Thank you for downloading

Using Response to Intervention Data to Advance Learning Outcomes

Amanda M. VanDerHeyden

from the Center on Innovations in Learning website
www.centeril.org

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

VanDerHeyden, A. M. (2013). Using response to intervention data to advance learning outcomes. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on innovations in learning* (pp. 20χ–22φ). Philadelphia, PA: Center on Innovations in Learning, Temple University; Charlotte, NC: Information Age Publishing. Retrieved from http://www.centeril.org/

Center on Innovations in Learning

Using Response to Intervention Data to Advance Learning Outcomes

Amanda M. VanDerHeyden

Response to intervention (RtI) is a system of service delivery that uses student data to evaluate and repair core instruction and to provide increasingly intensive intervention supplements to students who need it to meet expected learning outcomes. Universal screening is conducted to identify students who are likely to experience academic failure and to indicate the general adequacy of instruction for the system. Screening data are used to indicate a need for core instruction enhancements that affect all students and to evaluate the extent to which such enhancements improve the effects of instruction for all students. Universal screening data are also used to identify students who require supplemental instruction to attain important learning objectives. RtI (now often called multi-tiered systems of support) is generally presented as a filtered system whereby student data are collected to identify risk. Based on that data, increasingly intensive interventions are provided to subgroups of students, with (a) most students successfully responding to core instruction alone, (b) a small subset of students requiring supplemental intervention support to experience success, and (c) a smaller subset requiring intensive individualized intervention to attain important learning outcomes.

When used effectively, RtI systems generate data that indicate the general effectiveness of instruction in a system; that is, the percentage of students requiring intervention should be below 20% and should decrease over time with core instructional enhancements (O'Connor, Fulmer, Harty, & Bell, 2005; Shapiro & Clemens, 2009). RtI systems also generate data that may be used to identify children for special education eligibility, particularly under the category of specific learning disability (Kovaleski, VanDerHeyden, & Shapiro, in press).

RtI came to the forefront in the 1990s as an innovative means of using data to determine when and for which students instruction was working and what types of adjustments were needed for students who were not learning successfully. The RtI framework has its roots in precision teaching, direct instruction, curriculum-based measurement, and school-based consultation. It has been widely studied with encouraging results (Burns, Appleton, & Stehouwer, 2005), endorsed and recommended by numerous policy groups as a method of system reform (Batsche et al., 2005; Bradley, Danielson, & Hallahan, 2002; Donovan & Cross, 2002), and permitted as a method of eligibility determination under Individuals with Disabilities Education Act and state regulations (IDEA, 2004). While promising, however, the effects attained depend upon the quality with which components are implemented, and quality of implementation varies greatly across sites. This chapter will demonstrate how to use student performance data to make decisions about core instruction adequacy, to guide instructional enhancements to core instruction, to identify small groups and individual students for intervention, to guide small-group and individual intervention, and to evaluate the effects of instructional changes so that implementation can be managed effectively and desired student learning improvements can occur.

Key Action 1: Conduct Screening to Yield High-Quality Data

Universal screening is the starting point for any RtI implementation. Brief academic assessments are administered schoolwide, typically in reading and in mathematics, to characterize student performance by school, grade, and class. Universal screening data are a rich resource that is often underexploited by decision makers. Investing the time needed to ensure that the screening is conducted with sufficient quality to yield meaningful data is time well spent because screening data can be used to accomplish several objectives (which will be explained below). Adequate universal screening measures should (a) yield reliable scores, (b) forecast future learning success or outcomes, (c) be administered efficiently, and (d) reflect the mastery of key academic objectives (Kovaleski et al., in press). Screening measures can be selected to reflect a performance standard that children are already expected to have mastered because mastery of that skill or skills is an essential prerequisite to the instruction that students will experience. The screening must be administered correctly and scored accurately. For efficiency, curriculum-based measurement probes in reading and mathematics function well as screening devices, and it is possible for schoolwide screening to occur within a single day, requiring no more than 45 minutes in any class. Universal screening typically occurs three times per year. Screening scores should be organized by content area (e.g., reading, mathematics), school, grade, and class. Because screening yields data upon which important decisions will be based (e.g., who receives intervention), it is important to verify that high-quality screening has occurred. When training professionals to collect screening data, the following indicators may be used as a guide to determine that a professional has been adequately prepared to conduct high-quality screening.

Table 1. Key Action 1: Indicators That Trainee Is Proficient in Screening

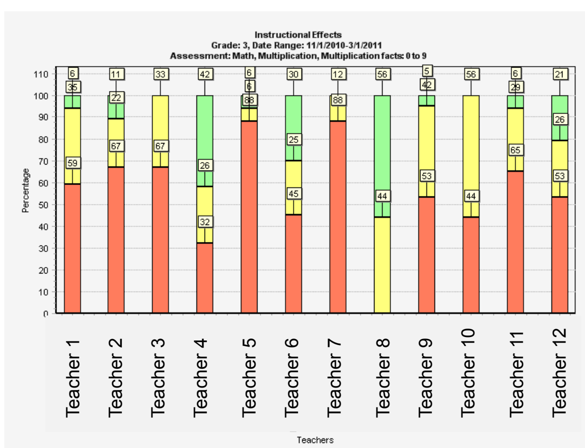
Faculty overview has been provided and screening materials selected.
Screening has been scheduled to occur on a single day, and screening schedule has been planned.
All materials for screening are available and have been organized by class, including a written protocol for screening.
Trainee has been observed to correctly administer and score screening materials.

