The Next Generation of Response to Intervention

Using data-fueled insight, decisions, and resources to respond to challenge, intervention, and growth
Reports and software screens may vary from those shown as enhancements are made.

©Copyright 2016 by Renaissance Learning, Inc. All rights reserved. Printed in the United States of America. All logos, designs, and brand names for Renaissance’s products and services, including but not limited to Accelerated Reader 360, Renaissance, Star, Star 360, Star Assessments, Star Custom, Star Early Literacy, Star Math, and Star Reading, are trademarks of Renaissance Learning, Inc., and its subsidiaries, registered, common law, or pending registration in the United States. All other product and company names should be considered the property of their respective companies and organizations.

This publication is protected by U.S. and international copyright laws. It is unlawful to duplicate or reproduce any copyrighted material without authorization from the copyright holder. For more information, contact:

RENAISSANCE
P.O. Box 8036
Wisconsin Rapids, WI 54495-8036
(800) 338-4204
www.renaissance.com
educatordevelopment@renaissance.com

08/18
Contents

1 Foreword
3 Introduction
4 Response to Intervention: A brief history
6 What we have learned
7 RTI in a next-generation world
8 Lessening the distance between assessment and instruction
10 Screening in high school
11 Renaissance Star Assessment® screening data and specific learning classifications
12 Using technology advancements to interact with vivid, visual data
13 Response to information
16 Goal setting, growth, and progress monitoring
18 Conclusion
19 References

Figures

5 Figure 1. RTI pyramid
9 Figure 2. The Star test blueprint works with CAT technology to determine item presentation
10 Figure 3. When selecting state benchmarks in Star Assessments, categories and cut scores are unique to each state
12 Figure 4. Sample view of data in Reading Dashboard
Foreword

By James E. Ysseldyke, Ph.D.

Since the 1970s, I have witnessed a struggle between two competing paradigms in assessment, instruction, and intervention practices in school settings. In that time, I have also witnessed and contributed to a major paradigm shift from a focus on inputs and professional processes to increasing emphasis on outcomes and benefits to students. The model of practice has changed from one in which indications of high-quality services ranged from number of assessments given, insightful interpretations, and general prescriptions for classification and placement to delivery of a multi-tiered system of supports (MTSS) emphasizing prevention, early identification and treatment, and intensive evidence-based interventions.

This represents a major shift from a correlational science focused on standardized testing, prediction, and placement to an experimental science focused on enhancing the competence of individual students through interventions designed to maximize learning and guided by short-run empiricism (i.e., problem solving and response to intervention). Such practices are characterized by careful assessment of students’ current academic skill standing, setting appropriate instructional goals, engaging in intensive direct instruction, monitoring student progress continuously, and adapting instruction based on feedback. Renaissance® has played a major role in the development of evidence-based processes, procedures, and products designed to aid this shift and to accelerate learning for all children of all ability levels and ethnic and social backgrounds, worldwide.

Teachers, school psychologists, school executives, and others use assessment information to make many kinds of educational decisions about students. Primary among these is a set of instructional decisions: deciding what and how to teach individual students. It is critical that school personnel focus on “assessment that matters.” There are three kinds of assessments that occur in school: (1) those that matter but are technically inadequate, (2) those that are technically adequate but do not matter, and (3) those that are technically adequate and matter. In my professional writing, I have repeatedly argued that unless our assessment efforts lead to instructional changes resulting in improved academic outcomes for students, the assessments do not matter. Assessment that is related to and supports the development of effective interventions is worthwhile and is clearly in the best interests of individuals, families, schools, and society.

For years, assessment practices in schools were dominated by a correlational paradigm in which efforts were made to identify and describe deficits, dysfunctions, disabilities, and disorders that were presumed to cause learning difficulties. Federal efforts to identify children with learning disabilities (LD), for example, required a significant discrepancy between ability and achievement commensurate with a disorder in one or more psychological processes thought to underlie learning. When researchers (including much of my work with colleagues at the University of Minnesota Institute for Research on Learning Disabilities) demonstrated the psychometric shortcomings of a discrepancy approach, the failures of a process or...
deficit perspective, and the absence of evidence for aptitude x treatment interactions, there was a move
toward an experimental paradigm and the use of information from student response to evidence-based
treatments or interventions in instructional planning. Use of the terminology Response to Intervention
(RTI) or Response to Instruction and Intervention (RTII) began in the professional literature in the 1990s.
RTI practices were formally permitted as an alternative in LD identification in the Individuals
with Disabilities Educational Improvement Act of 2004. Both of these developments prompted
significant shifts:

• From making predictions about students’ lives to making a difference in their lives
• From a deficit perspective that identifies weaknesses to a resilience perspective focusing on
  an individual’s strengths and on ways to modify instruction to remove barriers and increase the
  probability of success
• From sifting and sorting (classification) to multi-tiered serving

This next chapter of RTI focuses on making the link between assessment and instruction/intervention
even more direct and relevant, using specific goal-setting procedures, links to instructional hierarchies,
 improved progress-monitoring practices, and growth modeling, as well as taking into account ecological
 factors like mindfulness.

Renaissance® has been at the forefront of these changes, providing educators with products, processes,
 and professional development to fit improved practices in an RTI paradigm. The computer adaptive
 assessment practices and procedures described in this paper, as well as the growth models advocated,
 signify the present and future of progress monitoring within RTI. An RTI framework represents best
 practices in making assessment decisions that matter, and this next generation allows us to continue
 engaging in assessment practices that will make a difference in the lives of all the students we serve.
Introduction

Response to Intervention (RTI)—a key component within the broader multi-tiered system of supports (MTSS)—finds its roots in the late 1970s as an alternative way to identify students with specific learning disabilities (SLD). RTI challenged the established discrepancy model that compared student achievement with performance on IQ measures in determining eligibility for special education services (National Education Association, n.d.).

The discrepancy model is based on a normal curve and requires a substantial difference (i.e., discrepancy), between a student’s IQ measure and scores from one or more areas of academic achievement. The accepted “substantial difference” for identification of special services is at least two standard deviations. Additionally, the model requires a disorder in one or more of seven identified processes. According to Ysseldyke (2009), this addition led practitioners to focus on identification of disorders at the expense of skill deficit identification and remediation, a move that fostered the processing strengths and weaknesses (PSW) approach that remains today as an alternative to RTI.

