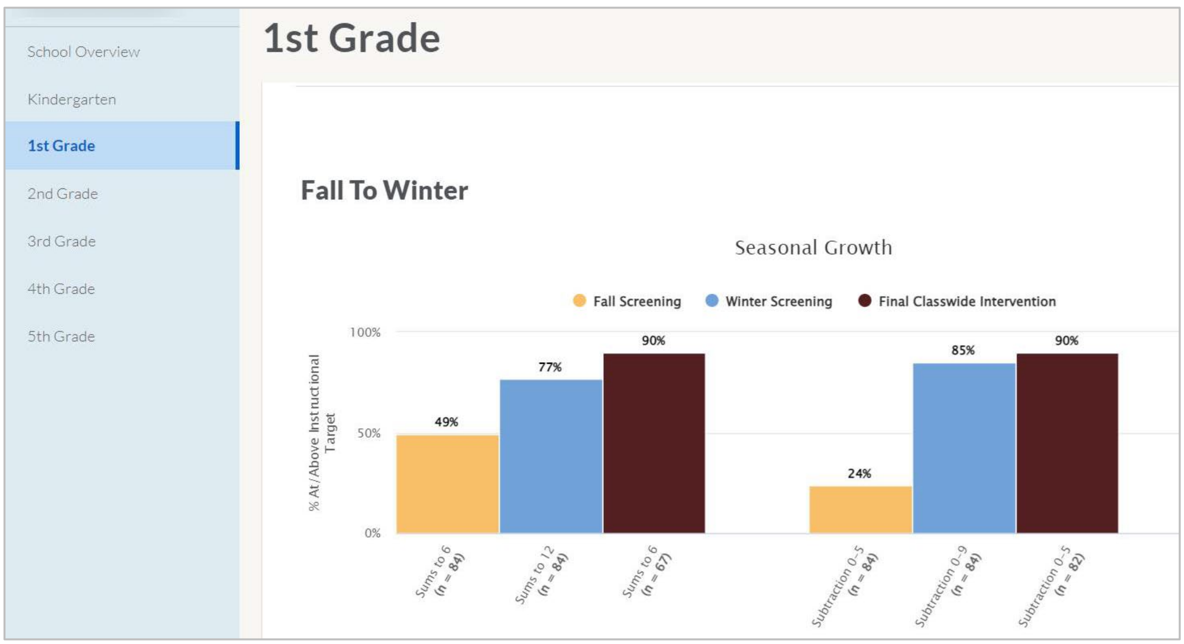


Figure 7.24. For all grade levels, coaches can see the percent of students who were instructional or higher on the fall screening and the final classwide intervention session for the same skill.

For most teachers, there is no more rewarding experience than to bring about learning gains for students. Antecedent supports are necessary, but performance feedback via a trusted coach is the most critical ingredient for any MTSS implementation effort. Performance feedback may sound frightening or heavy handed, but in fact, most teachers greatly appreciate the collaboration and support (although ideally the coach has the people skills needed to build and sustain trust and a culture of constant improvement). In classwide intervention, there are four common errors in implementation that coaches frequently need to deliver performance feedback on. These most common errors along with recommended coaching actions to correct are summarized in Table 7.4. Thus, using the coach dashboard to help teachers improve learning is ultimately a very rewarding experience for veteran and novice teachers alike. We often hear from systems that the work of implementing SpringMath produced a cultural transformation where teachers were inspired to collaborate and help each other to help all learners in the school grow.

Implementation errors and coach actions to improve

Not doing the intervention	Make intervention use fail-proof: Make sure teacher has materials. Make sure teacher knows HOW to implement. Make sure there is a scheduled time for intervention.
Students do not know how to follow the classwide intervention routine	Help the teacher retrain the students. Show the students how to get into working pairs, how to use the materials, how to provide high-quality feedback, and how to be engaged.
Teacher is not completing all steps of the intervention	Show the teacher the missed steps and provide the rationale. Papers must be scored during the intervention because that provides feedback to the student, provides the error correction opportunity, and provides goal attainment opportunity. The error correction component is important because it improves student accuracy for the next session.
Children seem bored with the intervention	Help the teacher include rewards to motivate students. Help the teacher display the median graph on his or her dashboard for the class to see their growth. Be sure the teacher is setting daily goals with the students!

Table 7.4

7.4 Support and training resources

SpringMath was designed to provide everything that a teacher needs to intensify math instruction efficiently and effectively in the classroom. Because many of the ingredients built into SpringMath can be more broadly useful to and used by the teacher for core instruction, the support portal has been designed to give access to these resources. In Support, teachers and coaches can find short how-to videos, videos of classes and students implementing SpringMath, printable resources such as number cards for playing games and progress-tracking sheets for students, implementation guides to get started and troubleshoot common challenges. Skill sequences for classwide intervention are provided along with static versions of all assessments and word problems and games. To access Support, the teacher or coach can click on their name in the top left corner and select Support from the drop-down menu. Resources are organized in the following categories on the main landing page for Support (Figure 7.25).

Figure 7.25. Teachers and coaches can access support items under each of these headings in Support.