Key Action 2: Interpret Screening Data Beginning With an Aerial View

Screening data can be examined to identify schoolwide, gradewide, and classwide problems. Decision makers should begin at the district or school level—the aerial view—and work their way down through the data to the grade, class, and, finally, individual students. A schoolwide learning problem is detected when more than half the grades within a school exhibit a gradewide problem. A gradewide problem is detected when more than half the classes in a grade exhibit a classwide problem. A classwide problem is defined as the median score for the class falling within the risk range associated with the screening tool.

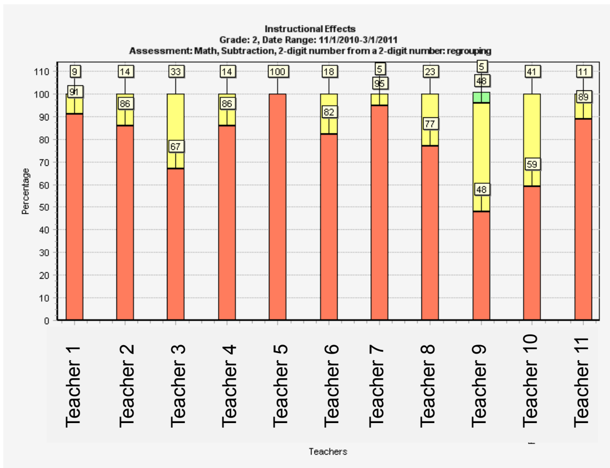
In the example shown in Figure 1, a gradewide problem in mathematics was detected by a screening that was conducted in February of the third-grade year after multiplication facts 0–9 had been taught and students were expected to demonstrate proficient performance of that skill. Figure 1 shows that, in 8 of 12 classes, the majority of students performed in the risk range and therefore constituted a classwide problem. Because more than half of the classes at this grade level scored in the risk range during screening, a gradewide problem is indicated. There is no need to look further at individual classes or individual students because the gradewide problem should be addressed first.

Figure 1. Instructional Effects, Grade 3. Assessment: Math, Multiplication, Multiplication Facts 0–9



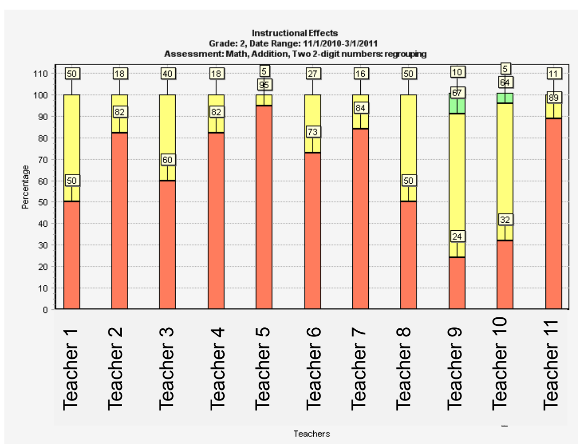
The data team should next examine other grade levels to determine if they show a similar gradewide learning problem in mathematics. Figure 2 shows the universal screening data for second-grade mathematics in the same school.

Figure 2. Instructional Effects, Grade 2. Assessment: Math, Subtraction, 2-Digit Number from a 2-Digit Number, Regrouping



Second grade was also administered a 2-digit addition probe with regrouping, with the results shown in Figure 3.

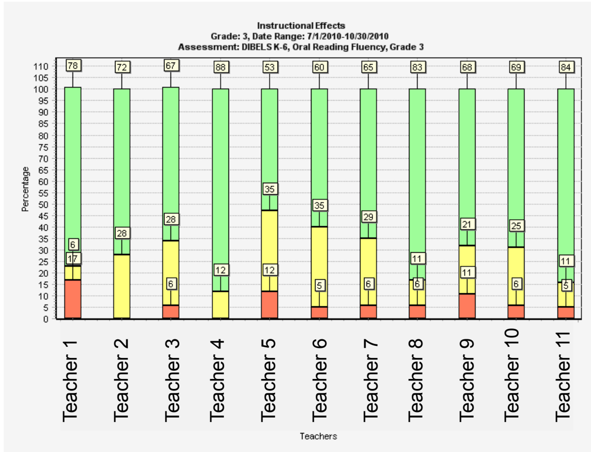
Figure 3. Instructional Effects, Grade 2. Assessment: Math, Addition, Two 2-Digit Numbers Regrouping



Thus, a schoolwide learning problem in mathematics was identified for this elementary school serving Grades 1 through 3.

Let's consider reading performance for the same grade level as indicated in Figure 4.

Figure 4. Instructional Effects, Grade 3. Assessment: DIBELS K–6, Oral Reading Fluency, Grade 3



Here we reach a different conclusion. In this example, screening reveals no classwide problem—and therefore no gradewide problem—in reading. For the school, then, only the schoolwide problem in mathematics needs to be addressed through systemic solutions. In reading, individual children can be selected for further assessment and possibly intervention. Data teams will want to verify that the screening task was appropriately selected at each grade level (the difficulty of the screening task was well aligned with standard learning expectations at each grade level). Systemic performance problems should be treated with systemic solutions, which will be briefly discussed in the next section. When training professionals to interpret screening data, the indicators presented in Table 2 may be used to examine mastery of screening data interpretation.