There are practical concerns related to the discrepancy model. Of particular concern is that it places greater focus on disorders over skill deficits, and thus rarely identifies children in early grades. As a result, students may struggle for some time before they are identified for services. In other words, children must fail—and fail for a while—before they receive specialized instruction. The model also does not identify students who may need immediate, but temporary, intervention to address an experiential or developmental gap rather than a long-range placement. Further, IQ and achievement metrics do not always consider the quality of instruction, whereas RTI asks, “Would this student benefit from a different approach?” Finally, using IQ and achievement results works to identify, but falls short in informing, the instructional process (The IRIS Center, 2007).

While these concerns are substantive, the most important reason for moving away from the discrepancy model was based on significant empirical support—most notable, the model’s lack of statistical reliability and the lack of empirical support for the effectiveness of interventions based on the model (Aaron, 1997; Fletcher et al., 1998; Francis et al., 2005; Stuebing, Fletcher, Branum-Martin, & Francis, 2012).

In part to address these concerns, the RTI model provided an alternative to established practices, including processes to:

- Screen all students to understand where each is in the progression of learning in reading, math, or other disciplines
- Identify students for intervention at the earliest possible moment before misconceptions become habits and repeated struggle feeds a fixed mindset
- Intervene early—in the early grades as well as early in conceptual development for disciplines throughout all grades
- Inform the instructional process based on empirically validated evidence of what students know and what they are ready to learn
- Provide the right resources for learning at the right point in instruction

RTI gained nationwide attention in 2004 when the reauthorization of the Individuals with Disabilities Education Act (IDEA) included this model for early intervention and provided funding in 2005 to initiate RTI...
in districts across the nation. At Renaissance, we have long believed that RTI is potentially the most promising educational development in many years—if it is understood and implemented in the right ways. We base this belief on more than two decades of hands-on experience with the essential element of RTI: educators who use student data in the classroom to accelerate learning for all (Renaissance, 2011).

While researchers were wisely investing more than a quarter century observing data-driven instruction for the purpose of improving education outcomes for all students, Renaissance was seeing this essential element work, time and again, for students of all ethnicities and socioeconomic status, at all grade levels in reading and mathematics. It is what we do: Our primary purpose is to accelerate learning for all children and adults of all ability levels and ethnic and social backgrounds, worldwide.

This paper acknowledges the RTI legacy established by those researchers who took the lead and guided our schools in data-fueled instruction. As a part of this legacy, Renaissance worked in concert with these leaders to develop assessments proven to be reliable, valid, and efficient. Most importantly, we engineered RTI technology tools and resources that lessened the distance from assessment to instruction.

This paper looks ahead to the next generation of RTI, noting the subtle, and not-so-subtle, shifts in focus from the first implementations in 2004 to what we see now in US schools well into our second decade in 21st century classrooms. This information within assumes readers have fundamental knowledge of RTI, the traditional three-tiered model, and the screen, intervene, and monitor-progress protocol. For in-depth information on the foundations of RTI, please see Making RTI Work: A Practical Guide to Using Data for a Successful Response to Intervention Program (Renaissance, 2011) and New Thinking in Response to Intervention (Shapiro, 2012). Let’s draw on lessons learned in the first generation of RTI and explore ways the next generation offers greater promise for each student, teacher, and educational leader.

Response to Intervention: A brief history

As noted by Salvia, Ysseldyke, and Witmer (2017),

The basic conceptual framework for RTI has existed in the psychological and educational literature for many years, with its foundation in the prevention sciences (Caplan, 1964), where physicians talked about primary, secondary, and tertiary prevention or treatment. In education and psychology, the concept likely originated in the early work of Lindsley (1964) on precision teaching, and it was first implemented as an assessment model by Beck (1979) in the Sacajawea Project in Great Falls, Montana. There are many models of RTI (Jimerson, Burns, & VanDerHeyden, 2016; National Association of State Directors of Special Education, 2005; Sugai & Horner, 2009) but they all share (1) multiple tiers of effective intervention service delivery, (2) ongoing progress monitoring, and (3) data collection/assessment to inform decisions at each tier. (pp. 174–175)

1970s–1980s

Some of the techniques used in most RTI implementations go back more than 30 years. In the 1970s and ‘80s, researchers such as Stanley Deno and Phyllis Mirkin (1977) found that short, frequent assessments helped manage Individual Education Plans (IEPs) for students in special education. Around the same time, Benjamin Bloom’s (1981) “mastery learning” experiments demonstrated that using formative assessment as a basis to modify curriculum and instruction dramatically improved average student performance.
1990s
Gresham (1991) likely was the first to use the initials RTI. The three-tier structure, often represented with a pyramid, originated early in the '90s with researchers like Sugai and Horner (1994), who were seeking ways to deal with behavioral problems in general education settings (see figure 1). In 1998, Black and William’s meta-analysis further documented how using assessment results to set goals and determine interventions improved performance and was particularly effective in reducing achievement gaps between subgroups.

2000s
As RTI was widely implemented, limitations associated with the discrepancy model (revisit Introduction, p. 6) were evident. During this time, Fuchs (2003) developed the dual-discrepancy model which examined each student’s level of achievement, expectations of the discipline, rate of student growth required to hit benchmarks in a timely manner, and response to instruction or intervention. It was the dual-discrepancy model and the noted success of tiered interventions that attracted federal funding in the amendments to the 1997 IDEA, with implementation funding available in the 2004 reauthorization and its alignment with No Child Left Behind (NCLB). In other words, RTI became an established program in schools across the nation.

Although RTI was established in the late 1970s, we refer to implementations since the early part of the 2000s as “the first generation of RTI”—in other words, the first fully funded and broadly implemented RTI programs. The following is a list of a few of the key areas of focus commonly observed in the first generation of RTI implementations:

• Research: Educational leaders sought research-based tools for RTI. Assessments for screening and progress monitoring required evidence of reliability and validity. Likewise, intervention programs met equally rigorous standards. Educators accessed independent and federally funded online tools charts from the National Center on Response to Intervention (NCRTI: www.rti4success.org) and the National Center on Intensive Intervention (NCII: www.intensiveintervention.org) to review reliability and validity ratings for screening tools and intervention programs.