Flexible Learning Plans and Resources

Flexible Learning Plans and Resources provides resources related to implementing SpringMath in scenarios where students are working remotely or in hybrid instructional models. Instructional calendars for each grade are also provided along with instructions about how to use the instructional calendars to guide or supplement core math instruction. For each grade, a 36-week calendar is provided that specifies skills to be mastered each week of instruction and provides assessments to assess skill mastery for that skill, a whole-class acquisition lesson to teach the skill, and a fluency-building lesson for that skill.

Resources for Teachers

In this section, we provide a quick-start guide for teachers, which includes video and printed documents and a letter to send home to parents at the start of implementation. How-to resources are provided for teachers for screening, classwide intervention, and individual interventions. Under Classwide Intervention, you can find whole-class acquisition lessons to use when teaching a new skill during core instruction, booster lessons to aid learning loss recovery, or just boost learning in math in your class, and game-like progress, tracking sheets for students to track their individual gains during classwide intervention. You can also find the skill sequences for each grade level (in case you are curious). SpringMath high school guides are provided here. Finally, under each section, training and troubleshooting guides and checklists are provided to promote effective use of SpringMath.

Resources for Coaches

In this section, we provide training materials to be used at the school level (slide decks, training videos), coach checklists, and written guides designed to help coaches provide consultation and leadership for their school's implementation. In this section, coaches can find all the how-to advice to help them effectively coach teachers and cultivate more effective SpringMath use.

Resources for Data Administrators

This section provides how-to guides and resources for those who set up or upload rosters, manage student rosters, and provide coach-level access to their staff. Data administrators are also able to edit recently entered scores.

Research Behind SpringMath

This section provides background information on the creation of SpringMath, explains the logic behind the functionality, and offers examples of results that districts are achieving. Much of this information is also covered in this technical manual.

Release Notes

This section documents all the changes that occur for each new version of SpringMath. These brief descriptions cover new features, bug fixes, and behind-the-scenes changes.

Assessment Materials

This section provides a list of the screening measures by grade and season and 10 versions of each assessment by grade and by season for all assessments contained within SpringMath.

Newsletters and FAQs

Prior versions of the SpringMath newsletter appear here along with some FAQs around procedural fluency, target setting, and accommodations.

Word Problems and Games

In this section, we provide an extensive collection of word problems, games, and flashcards that correspond to the SpringMath interventions. For example, for classwide intervention skills, a total of 993 unique word problems are available. Each grade has access to on average 225 (range, 126 to 333) word problems to use for associated classwide intervention skills. For nearly all skills in SpringMath, we provide 9-11 word problems that can be used to practice problem solving. Games are provided for nearly all skills in SpringMath as well, and these can be found here.

References

Burns, M. K., & Gibbons, K. (2012). Response to intervention implementation in elementary and secondary schools: Procedures to assure scientific-based practices (2nd ed.). New York: Routledge.

Bushell, D., Jr., & Baer, D. M. (1994). Measurably superior instruction means close, continual contact with the relevant outcome data: Revolutionary! In R. Gardner, III, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshleman, & T. A. Grossi (Eds.), Behavior analysis in education: Focus on measurably superior instruction (pp. 3-10). Monterey, CA: Brooks/Cole.

Daniels, A. (2013). Feedback in education: On whom and for what. In R. Detrich, R. Keyworth, & J. States (Eds.), Advances in evidence-based education, Volume 3, Performance feedback: Using data to improve education performance (pp. 77-96). Oakland, CA: The Wing Institute.

Education Research & Consulting, Inc. (2013). SpringMath (version 2.2.1) Computer Software. Sourcewell. springmath.org

Fixsen, D. L., & Blasé, K. A. (1993). Creating new realities: Program development and dissemination. Journal of Applied Behavior Analysis, 26, 597-615. doi.org/10.1901/jaba.1993.26-597

Gilbertson, D., Witt, J. C., Singletary, L. L., & VanDerHeyden, A. M. (2007). Improving teacher use of interventions: Effects of response dependent performance feedback on teacher implementation of a peer tutoring intervention. Journal of Behavioral Education, 16(4), 311–326. doi.org/10.1007/s10864-007-09043-0

Heifetz, R. A. (1994). Leadership without easy answers. Cambridge, MA: Harvard University Press.

Joyce, B. R., & Showers, B. (1981). Transfer of Training: The Contribution of "Coaching." Journal of Education, 163, 163-172. doi.org/10.1177/002205748116300208

Mortenson, B. P., & Witt, J. C. (1998). The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. School Psychology Review, 27(4), 613–627. doi.org/10.1080/02796015.1998.12085942

National Implementation Research Network. nirn.fpg.unc.edu/national-implementation-research-network

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L., Resetar, J. L., & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. School Psychology Review, 34(1), 87–106. doi.org/10.1080/02796015.2005.12086277

Schein, E. H. (2013). Humble Inquiry: The Gentle Art of Asking instead of Telling. San Francisco, CA: Berret-Koehler Publishers, Inc.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. Journal of Applied Behavior Analysis, 10, 349-367.

VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. American Education Research Journal, 49, 1251-1284. doi.org/10.3102/0002831212462736

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. School Psychology Quarterly, 13(2), 141–154. doi.org/10.1037/h0088978

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. Journal of Applied Behavior Analysis, 30(4), 693-696. doi.org/10.1901/jaba.1997.30-693

Learn more or set up a demo at

springmath.org

Prepared by SpringMath by Sourcewell