Table 2. Key Action 2: Indicators That Trainee Is Proficient in Data Interpretation

Trainee has ruled out school-level, grade-level, and whole-class performance problems prior to selecting individual children for follow-up assessment and possibly intervention.
Data have been organized by grade and by class.
Data have been examined for identified vulnerable or high-risk groups of students to identify potential performance patterns (e.g., high numbers of new students scoring in the risk range, disproportionately high numbers of special education students scoring in the risk range).

Key Action 3: Treat Systemic Problems With Systemic Solutions

Systemic problems deserve systemic solutions. So when a schoolwide learning problem is detected, the first step to be taken by the data team should be to verify that research-supported curriculum materials are available to all teachers. The data team should also verify that the teachers understand what learning outcomes are expected of students and have a clear calendar of instruction that specifies the time points by which certain learning outcomes will have been attained. Next, the data team should examine the quality of instruction in the classroom; team observations should answer questions such as the following:

- Is adequate instructional time allocated?
- Are students actively engaged during the instructional period?
- Does the teacher have a system for knowing which skills students have mastered and which skills require additional support to reach mastery?
- Does the teacher align instructional efforts with student needs (e.g., acquisition supports for skills that have not been established, fluency-building supports when student responses are accurate but slow, systematic practice applying learned skills to solve more complex problems or in different contexts)?

The data team should establish priorities for improvement and determine a timeline. So, if a schoolwide problem were detected at all grade levels, the data team may choose to begin a classwide intervention for all classes at one grade level, while simply monitoring performance weekly in the other grades and providing feedback to teachers. If systemic performance problems were detected in reading and mathematics, the data team may choose to target one content area initially and add the second only after improvements are attained for the first. Both of these approaches allow for a staggered or incremental solution implementation, which allows the data team to implement the intervention with quality, ensure that performance gains occur, troubleshoot any implementation challenges, and expand to new areas as capacity for implementation is increased.

In each class with a classwide problem (i.e., median score in the risk range), a classwide supplemental intervention should be conducted. Building fluency in prerequisite skills (i.e., skills that have been taught but which students have not mastered and which are required for successful goal-level performance) is an ideal target for a classwide intervention. Classwide intervention can occur daily within about 20 minutes and can produce large returns on proximal (targeted skills) and distal (more comprehensive or multicomponent skills, including content and skills not directly taught during the intervention) measures (Codding, Chan-Iannetta, Palmer, & Lukito, 2009; Fuchs, Fuchs, Mathes, & Simmons, 1997; VanDerHeyden, McLaughlin, Algina, & Snyder, 2012). When a classwide intervention has been initiated, progress monitoring should occur weekly. Weekly progress monitoring is used to determine when to advance task difficulty of the intervention and to signal the need for in-class coaching to support the fidelity of intervention implementation.

Data teams should examine and respond to implementation effects each month. The data demonstrating a systemic problem and the intervention data reflecting improvements gained through intervention should be shared with decision makers in the system's feeder pattern. Instructional leaders should consider and identify ways to prevent the same problem in the future and provide supports to ensure maintenance of intervention gains over time and across grade levels. One common need identified during multigrade and multischool troubleshooting sessions is an increased rigor of learning expectations and practice opportunities at earlier grade levels. Improved rigor will "reduce the load" experienced at subsequent grade levels and help prevent the emergence of gradewide performance problems. Progress monitoring data should reflect that at-risk performance by demographic categories becomes proportionate over time with intervention improvements. The percentage of students not at risk should increase following intervention. Systems can define and track their own long-term outcomes, such as the percentage of students enrolling in and passing algebra, advanced placement course enrollments and advanced placement test scores, and the percentage of students taking and meeting the ACT benchmarks for college readiness.

When an isolated classwide learning problem is detected (the majority of classes at a grade level are doing fine, but a minority of classes—one or more—have more than half of their students in the risk range at screening), classwide intervention can be started immediately. While training the teacher (or teachers) to implement a classwide intervention, the coach can assist the teacher to improve core instructional procedures (e.g., Does the teacher follow the master schedule? Does the lesson plan include time for establishing new skills, verifying understanding of new skills and information, providing guided practice with corrective feedback for new skills, providing fluency-building support for established skills, monitoring student performance for mastery, and providing structured support to generalize skills and connect newly learned information to existing knowledge?). The classwide intervention can be used to establish mastery-level performance of prerequisite skills and serve as a training vehicle to provide the teacher with an expanded skill set to enhance the quality of core instruction. The classwide intervention can be delivered following a standard, scripted intervention protocol (e.g., Vanderbilt Kennedy Center, n.d.; http://www.gosbr.net/).

Children who successfully respond to intervention should surpass the screening risk criterion at higher rates on subsequent screenings. Students receiving intervention should also pass the year-end accountability tests at higher rates following intervention. Unsuccessful responders should qualify for more intensive instruction at higher rates. Students successfully responding to intervention and students not successfully responding to intervention should be proportionate by demographics.

Only after systemic problems have been ruled out should individual children be considered for intervention support. The utility of a decision rule to determine academic risk status is affected by the prevalence of risk in the group within which the decision rule will be applied. Providing classwide intervention in classes where a classwide problem has been identified is more efficient than working with individual children, more effective in terms of showing learning gains for all students, and results in more accurate decision making in subsequent risk decisions because it reduces prevalence of risk within the group. Indicators that a professional has been adequately trained and equipped to deploy systemwide interventions are provided in Table 3.