• Technology: Schools retrofitted labs and expanded student access to computers to keep up with data-generation requirements, which aligned with Kurns and Tilly’s observation that “in the absence of technology, the data burden becomes unmanageable” (2008).

• Students’ response to intervention or instruction: The overarching purpose of RTI is to improve student outcomes. Monitoring students’ responsiveness to intervention or instruction and making instructional decisions based on monitoring data help educators realize that purpose. Educators seeking to implement the tools reviewed by NCRTI and NCII found research and guidance from RTI experts at the RTI Action Network.2

• Achievement: Successful interventions reported achievement gains toward established benchmarks.

• Reading: Some of the earliest academic implementations of RTI focused on reading at the primary grades (National Association of Special Education Teachers, 2006).

• Elementary: Typically, educators found a greater degree of scheduling flexibility in the elementary grades. Compartmentalized classes in middle school and the structure of high school schedules

---

1 NCRTI, which was federally funded through 2012, is now maintained and supported by American Institutes of Research (AIR), a nonprofit, nonpartisan research and technical support organization. As of press time, the NCII is funded by the U. S. Department of Education’s Office of Special Education Programs, and like NCRTI, is managed by AIR.

2 Per www.rtinetwork.org, the “RTI Action Network, part of the National Center for Learning Disabilities, is funded by the Cisco Foundation and in partnership with the nation’s leading education associations and top RTI experts.”
built around courses rather than subjects or grades confounded first-generation efforts to provide assessment, intervention, and progress monitoring at the secondary level.

RTI thrived, in part, because of the specificity of these focus areas as they became implementation building blocks: find what works, retrofit facilities to manage the data load, screen, plan for interventions, and monitor the progress of students engaged in the interventions. The day-to-day business of teaching and learning became more quantified, systematic, diagnostic and prescriptive, as well as more observable and replicable from classroom to classroom. RTI has endured and is now adapting to a new era in our schools—one of next-generation standards, 21st century skills, greater attention to local assessment concerns, and global achievement metrics.

What we have learned

We have learned that RTI requires continuous assessment but is ultimately about lessening the distance between assessment and informed instruction. According to Ysseldyke (2009), assessments that do not inform instruction do not matter and are of heuristic value only. He believes we have moved from using assessment data to make predictions about students’ lives (such as who might succeed in school) to using these data to make a difference in students’ lives by improving instruction, enhancing competence, and leading them to positive outcomes.

We have learned that the academic partnership between a teacher and a learner is the dominant achievement variable in both core instruction and intervention (Barber & Mourshed, 2007; Chetty et al., 2011; Coleman et al., 1966; Wright, Horn, & Sanders, 1997). We know we must acknowledge the impact of influences on achievement outside of school, such as homelife, peers, and socioeconomic status; however, researchers have found there is greater variability of effectiveness within schools than between schools (Nye, Konstantopoulos, & Hedges, 2004). As Coleman et al. explained, “Variations in facilities and curriculums of schools account for relatively little variation in pupil achievement. The quality of teachers shows a stronger relationship to pupil achievement” (p. 22). Further, he observed this impact was progressively greater at higher grades, indicating a cumulative impact of effectiveness in fostering student achievement.

We are continuing to learn that academic partnerships grounded in growth mindsets (Dweck, 2007) are critical to successful RTI implementations. The plural “mindsets” is intentional. In addition to the student mindset, teachers’ beliefs about their students’ abilities to achieve can impact not only instruction, but also the level of student engagement and motivation (Dweck, 2015; Farrell, March, & Bertrand, 2015).

Thanks to advancements in brain imagery, we now see the physical association between mindset and mastery. A Stanford University study using functional magnetic resonance imaging (fMRI) found that subjects who scored higher on an assessment of positive mindset showed more brain activity throughout areas associated with problem solving and recall. Likewise, a lower positive mindset was associated with lower math performance. As Stanford researcher Lang Chen stated, “Beyond the emotional or even motivational story of positive mindset, there may be cognitive functions supporting the story” (Sparks, 2015).

Documenting achievement gains, such as an uptick in a student’s percentile rank or grade-level equivalent score, is key to any RTI implementation. We have learned that growth metrics provide context for those gains. The student growth percentile (SGP) score (see Betebenner, 2011) is a pre–post metric that reports a student’s academic growth over time relative to the academic growth, over the same period of time, of peers with similar histories of achievement or struggle. This is distinct from a single point-in-time
achievement metric that reports academic growth relative to all peers within a grade level, at that single point in time, regardless of established score history. (For more on SGP, see Goal setting, growth, and progress monitoring, p. 21).

The need to understand both sides of a learner’s performance is supported by the 2015 Every Student Succeeds Act (ESSA), which authorized use of “student achievement, which may include student growth” as evidence of student progress and state accountability (in ESSA, see pp. 33, 35, 41, 79, 119, 126).

All that we have learned deepens our understanding of student potential and the power of RTI in the hands of skillful educators. Now let’s look at the next generation of RTI implementation and explore ways to apply what we have learned.

RTI in a next-generation world

The implementation of next-generation standards has generated greater public interest in what students learn, the ways they are taught, and how—and how often—their learning is assessed. This next generation of RTI is firmly rooted in the foundational principles established by RTI leaders, and those foundational principles have yielded substantial evidence of effectiveness.

In a synthesis of more than 20 years of education research, educational researcher John Hattie (2012)\(^3\) and his team used a statistical model to determine effect sizes of programs and practices associated with student achievement.\(^4\) RTI is ranked third among 150 influences on achievement in Hattie’s work, with an effect size of \(d = 1.07\) (p. 266). The magnitude of RTI’s effect size implies that administrators, teachers, and students engaged in a skillfully implemented RTI program can expect consistent student gains, which, over time, hold the potential to make a significant difference in students’ lives.

The concept of a “skillfully implemented” RTI program is critical in achieving this much improvement. In fall 2015, the Institute of Education Sciences published an evaluation of RTI practices for reading in grades 1–3. Among a large subset of schools in the study, RTI implementations did not return expected results. Researchers noted variances in RTI practices among these schools, including placing all students, regardless of reading level, into an intervention program (to address scheduling concerns) and providing intervention services during—rather than in addition to—core instruction. Study authors Balu et al. noted these factors may have “displaced instruction time and replaced some small-group or other instruction services with intervention service” (p. 11). As the Balu et al. study suggests, RTI programs lacking in implementation integrity are at risk to fall short of the \(d = 1.07\) ranking reported by Hattie (2012).