Table 3. Key Action 3: Indicators That Trainee Is Proficient in Treating Systemic Solutions

Classwide interventions have been started in classes with classwide problems.
Data teams have examined core instructional procedures in classes and grades with systemic problems.
Vertical teaming has occurred across grades and schools within the feeder pattern to share screening data and systemic intervention data.
Lower percentages of students fall in the risk range across consecutive screenings within and across years.
Higher percentages of students meet the proficiency criterion on the year-end accountability measure over time.
Historically vulnerable students show learning gains and fall into the risk range at lower rates. Performance gaps between those at risk and not at risk are reduced with intervention.
All students, including those in the higher performing groups and the lower performing groups, show gains with intervention and over time.
Students found to be at risk become proportionate by demographics with interventions.

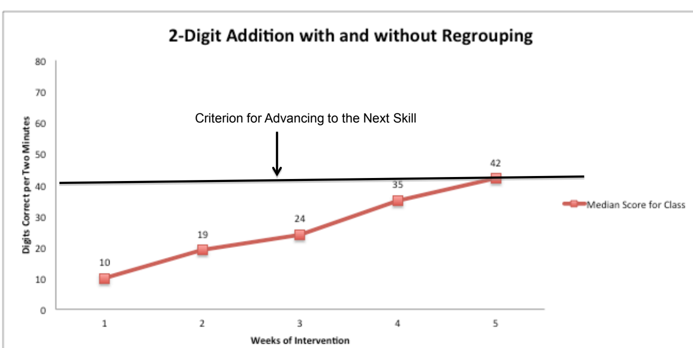
Key Action 4: Monitor Implemented Solution Effects and Manage Implementation Effectively

Once systemic problems have been detected and addressed through intervention, the data team can identify individual students for assessment and intervention. Individual children falling in the risk range should participate in brief follow-up assessments to verify risk-range performance, test the effect of rewards on performance, reduced task difficulty, and brief instructional trials on learning. This type of assessment is referred to as "functional assessment" or "brief experimental analysis" in the intervention literature (Daly, Witt, Martens, & Dool, 1997; Wagner, McComas, Bollman, & Holton, 2006) and explains the process of aligning instructional strategies with student skills for optimal intervention effects. The purpose of functional assessment is to identify an intervention that will work (i.e., has a functional relationship with student learning) when the intervention is properly used. If several children at a given grade level perform similarly (require instruction on the same content and subskill, require the same type of instruction), those students may be organized into a small group for supplemental instruction. Small-group intervention should occur daily with weekly progress monitoring and weekly adjustments as students' performances change. So called "standard protocol interventions" can be especially useful supplemental interventions (i.e., Tier 2) that generally involve teaching grade-level skills in a more explicit fashion with more opportunities to practice and receive corrective feedback. Children whose scores improve outside of the risk range on the lesson objectives and the screening criterion can be exited from the small-group intervention. Some children will require individualized intervention (i.e., Tier 3) to attain expected learning outcomes. These children should participate in an individual functional assessment to develop and test an intervention that will be conducted individually each day. During the individual assessment session, intervention targets should be specified, an effective intervention should be identified, and baseline performance should be quantified. Intervention progress should be examined weekly (five data points per week), and the intervention should be adjusted to accelerate growth when needed. Individual growth should be detectable within about 2 weeks. After ruling out poor fidelity when the intervention does not produce growth, data teams should troubleshoot and adjust the intervention (Fixsen & Blasé, 1993; Noell & Gansle, 2006).

In the implementation of an intervention, the lack of fidelity to its design is a persistent and ubiquitous threat. To prevent fidelity problems, coaches should provide adequate support for correct intervention implementation with ongoing monitoring of student outcomes. Where student outcomes lag, in-class coaching support, following a process known as "performance feedback" (Noell et al., 2005), should be provided.

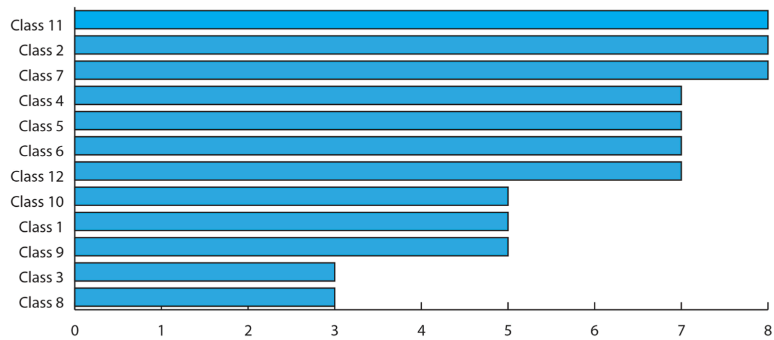
Each week, the data team should examine growth in each classroom to verify that gains are being made. Where gains are not occurring, a trainer or coach should visit the classroom during intervention to verify correct intervention use or provide support and coaching for stronger intervention implementation (i.e., provide performance feedback). Figure 5 below presents the data from a class that is working on a particular skill target; growth each week is monitored and reflects steady, upward gains and the class's meeting the goal within a few weeks of the start of intervention.

Figure 5. 2-Digit Addition With and Without Regrouping



Where classwide intervention is occurring in many classes, data teams should identify those that "lag" relative to classes in the same grade at the same school in terms of the number of trials needed to reach the criterion (or duration of time required to meet the goal). That is, lagging classes can be identified by tracking the number of targeted skills mastered by class, as demonstrated in Figure 6. Maintaining these data is an efficient way to identify lagging classes each week so that a trainer or coach can visit those classes and provide support for greater intervention gains. In Figure 6, Classes 3 and 8 are lagging behind the other classes—all given the same classwide intervention—in terms of skill gains. Classes 1, 9, and 10 also require in-class coaching and support to maximize intervention gains.

Figure 6. Number of Skills Mastered



Interventions that are not actively managed for fidelity and consistency will not be effective. One of the most important functions of the data team at a school is to actively manage intervention implementation, which includes monitoring intervention effects and providing support in classrooms where gains are not observed. In Table 4, indicators are provided for effective intervention implementation management that could be useful when training professionals to manage intervention or evaluating the quality with which interventions are being managed in a system.

Table 4. Key Action 4: Indicators That Trainee Is Proficient in Monitoring Intervention Effects and Managing Implementation

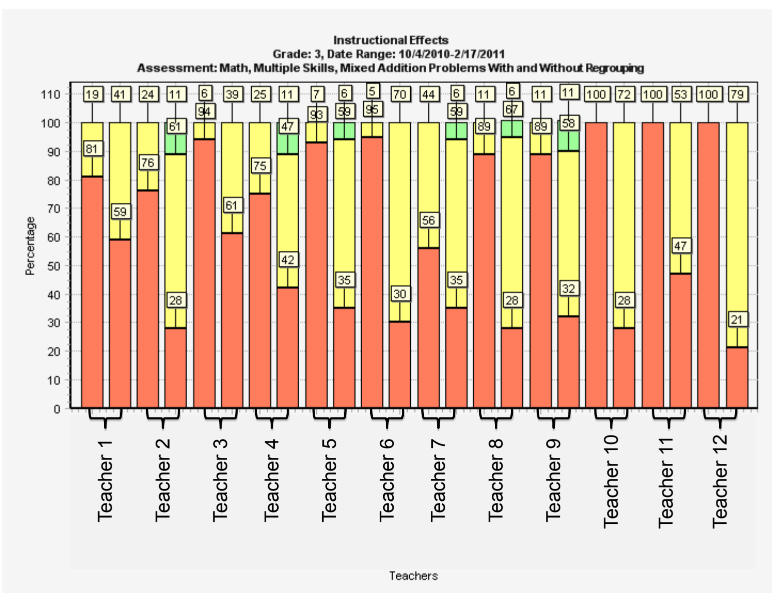
Interventions have written protocols available for teachers to use.
The teacher has been provided with all needed materials to conduct the intervention and has demonstrated correct and independent use of the intervention prior to being considered trained.
An in-class trainer or coach is available to model correct intervention use and provide in-vivo training for the teacher.
A tracking log is available showing at a glance who is experiencing intervention in the school.

A master schedule is followed to deliver classwide, small-group, and individual interventions.
Weekly progress monitoring data are collected for all children experiencing intervention.
Progress monitoring data are graphed, and interventions are adjusted weekly with in-class support where growth is not occurring as anticipated.

Key Action 5: Conduct Follow-up Screening to Verify Improvements

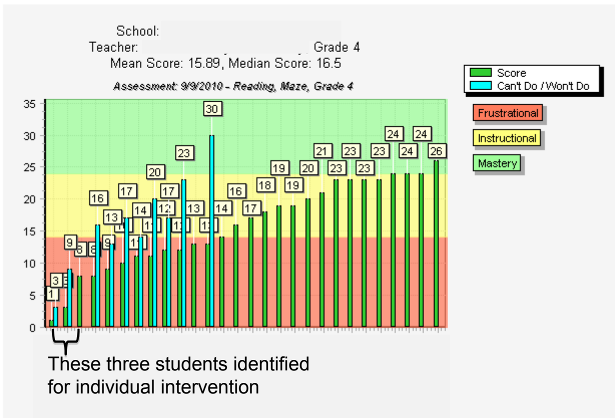
Intervention should produce appreciable effects for students in the school. Subsequent screenings should show that fewer children fall into the risk range, as shown in Figure 7. In Figure 7, the paired bars show the percentage of students at risk (orange) during the fall and winter screening, respectively, for each teacher. So, for example, for Teacher 1, 81% of children in her class score in the risk range during the fall screening, and 59% of children score in the risk range during the winter. Comparing fall and winter screenings reveals that for all 12 teachers the percentage of students at risk is decreasing with intervention.

Figure 7. Instructional Effects, Grade 3. Assessment: Math, Multiple Skills, Mixed Addition Problems With and Without Regrouping



Similar data could be accumulated for all grades to show schoolwide progress. In any case, individual students who experience intervention should perform above the risk range on subsequent screenings and score at higher rates in the proficient range on the year-end accountability measure. Figure 8, a classwide screening graph, shows the baseline reading performance of students in a fourth-grade class (i.e., green bars). A follow-up assessment tested the effect of incentives on performance, and the blue bars next to the green bars show the students' scores upon being given an opportunity to earn a reward for beating the previous score. Based on the screening and follow-up assessment with incentives, three children were identified for individual reading intervention in this class.

Figure 8. Assessment 9/9/2010—Reading, Maze, Grade 4



These three children participated in intervention that was actively managed. The subsequent classwide screening graph (for the winter screening) is shown in Figure 9. Here we see that two of the three children exposed to intervention now perform outside of the risk range for the class at the winter screening. Performance outside of the risk range during subsequent screenings is an indicator with great consequential and social validity and indicates that the interventions are having positive effects on important outcomes for the school. Key indicators of adequate use of follow-up screening data to verify intervention gains are provided in Table 5.