Implementing an RTI model with integrity relies on strong, differentiated Tier 1 (core) instruction, effective interventions delivered in Tier 2 (small group) targeted instruction, and accurate use of assessment data for universal screening, informing instruction, and engaging teachers in progress monitoring. In the next generation of RTI, educators seek assessments with the greatest impact on informing instruction and the least impact on instructional time.

---

3 While Hattie’s work is widely accepted, other equally impactful education researchers have published concerns related to his findings and inferences drawn from them. For commentary in support of, and in question of, Hattie’s work see Kamenetz (2015) and Kukic (2013).

4 An effect size reflects the degree of change in relation to the standard deviation (SD). An effect size of 0.5 (\(d = 0.5\)) reflects a difference of 0.5 SD—equivalent to 1 year of growth.
Lessening the distance between assessment and instruction

As a response to concerns about over testing in American schools, the Council of Chief State School Officers (2015) published a framework for state agencies to improve the quality of assessment in schools and reduce the burden of testing time—in particular, the loss of instructional time. Reclaiming instructional time is a complex endeavor that cannot be achieved solely by implementing strict limits on numbers of assessments or time set aside for them. Focusing on numbers and hours may reduce assessment time, but the issue at hand is reclaiming instructional time while increasing the impact of that instruction. Additional instructional time is positively correlated with achievement gains; however, the quality of curriculum, instruction, and student engagement affects the degree of achievement gains (Baker, Fabrega, Galindo, & Mishook, 2004; Batsche, 2007; Gallup, 2014; Jez & Wassmer, 2015).

RTI requires a lot of assessment, but essentially it is about using assessment data to make informed decisions about curriculum, instruction, and student engagement. Effective RTI implementations include data from multiple measures, which may include both traditional and adaptive assessments.

Traditional tests
Briefly stated, a traditional (fixed-form) test is built from a selected set of skills. Once these skills are identified, typically 5–7 items are developed for each skill represented on the test. With a fixed-form test, all students respond to the same items delivered in the same sequence. Curriculum based measures (CBMs) are traditionally designed, fixed-form tests that can play an important role in measuring specific skill fluency in reading and mathematics. CBMs allow for brief, one-skill-at-a-time measures. By design, CBMs use a handful of items and limit the number of skills assessed.

Because both traditionally designed fixed-form and CBM measures use a predetermined set of skills, they may limit the ability to quickly and efficiently access information about students’ achievement levels beyond the skills represented on each assessment (Catts et al., 2009). To gain adequate information to meet the learning needs of students who did and did not demonstrate mastery of the skills represented on a test, educators must administer additional assessments. However, this becomes impractical as it may negatively impact instructional time and actually increase the distance between assessment and instruction.

Computer adaptive tests (CATs)
A computer adaptive test adjusts the difficulty of each student’s individual assessment by selecting items or modules based on the student’s earlier performance during the test. CATs, including Renaissance Star Early Literacy®, Renaissance Star Reading®, and Renaissance Star Math®, typically use item response theory (IRT) as the basis to match a student’s knowledge or skill level to adaptive item selection and scoring. IRT relies heavily on a highly calibrated scale, placing the ability of the student and the difficulty of items on the same scale (Renaissance, 2014).

Typically, each question in a CAT has ample response time so test scores are based on the accuracy, rather than the speed, of each student’s response. Because each student receives a set of items near his or her own ability level, CATs can be two or more times as efficient as traditionally designed tests in measuring student achievement. In this way, adaptive tests play a critical role in reclaiming instructional time; however, it is the information gained from the adaptive test that brings added value to increased instructional time.

According to Shapiro (2012), with an adaptive test, student responses “cue shifts in subsequent items and result in an indication of the skills attained and not attained across an assessment domain.” By design, an adaptive test quickly pinpoints each student’s academic achievement level across a developmental continuum.
Renaissance, in consultation with nationally recognized experts, has created learning progressions for reading and math that present the developmental continuum for both disciplines. Empirical testing has found a strong statistical link between the progression of skills in these learning progressions and the assessed difficulty level of Star Assessment test items, meaning educators can use scores from the assessments to identify both what students know and what they are ready to learn. A Star assessment’s blueprint works with CAT technology to determine which items are presented to each student. While each Star test event is unique, the blueprint ensures that a certain number of items from the domains and skill sets are presented to each student (Renaissance, 2014; see figure 2). It is the statistical link between the learning progressions and Star that allows educators to increase the impact of instruction and significantly lessen the distance between assessment and instruction.

Figure 2. The Star test blueprint works with CAT technology to determine item presentation

Computer adaptive tests emerged in the first generation of RTI as an effective and efficient option for screening and progress monitoring. Based on evidence of the reliability, validity, and sensitivity of Star Assessments® to student growth, these assessments continue to make an important contribution to RTI in this next generation of data-fueled decision making (Johnson & Thompson, 2016; Kovaleski, VanDerHeyden, & Shapiro, 2013; Shapiro, 2012).

Screening in a next-generation implementation

Screening in the next generation of RTI holds the same overarching purpose as screening in the first generation; that is to consider assessment data in decisions about how to best serve students, whether they are working at or above grade level or are in need of support. In the first generation of RTI, many states and researchers interpreted “working at grade level” and “proficient” to generally correspond with the 40th percentile (Renaissance, 2013). It is important to note that the 40th percentile is widely considered the minimum benchmark, or entry-level score into proficiency (Renaissance, 2013); however, as noted by Mellard and Johnson (2008), “because screening does not directly result in a diagnosis, it is better for a screening instrument to err on the side of false positives or identify students as at risk that might not be at risk.”

While the benchmark should not be set lower than the 40th percentile regardless of the challenges associated with current low performance across the campus or district (Shapiro, in Renaissance, 2013), there is no established limit as to how high a proficiency benchmark can be set. In high-performing districts or campuses, the 50th percentile or higher might be reasonable. Summative assessments designed to measure proficiency on next generation, more rigorous standards often have a proficiency threshold that equates to the 60th percentile or higher—in some states, significantly higher.