Figure 9. Assessment 1/21/2011—Reading, Maze, Grade 4

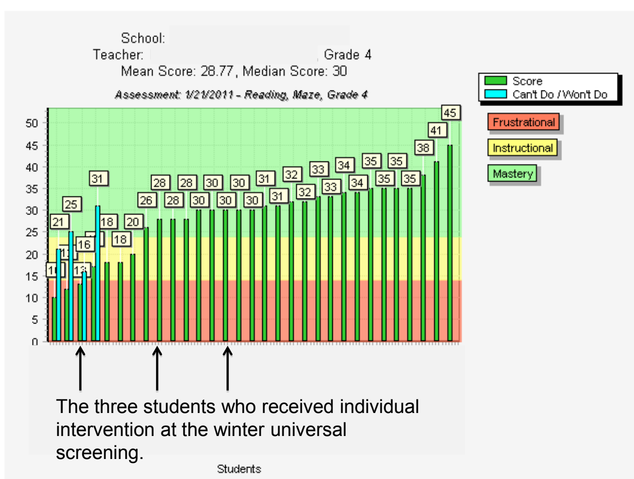


Table 5. Key Action 5: Indicators that Trainee is Proficient in Organizing Follow-up Data to Verify Improvements

The data team organizes data across consecutive screenings to show a reduced risk status accompanying the intervention.
Intervention effects are monitored for vulnerable or at-risk students over time.

When data are used to track instructional effects schoolwide and to make adjustments to instruction, learning outcomes can be accelerated. In the preceding case example, Figures 1–9, I have illustrated how universal screening data can be used to identify systemic problems, to monitor and manage intervention effects, and to evaluate intervention effects for the system. This case example ends with a caveat. Three of the most common errors in data-driven instructional decision making are (a) to collect too much of the wrong data, (b) to fail to expect intervention integrity errors, and (c) to fail to actively manage intervention to avoid fidelity errors. In this case example described above, the process for active management of intervention has been highlighted. The remaining space will used to explain how to avoid overassessment.

One of the most common implementation pitfalls in RtI is overassessment. Overassessment involves collecting data that provides redundant information or does not provide useful information. In many RtI systems, implementers conduct multiple screenings to determine which students are at risk. Overassessment is a costly waste of resources and comes with a direct cost to instructional time. Further, overassessment reduces the probability that the data will be used because implementers are overwhelmed by so much data and unsure how to translate the data into actions that make a difference.

To avoid the pitfall of overassessment, data teams should take an assessment inventory and verify that each assessment has a unique purpose. Further, data teams should verify that the intended purpose is served by each assessment in the least costly way possible. Where multiple assessments are being administered to inform the same decision, the data team should use local data to examine which measures provide the best utility for decision making. Data teams should examine local data to verify that publisher-suggested cutscores are serving the decisions well (accurate, sensitive, and efficient).

One metric for comparing the utility of each test measure is the AUC. The AUC stands for “area under the curve,” and it is the probability that the results of a given test will rank a student who fails the criterion lower than it would rank a student who passes the criterion. It also is equivalent to the average sensitivity over all false positive rates. AUCs range from .5 (no value) to 1.0 (perfect value), and some groups (e.g., rti4success) recommend an AUC of at least .80 to consider a test potentially useful. The AUC is derived from a receiver operating characteristics (ROC) curve analysis. As shown in Figure 10 and Table 6, ROC analysis considers the full range of available screening scores and, for all possible decision thresholds (i.e., the number of unique test scores minus 1) on the test, plots the sensitivity of that score against the false positive rate for that score if it were used as the cutoff value in predicting the criterion (in this case, proficient performance on the year-end accountability measure). Thus, data teams can scan the associated AUC values and identify those with the greatest relative predictive value. In Figure 10, ROC curves have been plotted for each of the possible screening measures so that data teams can visually identify the relative merit of each screening. Generally speaking, as in Figure 10, trend lines closer to the upper left quadrant of the graph (as high vertically as possible indicating very strong sensitivity and as close to the y-axis as possible indicating very few false positive errors) are stronger and will have stronger AUC values; screening measures analyzed in Figure 10 show little difference among their ROCs and equally high relative predictive value.

Figure 10. ROC Curve

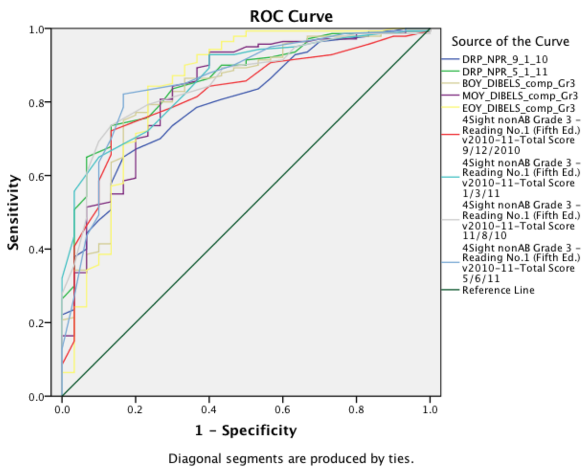


Table 6. Receiver Operating Characteristics (ROC) Curve Analysis

Screening	Correlation With State Annual Proficiency Assessment in Reading	Percentage Nonproficient (Nonproficient on State Test = 23%)	AUC
DRP Fall	.74	32%	.797
DRP Spring	.79	28%	.857
DIBELS Fall	.66	19%	.827
DIBELS Winter	.78	26%	.832
DIBELS Spring	.77	25%	.841
4Sight Fall 1	.72	34%	.816
4Sight Fall 2	.79	24%	.856
4Sight Winter	.76	18%	.852
4Sight Spring	.78	10%	.855

The value of each screening measure can be further evaluated by considering the measure's sensitivity and specificity, as indicated in Table 7. To do this, we must tabulate all the cases in the sample and identify whether the screening measure was passed or failed and whether the criterion measure was passed or failed.

Table 7. Predictive Value of Screening Tools as Determined by Sensitivity, Specificity, and Likelihood Ratios

Screening	Sensitivity[a]	Specificity[a]	Positive Likelihood Ratio[b]	Negative Likelihood Ratio[b]	Posttest Probability of Failing Year-End Test for those who FAILED Screener[b]	Posttest Probability of Failing Year-End Test for those who PASSED Screener[b]
DRP Fall	.70	.80	3.5	.38	51%	10%
DRP Spring	.30	.66	.88	1.06	21%	24%
DIBELS Fall	.58	.89	5.27	.47	61%	12%
DIBELS Winter	.74	.89	6.72	.29	67%	8%
DIBELS Spring	.75	.91	8.33	.27	71%	8%
4Sight Fall 1	.77	.79	3.67	.29	52%	8%
4Sight Fall 2	.68	.89	6.18	.36	65%	10%
4Sight Winter	.58	.97	9.67	.45	74%	12%
4Sight Spring	.37	.99	37.00	.64	92%	16%

[a]Sensitivity = number of correctly identified positives (i.e., correctly identified students with nonproficient year-end test scores) divided by the total number of positives (i.e., total number of students with nonproficient year-end test scores).

Specificity = number of correctly identified negatives (i.e., correctly identified students with proficient year-end test scores) divided by the total number of negatives (i.e., total number of students with proficient year-end test scores).

[b]Positive Likelihood Ratio = (sensitivity) / (1 – specificity)
Negative Likelihood Ratio = (1 – sensitivity) / (specificity)
Pretest Probability = .23 (23% of students failed the year-end test)
Pretest Odds = Pretest Probability / (1-Pretest Probability) or .23 / .77 = .30
Posttest Odds = Pretest Odds x Likelihood Ratio
Posttest Probability = Posttest Odds / (1 + Post test Odds)

Sensitivity is the power of the test to detect positives and is calculated as the number of correctly identified positives (test positive plus gold-standard positive cases) divided by the total number of gold-standard positives. Specificity is the power of the test to detect negatives and is calculated as the total number of correctly identified negatives (test negative plus gold standard negative cases) divided by the total number of gold-standard negatives. False-positive errors in this example are cases that were predicted to fail the year-end accountability measure based on the screening score but actually passed the year-end accountability measure. False-negative errors are cases that were predicted to pass the year-end accountability measure based on the screening score but actually failed the year-end accountability measure.

From sensitivity and specificity values, likelihood ratios can be calculated. The positive likelihood ratio (ratio of true positives to false positives) is computed as sensitivity divided by (1-specificity). The negative likelihood ratio (ratio of false negatives to true negatives) is computed as (1-sensitivity) divided by specificity. Sensitivity pertains only to those cases that are test-positive and/or criterion-positive while specificity pertains only to those cases that are test-negative and/or criterion-negative. Thus, sensitivity and specificity cannot be considered in isolation from one another; instead, sensitivity and specificity must be considered in tandem. Likelihood ratios provide a single value that incorporates sensitivity and specificity and allow for the calculation of posttest probability for test-positive and test-negative cases. Posttest probabilities are important because knowing that a test is capable of detecting 50% of actual positives (i.e., sensitivity = .50) or knowing how much more likely a positive result is for a person who truly has a condition (i.e., positive likelihood ratio) gives us good information when selecting a test for use in a particular context, but tells users nothing about how to interpret the test findings clinically for a given case or set of cases. The posttest probability values allow users to readily communicate to teachers what the calculated probability of failing the year-end test is for students who have failed the screening test and for students who have passed the screening test in a way that is superior to positive predictive value and negative predictive value (see VanDerHeyden, 2010a, 2010b, 2011 for a more complete analysis of the limitations of positive and negative predictive value). In the range of considered screening instruments in Table 7, the data team should discuss the cost associated with the time and materials needed to conduct each screening. The lowest-cost options can be identified for fall, winter, and spring. Next, the data team should discuss and ask teachers about what other useful data may be garnered from each screening score and determine whether teachers prefer one screening over another. Finally, the data team should identify the least costly measure that provides the most useful information at each screening occasion. Given the data in Table 7, the DIBELS and 4Sight screening measures could be supported as viable screening tools for use in the school (meaning one of those two should be selected for use, and the other, along with the DRP screening measure, could and should be discontinued).

Conclusion

Student performance data offers an efficient and accurate guide for instructional actions. If one wants to know whether a program of instruction is effective, there is no better metric than the student's learning (Bushell & Baer, 1994; Deno & Mirkin, 1977). RtI is a framework for using student performance data to reach actionable conclusions for system improvement. There is no question that RtI systems work, but making them work requires that key components be implemented well. In this chapter we have highlighted key actions, including:

- conducting screening to yield high-quality data
- interpreting screening data beginning with an aerial view
- treating systemic problems with systemic solutions
- monitoring implemented solution effects and managing implementation effectively
- conducting follow-up screening to verify improvements

Key indicators of successful completion of each of these actions were provided in Tables 1–5 within the chapter and are summarized here for convenience.