In the next generation of RTI, educators equipped with insight about students, standards expectations, and rigors of the summative exam may adjust the default screening benchmarks and cut scores in Star. By default, Star Assessments use four categories that correspond to established cut scores: At/Above Benchmark (40th PR or higher), On Watch (25th–39th PR), Intervention (10th–24th PR), and Urgent Intervention (below 10th PR). States that extend RTI to students identified for advanced placement may...
consider adding a fifth benchmark category, Above Benchmark (cut score determined by campus or district). Figure 3 shows sample report data from the same fifth-grade Star Reading screening event; however, the benchmark structure and cut-score selection impact how these data are displayed and may influence decisions about instructional grouping.

Figure 3. When selecting state benchmarks in Star Assessments®, categories and cut scores are unique to each state

Data reflect traditionally accepted RTI benchmark categories and cut scores: At/ Above benchmark, On Watch, Intervention, and Urgent Intervention.

Data reflect a fifth benchmark category and cut score for Above Benchmark.

Data reflect benchmark categories and cut scores that correspond to the state summative exam.

Screening in high school

In the first generation of RTI, the focus was on reading in the primary grades; however, RTI has now moved from a program to a way of doing business, and this business extends across the content areas and well into high school. Implementing RTI at the high school level includes essential components, including high-quality, differentiated core instruction, universal screening, continuous assessment to fuel decision-making, and tiered interventions (Windram, Ballmer, & Johnson, 2011); however, effective practice will differ from a primary implementation due to structural and organization differences. Regardless of these differences, the overarching purpose of RTI at all grade levels is to improve student performance.

In a high school implementation, the focus may be on all students, or a subset of 9th–10th grade students in reading and mathematics (see Ehren, 2009; U.S. Department of Education, n.d.). Another high school may implement RTI across the content areas and extend into non-cognitive areas, such as behavior, attendance, and attitudes about learning (Duffy & Scala, 2012). In the latter example, general outcome measures, such as Star Reading and Star Math, would be used with quantitative measures of attendance and graduation rate, as well as qualitative measures of attitude and mindsets.

Further, high schools may vary in their approaches to screening. In high school, the term “universal screening” includes variations on what counts as universal. Within the literature, multiple protocols for universal screening have been noted (Duffy, 2007; Duffy & Scala, 2012; Hughes & Dexter, n.d.). This paper focuses on RTI for reading and math, with the following protocols: (1) screen all grades three times per year in reading and math, (2) screen 9th–10th grade students three times per year in reading and math, (3) implement targeted screening for students who enter high school with a history of academic challenge, and (4) forgo screening and make decisions based on common assessment history.

Each high school’s protocol for screening is designed from a deep understanding of both the purpose for the RTI implementation and the ways that improvement is documented:

• In high schools where RTI is implemented for all students, universal screening with Star Assessments three times per year provides evidence to document student improvement in reading and mathematics.
• In high schools where the RTI implementation is focused on reading and math improvement in grades 9 and 10, universal screening three times per year supports this effort.

• In high schools focused strictly on students with a history of intervention, whether well above or significantly below proficiency benchmarks, targeted screening supports the implementation.
  - As with the two previous protocols, Star Assessments provide academic and growth metrics to monitor student improvement in reading and mathematics.
  - When working with students well above benchmark, particularly at the high school level, the SGP score provides insight into growth and adds context to achievement. Consistent academic gains for high-performing students become more difficult to document over time, and in some cases these students “may begin to lose ground as they progress through school” (Olszewski-Kubilius & Clarenbach, 2012). Growth metrics, however, provide evidence of growth over time and within school year.

• For high schools using the common-assessment-history protocol, successful implementations require additional measures to gauge proficiency on current benchmarks and monitor student growth. These metrics may be unavailable or difficult to gauge solely through a history of common assessments. Mining common assessment history provides some support for working with students in Tier 2, but may fall short in equipping educators with insight for differentiated instruction at Tier 1 (Duffy & Scala, 2012).

**Star™ screening data and specific learning classifications**

Star Assessments are general outcome measures that can be used directly, or in concert with additional measures, in discussions regarding learning differences. Star data directly assess and progress monitor three areas of specific learning disabilities: reading comprehension, math calculation, and math problem solving.5

- Star Reading—reading comprehension; valid estimate of oral reading fluency for grades 1–4
- Star Math—math calculation and/or math problem solving
- Star Early Literacy—early reading behavior

For more information, including research and detailed guidance, see *Using Star Reading Data for Progress Monitoring and Specific Learning Disability Identification* (Renaissance, 2015b).

**Using technology advancements to interact with vivid, visual data**

Among the first steps in lessening the distance between assessment and instruction is expanding use of universal screening data from informing placement decisions to informing daily instruction. While the idea of using assessment data to inform instruction has always been at the core of RTI, access to vivid, visual screening information provides additional ways to interact with these data beyond spreadsheets and graphs.

**Viewing data as color, shape, and size**

Technology as a means to administer assessments and store data is well established in RTI. In a next-generation RTI implementation, technology also allows educators to develop insight by interacting with data where color, shape, position, and size reflect current and historical data, predictive analytics,

---

5 According to the federal Individuals with Disabilities Education Improvement Act (IDEA) of 2004 (Section 300.8(c)(10), the eight areas of specific learning disabilities are: oral expression, listening comprehension, written expression, basic reading skill, reading fluency skills, reading comprehension, math calculation, and math problem solving/reasoning.
achievement and growth metrics, comparative data, and proficiency toward established benchmarks. For example, with the Star Reading data available in the Reading Dashboard (see figure 4), educators have access to multiple data points, including:

• Benchmarks—indicated by the variations of gray in the background
  - In this example, benchmarks that correspond to the state summative exam are represented.
• Achievement—indicated by blue circles
  - The large blue circle represents this student’s first Star scaled score of the school year.
  - The smaller blue circles represent each additional Star test within this school year.
• Growth—represented by a diamond
  - In this example, the SGP diamond is yellow, which means this student’s growth is typical.
• Range of potential growth—represented by the growth fan
  - Yellow is considered typical growth, green is higher than typical growth, and red is lower than typical growth.
• Growth targets—indicated by purple, blue, and orange labels
  - Target represents the Star scaled score the student needs by the end of the current school year to be on track for proficiency (in this case, on the state summative exam) within three years.
  - Projected and Predicted reflect the likely end-of-school-year Star scaled score for a student growing at a typical rate (i.e., the 50th SGP), as determined by the first score of the school year, the most recent Star scaled score, and one prior score from a previous screening window.
• Cohort comparisons—represented by the row of colored boxes (SGP differences, top of visual)
  - The + or – signs indicate the student’s SGP is above or below the median SGP for the cohort.