Action Principles

Key Action 1: Conduct Screening to Yield High-Quality Data

a. Faculty overview has been provided and screening materials selected.
b. Screening has been scheduled to occur on a single day, and screening schedule has been planned.
c. All materials for screening are available and have been organized by class, including a written protocol for screening.
d. Trainee has been observed to correctly administer and score screening materials.

Key Action 2: Interpret Screening Data Beginning With an Aerial View

a. Trainee has ruled out school-level, grade-level, and whole-class performance problems prior to selecting individual children for follow-up assessment and possibly intervention.
b. Data have been organized by grade and by class.
c. Data have been examined for identified vulnerable or high-risk groups of students to identify potential performance patterns (e.g., high numbers of new students scoring in the risk range, disproportionately high numbers of special education students scoring in the risk range).

Key Action 3: Treat Systemic Problems With Systemic Solutions

a. Classwide interventions have been started in classes with classwide problems.

b. Data teams have examined core instructional procedures in classes and grades with systemic problems.
c. Vertical teaming has occurred across grades and schools within the feeder pattern to share screening data and systemic intervention data.
d. Lower percentages of students fall in the risk range across consecutive screenings within and across years.
e. Higher percentages of students meet the proficiency criterion on the year-end accountability measure over time.
f. Historically vulnerable students show learning gains and fall into the risk range at lower rates. Performance gaps between those at risk and not at risk are reduced with intervention.
g. All students, including those in the higher performing groups and the lower performing groups, show gains with intervention and over time.
h. Students found to be at risk become proportionate by demographics with interventions.

Key Action 4: Monitor Implemented Solution Effects and Manage Implementation Effectively

a. Interventions have written protocols available for teachers to use.
b. The teacher has been provided with all needed materials to conduct the intervention and has demonstrated correct and independent use of the intervention prior to being considered trained.
c. An in-class trainer or coach is available to model correct intervention use and provide in-vivo training for the teacher.
d. A tracking log is available showing at a glance who is experiencing intervention in the school.
e. A master schedule is followed to deliver classwide, small-group, and individual interventions.
f. Weekly progress monitoring data are collected for all children experiencing intervention.
g. Progress monitoring data are graphed, and interventions are adjusted weekly with in-class support where growth is not occurring as anticipated.

Key Action 5: Conduct Follow-up Screening to Verify Improvements

a. The data team organizes data across consecutive screenings to show a reduced risk status accompanying the intervention.
b. Intervention effects are monitored for vulnerable or at-risk students over time.

References

Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovaleski, J. F., Prasse, D.,...Tilly, W. D. (2005). *IDEA 2004 and response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education.

Bradley, R., Danielson, L., & Hallahan, D. P. (Eds.). (2002). *Identification of learning disabilities: Research to practice*. Mahwah, NJ: Lawrence Erlbaum.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analysis of response-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23*, 381–394.

Bushell, D., & Baer, D. M. (1994). Measurably superior instruction means close, continual contact with the relevant outcome data. Revolutionary! In R. Gardner, D. M. Sainato, J. O. Cooper, & T. E. Heron (Eds.), *Behavior analysis in education* (pp. 3–10). Belmont, CA: Wadsworth.

Codding, R. S., Chan-Iannetta, L., Palmer, M., & Lukito, G. (2009). Examining a classwide application of cover-copy-compare with and without goal setting to enhance mathematics fluency. *School Psychology Quarterly, 24*, 173–185. doi:10.1037/a0017192

Daly, E. J., III, Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*, 554–574.

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

Donovan, S., & Cross, C. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.

Fixsen, D. L., & Blasé, K. A. (1993). Creating new realities: Program development and dissemination. *Journal of Applied Behavior Analysis, 26*, 597–615.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*, 174–206.

Individuals with Disabilities Education Act of 2004, Pub. L. No. 108-466. (2004).

Kovaleski, J., VanDerHeyden, A. M., & Shapiro, E. (in press). *The RtI approach to evaluating learning disabilities*. New York, NY: Guilford.

Noell, G. H., & Gansle, K. A. (2006). Assuring the form has substance: Treatment plan implementation as the foundation of assessing response to intervention. *Assessment for Effective Intervention, 32*, 32–39.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. 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.

O'Connor, R. E., Fulmer, D., Harty, K., & Bell, K. (2005). Layers of reading intervention in kindergarten through third grade: Changes in teaching and child outcomes. *Journal of Learning Disabilities, 38*, 440–455. doi: 10.1177/00222194050380050701

Shapiro, E. S., & Clemens, N. H. (2009). A conceptual model for evaluating system effects of response to intervention. *Assessment for Effective Intervention, 35*, 3–16. doi: 10.1177/1534508408330080

Vanderbilt Kennedy Center. (n.d.). *Peer-assisted learning strategies* (PALS) [Website]. Nashville, TN: Author. Retrieved from http://kc.vanderbilt.edu/pals

VanDerHeyden, A. M. (2010a). Determining early mathematical risk: Ideas for extending the research [Invited commentary]. *School Psychology Review, 39*, 196–202.

VanDerHeyden, A. M. (2010b). Use of classification agreement analyses to evaluate RTI implementation. *Theory into Practice, 49*, 281–288.

VanDerHeyden, A. M. (2011). Technical adequacy of RtI decisions. *Exceptional Children, 77*, 335–350.

VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. *American Educational Research Journal, 49*(6), 1251–1284.

Wagner, D., McComas, J. J., Bollman, K., & Holton, E. (2006). The use of functional reading analysis to identify effective reading interventions. *Assessment for Effective Intervention, 32*, 40–49.