As evidenced by the success of RTI, it is clear educators are skilled in working with data and drawing inferences for guiding students. Technology enhancements do not change the data; however, they do fundamentally change access to the data and the ability to interact with data electronically. With increased access, educators now use continuous assessment (i.e., data from screening, day-to-day assignments, and observations about each student) to support differentiated instruction, which is crucial to RTI.

**Continuous assessment and data-fueled differentiated instruction**

The Algozzine-Ysseldyke Model of Instruction (see Algozzine, Ysseldyke, & Elliott, 1997) describes the relationship between continuous assessment and data-fueled differentiated instruction. This model identifies four evidenced-based activities involved in differentiated instruction:

• Planning instruction
  - Deciding what to teach, how to teach, and communicating realistic expectations (goals)
• Managing instruction
  - Preparing the classroom for instruction, using time productively, and establishing a positive learning environment
• Delivering instruction
  - Developing strategies for presenting information, monitoring instructional presentations, and adjusting presentations (adapting instruction)
• Evaluating instruction
  - Monitoring student understanding, paying close attention to engaged time, keeping records of student progress, and using data to make decisions

Star Assessment data guide planning, delivering, and evaluating instruction within the Algozzine-Ysseldyke Instructional Model in the following ways:

• Plan: The statistical link between Star scaled scores and the learning progressions identifies skills students are ready to learn. To guide planning, these skills are presented in teachable order.
• Deliver: Star’s statistical link to the learning progressions provides teachers with access to instructional and student resources for identified skills.
• Evaluate: Teachers work with Star dashboard data to monitor progress (1) on achievement and growth, (2) toward meeting goals, and (2) toward mastering objectives (with Renaissance reading and math practice solutions).

All RTI implementations, whether past or present, are focused on solid, differentiated instruction in Tier 1. Star screening data support differentiation in Tier 1, which is effective in bridging learning gaps—thereby reducing numbers of students in need of small-group intervention in Tier 2 (Fuchs, Fuchs, & Vaughn, 2014; Jones, Yssle, & Grant, 2012; Lipson & Wixson, 2012).

Response to information

In every RTI implementation, educators must quantify evidence of students’ responsiveness to selected interventions and take action based on that evidence. In RTI, assessment, instruction, and intervention are three intricately interrelated actions. In this next generation of RTI, this proven model holds true as a way to provide comprehensive support to students, linking assessment and instruction in ways that inform educators’ decisions about how best to teach students.

Educators’ response to information includes responding to what we are learning about non-cognitive influences on achievement. Continuous access to quantitative assessment, instructional data, intervention data, and data on non-cognitive influences provides a platform for educators to respond to intricately interrelated pieces of information. For the first time, educators are able to access information about achievement, growth, and mindset.

For example, Carol Dweck (2015, 2016), who pioneered thinking about growth mindsets, directs educators to deepen expertise in recognizing certain triggers that may alter a student’s thinking from a growth mindset to a fixed one. She suggests teachers can learn to recognize these triggers and take action before a mindset interferes with academic growth. As mentioned, the Stanford University study on mindset used fMRI to document the influence of a positive mindset on mathematical problem solving (Sparks, 2015), suggesting there may be cognitive functions supporting the impact of mindset on achievement. In an earlier study, researchers also used fMRI to monitor students’ responses to feedback (van Duijvenvoorde, Zanolie, Rombouts, Rajmakers, & Crone, 2008) and found that a student’s age may impact how he or she responds to instructional feedback.

With brain imagery technology, we can deepen our understanding of how certain non-cognitive influences affect achievement in daily interactions between teachers and students.

Progress monitoring in the next generation of RTI

Progress monitoring requires a quantitative assessment to measure student improvement, and educators
are now finding that flexibility in assessing for the purposes of progress monitoring is critical (Shapiro, 2008). Even as educators look to multiple metrics and qualitative indicators of student improvement, the foundational thinking about interventions and monitoring progress remains intact. When reliable and valid data—considered in context of the educator’s knowledge about the student—indicate a need for support beyond differentiated, core instruction (Tier 1), educators determine Tier 2 interventions and establish a means to monitor progress.

Continuous assessment data from valid and reliable progress monitoring tools provide insight related to the power of the intervention. Educators often do not have time to rigorously evaluate assessments to determine if they are reliable and valid for progress monitoring. Fortunately, independent reviews of progress monitoring measures, such as Star Assessments, have been conducted by the National Center on Intensive Intervention (2016). The NCII reports there is convincing evidence that these assessments are sensitive to student improvement; thereby effective in progress monitoring. To access this evidence, go to the NCII tools chart at http://www.intensiveintervention.org/chart/progress-monitoring.

Discussions about how best to monitor student progress often focus on three key areas:

- **Time**—how long students should be progress monitored
- **Intensity**—how often they should be progress monitored during that time
- **Metrics**—what kinds of data yield the most impactful information

All three areas, however, are subjects of intense debate.

**Time**

The rule of thumb for how long students should be progress monitored has been six to eight weeks, but there is little empirical evidence to support that rule. Several studies (see Shapiro, 2013) cast doubt on this guideline and suggest it should be much longer. These studies are almost exclusively focused on curriculum-based measurement of reading (CBM-R) measures.

There is little doubt that CBM-R assessments can be useful in helping teachers measure whether students are mastering the intricacies of reading. Empirical evidence supporting the use of CBM-R measures has been noted in three areas: (1) to make normative comparisons to peers on level of performance, (2) to document annual growth relative to peers (if the same passages are used within and between students), and (3) to compare relative differences across groups. The empirical evidence for using CBM-R data for decision making, however, is lacking when these data are used as the primary outcome measure to evaluate individual student growth over short periods of time (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013; Shapiro, 2013).

Research suggests CATs are a sound choice for progress monitoring student performance in RTI settings. As noted by Bray & Kehle (2011), “RTI is a process of providing high quality interventions that are matched to student need, and uses frequent progress monitoring of student response to interventions to assist in making important educational decisions” (p. 616). In 2012, Shapiro wrote that, “Star measures offer an important and potentially valuable contribution to RTI” (p. 20), and because progress-monitoring feedback tells educators which interventions are helping students most, these “measures must be frequent, sensitive to instructional change over a short period of time, predictive of overall success as measured by the benchmark assessment, and able to drive instructional decisions” (p. 9). Decades of research have shown that CATs can be considerably more efficient than conventional tests that present all students with the same test questions (e.g., Lord, 1980; McBride & Martin, 1983).

In some RTI implementations, students in Tier 2 work within an intervention program 30 minutes per day, five days per week. In other implementations, the intervention program is based on three days per week for about 30–45 minutes per day (Shapiro, 2008). Although existing literature falls short of specific recommendations for the length of time of a Tier 2 intervention, consistency is noted in two areas:
• Tier 2 interventions are most often delivered 30 minutes per session with three to five sessions per week (Gersten et al., 2009; Harlacher, Walker, & Sanford, 2010).
• Tier 2 intervention is in addition to—and should not displace—Tier 1 core instructional time (Balu et al., 2015).

Intensity
How often students in a Tier 2 intervention should be progress monitored varies in the literature from monthly, bimonthly, or every two weeks (Brown-Chidsey & Steege, 2010; Kaminski, Cumming, Powell-Smith, & Good, 2008). Other researchers suggest as often as weekly (Johnson, Mellard, Fuchs, & McKnight, 2006).

In next-generation RTI implementations, there is a distinct place for monitoring students’ developing progress. Traditionally, teacher-made quizzes and/or skill probes provide this type of evidence. With advancements in technology, these same time-honored metrics take on greater power. Renaissance Star Custom®, for example, is a cloud-based assessment platform that provides a way for teachers, and students, to probe and monitor developing skill mastery. For a deeper discussion on monitoring developing progress, please see A few thoughts on progress monitoring by Dr. Gene Kerns, Chief Academic Officer, Renaissance (box at right).

Metrics
Educators typically select data to use in progress-monitoring decisions within the context of the goal established for the intervention, such as:
• Competency—reach a predetermined score on a Star Assessment
• Proficiency—reach an established proficiency benchmark on a summative exam (or a level on Star that is statistically equivalent to the summative benchmark)
• Normative data—meet or exceed a designated benchmark, such as the 40th percentile
• Growth—exhibit typical or greater-than-typical growth based on achievement history among grade-level academic peers

In this next generation of RTI, educators look to sophisticated growth models like SGP, which is measurable within Star Assessments. With access to achievement and growth data in Star, educators have at least two options for setting growth as a progress-monitoring goal. For example, an educator may set a goal for students in Tier 2 interventions to grow at a faster pace than same-grade, academic peers. As another option, the educator may use a growth measure like SGP to understand whether meeting a competency-based or normatively-based goal is a realistic expectation. Using a growth metric like SGP is a way to “better interpret change over time” (Kovaleski, VanDerHeyden, & Shapiro, 2013, p. 77; see also Johnson & Thompson, 2016).

A few thoughts on progress monitoring
Dr. Gene Kerns
Chief Academic Officer, Renaissance

Progress monitoring is one of the most psychometrically advanced tasks educators undertake when using the RTI model. If tools are not chosen wisely and implemented with consideration, this process can displace valuable instructional time.

Progress monitoring is a formal protocol necessary to collect valid and reliable data to chart students’ performance against expected outcomes. This process informs decision making around whether to continue, end, or modify a given intervention.

Some educators struggle to delineate the progress-monitoring protocol from regular assessment practices embedded in instruction. One way to distinguish these practices is to use the term “progress monitoring” when referring to the formal protocol and to use “formative assessment” for everything embedded into instruction.

It is important to recognize that the progress-monitoring protocol has many more specific requirements than any formative tool that might be used because of the significance of the decisions being made.
Goal setting, growth, and progress monitoring

As mentioned, Shapiro (2012) provides detailed information on using Star Assessment achievement data for progress monitoring in the first generation in RTI, and in this next generation, his guidance remains impactful. For the purposes of this paper, we shift now to focus on growth metrics as appropriate measures for progress monitoring.

In virtually any arena—education, athletics, music, business, and so forth—individuals determined to improve their performance (or the performance of their students/team) often begin with setting a goal. Setting achievement goals, monitoring progress toward goals, and making adjustments to instruction designed to meet the goal, or to the goal itself, have long been hallmarks of RTI implementations (Fox, Carta, Strain, Dunlap, & Hemmeter, 2009).

For an excellent summary of best practices in goal setting see Shapiro and Guard (2014), who note that the first important practice is to identify a challenging, yet realistic, achievement goal for each student. Progress monitoring is then done in the context of the student's progress toward that goal. However, there has been little research guidance on how to identify goals, leaving it to be a somewhat arbitrary process. The following summarizes the latest thinking on goal setting, particularly the role of student growth metrics in first informing goals and then evaluating progress.

Shapiro and Guard (2014) describe two frameworks for thinking about progress-monitoring goals. The first framework supports a competency-based approach and includes goals such as, “I want my student to reach the Proficient level on the state summative test in April,” or “I want my student to read at Level ___ by January.” The second framework—a normative approach—compares a student to his grade-level peers. This type of goal requires that assessments be formally normed or have very large databases that aggregate data for educators to make informed choices. An example of a normative goal might be, "I want my student to reach the 40th percentile in math by the end of the school year." The 40th percentile can be thought of as a very rough guideline for "working at grade level," according to RTI experts (see Screening in a next-generation implementation, p. 13, and Shapiro in Renaissance, 2013).

The concept of student growth, or rate of improvement, is not new in RTI. Regardless of the goal chosen by the educator, progress must be tracked from a starting point, such as a screening/benchmarking test, and observations made about how much the student is growing while working toward the goal. If the rate of improvement exceeds expectations, the goal can be increased, or this information may support a teacher’s decision to allow the student to exit the intervention. If improvement is lower than is required to hit the goal, instruction can be changed or intensified. Again, these are not new ideas; what is relatively new is the role that growth measures can play in this process.

In recent years, educators have had access to robust growth models built on massive databases of student performance in reading and math, and these models have proven helpful to educators in both informing goal selection and monitoring progress. The most prevalent of these models is SGP (Betebenner, 2011), which is reported by many state summative assessments and by Star Early Literacy, Star Reading, and Star Math:

"SGPs are a norm-referenced quantification of individual student growth derived using quantile regression techniques. An SGP is a moment-in-time score that compares a student's growth from one period to the next to that of his or her academic peers—defined as students in the same grade with a similar score history. SGPs range from 1–99 and interpretation is similar to percentile rank (PR) scores: lower numbers indicate lower relative growth and higher numbers show higher relative growth. For example, an SGP of 75 means that the student's growth exceeds the growth of 75 percent of students with a similar score history. Typical growth is an SGP of 50, though state and local policy makers may define a range for typical growth." (Renaissance, 2016, p. 2)

---

6 Initial efforts to incorporate growth in RTI involved relatively small databases of student data from rather limited CBM-Reading measures.
Educators now use SGP for two purposes related to goal setting and progress monitoring:

- **SGP as the goal.** This can be thought of as an extension of the normative approach to goal setting. Because SGPs measure growth against national academic peers with similar achievement histories, some educators aim for a student to grow more than what is expected (typical growth is SGP 50). They may identify SGP of 65 or 75, for example, as an ambitious goal.

- **SGP as a reality check.** Whether an educator uses a competency-based approach or a normative approach to goal setting, SGP can be useful as a general indication of the likelihood of success. For instance, if an SGP of 90 is associated with reaching a particular goal, such as attaining proficient status on the end-of-year state test, the educator can readily tell that just 10 percent of similar students nationally would be expected to attain that goal. The educator is free to select the student’s goals, but knows whether a goal is very ambitious given the student’s recent history. Conversely, if the goal selected is associated with an SGP of 19, the educator will understand that about 81 percent of the student’s national academic peers have met or exceeded that amount of growth and, thus, the goal is not very ambitious.

Renaissance believes teachers know their students and are in the best position to set challenging, but realistic, goals. We do not recommend specific goals but instead provide contextual information and tools to help teachers help their students.

Other relatively new advances and knowledge in the area of goal setting include the following:

- Many formative and daily practice tools include personalized goal-setting features. For instance, Renaissance Accelerated Reader 360® allows teachers to set personalized reading-practice goals regarding how much students read and how well they understand what they read. A recent analysis of Accelerated Reader goal-setting data supports prior findings on the value of goal setting and progress monitoring: students who met set goals read more, enjoyed higher comprehension, and achieved higher levels of academic growth than similar students who did not have reading-practice goals (Renaissance Learning, 2015a).

- Although goals are most often set for students who are struggling or at-risk, it is worthwhile to set goals for all students, even those who are very high achieving (Johnsen & Coleman, 2012). A large-scale review of goals by Haas, Stickney, and Ysseldyke (2016) found that 30 percent of all goals were set for students with fall/pretest achievement scores above the 50th percentile, challenging assumptions that progress monitoring goal setting is limited. Haas et al. looked at reading, math, and early literacy goal and achievement data involving more than 360,000 students and found:
  - Most teachers (72%) chose moderate or typical rates of growth for students (roughly equating to SGP 50)
  - Of students with ambitious goals (roughly SGP 75) set by their teacher, a surprisingly high percentage met their goal, a finding that is consistent with other studies (e.g., Fuchs, Fuchs, & Deno, 1985) that ambitious goals may be associated with growth.
  - In reading and math, students who met set achievement goals significantly outperformed students without set goals. This is consistent with the literature on goal setting (e.g., Hattie, 2009) that indicates setting goals is associated with accelerated achievement.

In this next generation of RTI, deeper thinking about goal setting considers reflective, contextual goals that support both students who learn with ease and those who learn with determination. Non-cognitive goals, such as adapting a different attitude about learning, are taken into consideration with achievement thresholds and proficiency benchmarks. We now know of the physical and visible connection between mindset and mastery. Growth goals provide insight into success that has previously been difficult to document. Likewise, these goals offer challenge to learners with a history of high achievement. Like

---

7 One feature of the Star Goal Setting Wizard is a “catch-up, keep-up” indicator, which is tied to the state summative test via the Star scores estimated as equivalent to the state's "proficient" category (may also be called "passing," "met standards," and so forth). This indicator reports the minimum amount of growth in scaled score points per week and the SGP a student would need to reach (if they are below) or maintain proficiency over the next three years.
screening, differentiation in Tier 1 instruction, and progress monitoring, setting growth goals offers next-generation ways to support each learner.

Conclusion

Response to Intervention has endured and evolved over the past decade to support next-generation thinking about standards, assessing student achievement, and implementation integrity. Conversations about using technology to amass assessment data now focus on capturing the power of technology to lessen the distance between assessment and instruction. Continuous assessment and access to learning analytics fuel decisions about core instruction, differentiation, and intervention, not only lessening the distance between assessment and instruction, but making them almost indistinguishable.

RTI implementations have evolved to support all students—from the primary grades through high school. In many districts, RTI is more widely implemented across the content areas. In other districts, the RTI framework supports growth in non-cognitive influences on learning.

Because brain-imagery research allows us to see to a greater extent how we learn, there is more potential for accelerating learning. This insight has inspired us to focus as much on how we learn as the most impactful way to lead students to learning. As RTI continues to progress into the next generation, educators will monitor ways their students respond to specific interventions and make thoughtful, impactful data-based decisions augmented by student observation.

RTI has evolved to focus even more on mindsets and attitudes about learning. Educators who understand mindset triggers, and strategies to quiet these triggers (Dweck, 2016), have the power to engage students in a skillfully designed, momentary intervention so seamless, that few—beyond the teacher and learner—may notice.

Certainly, skillful implementation of RTI programs will continue to evolve as we continue to deepen our understandings about leading students to achieve. We owe much to those who guided America’s schools to embrace data-fueled insight for decision making. From the late 1970s through today, advancements in learning science, data analytics, and brain imagery have provided continuous insights about how we learn. As we continue to hone our data-fueled decision making, we carry forward concepts about growth, mindset, insight, and continuous assessment. At the forefront, we continue to move from using assessment data to make predictions about students’ academic lives to make a difference in their everyday lives.
References


Shapiro, E. S., & Guard, K. B. (2014). Best practices in setting progress monitoring goals for academic skill improvement. In P. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Student level services*. Bethesda, MD: National Association of School Psychologists.


