Chapter 4

Are We Making the Differences That Matter in Education?

AMANDA VANDERHEYDEN
Educational Research and Consulting, Inc.

ABSTRACT: *This paper argues that ineffective practices in schools carry a high price for consumers and suggests that school systems consider the measurable yield in terms of gains in student achievement for their schooling effort. Student performance data can be used to evaluate efforts, set instructional targets, and plan instructional changes. The routine reporting of student achievement gains is also a very powerful way to solve the tension that can exist between stakeholders in the schools and those who must run the schools using limited resources. The data are transparent for all to interpret and any corrective action can be evaluated by all interested parties. This paper contends that the structure of multi-tiered intervention services or response to intervention (RtI) systems in schools (that is, screening, providing intervention, and monitoring progress to verify that the interventions worked as planned) offers great opportunity for determining whether or not an educational effort changes the odds of student success. Schools can and should examine whether the use of assessments and interventions in their schools reduces risk of learning failures over time for all students and for students who are thought to be especially vulnerable. Given the historically great investments that have been made in education and the current economic climate pushing for spending reductions, policy makers and local decision makers must avoid the "more is better" logic and instead seek information about which investments (assessments, interventions) yield the greatest return in student learning. Systems must also consistently engage in those actions that are demonstrated to yield a high return in terms of student learning.*

THE HIGH COST OF INEFFECTIVE PRACTICES

By my calculations, in my own son's school, taxpayers spent about $20 per student per day of the school year last year. That is an extraordinary sum of money when you stop to think about it. Between 1970 and 2007, average spending in the United States increased from $4,210 to $10, 041 per pupil per year (Snyder & Dillow, 2010). Given the well-documented history of rapidly increasing investments in public education, it is puzzling that so many Americans suggest that perhaps we are not spending enough money. Spending more money is often offered up as the best solution for making our schools more effective. What seems lost in the debate over whether we are spending too much or too little is the relevance of what we are spending the money on. There seems to be little attention paid to whether or not we are funding what works.

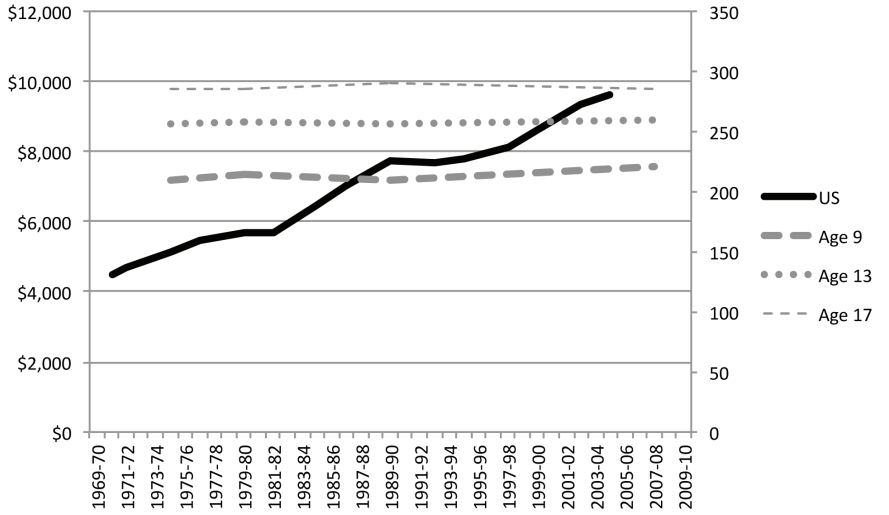

Figure 1. The relationship between spending and reading scores.

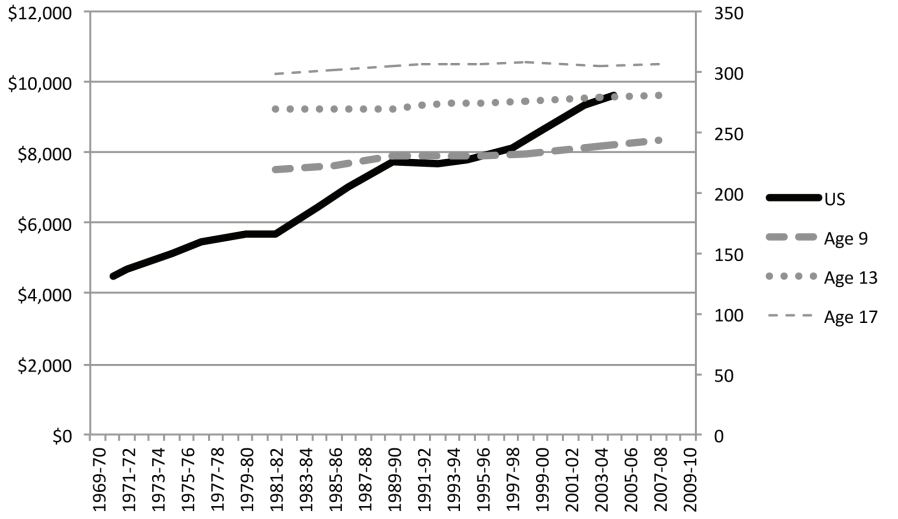

Figure 2. The relation between math scores and spending.

Part of the trouble with having a discussion about whether we are funding what works is that the goals of educational efforts have often been poorly defined (Bushell & Baer, 1994). That is, we have not determined what it means when schooling is successful or what results we want. So we place undue focus on the process and almost completely neglect the outcome, and this leads us to superficial solutions that are not necessarily causally related to improved student outcomes (e.g., reducing class size, increasing time in school). The focus on process and the neglect of outcome have fueled debate that is often filled with vitriol and passion, but is of little use to the students the debaters are supposed to be concerned about.

Focusing on process as opposed to outcome has also fueled tension between school systems and parents. Mistrust is bred and communication suffers when parents approach the school with an agenda of wanting a particular service and view the school as either giving them what they want or not. Similarly, schools may be slow to share information with parents because they do not want parents to interfere with the process they wish to use. This arrangement does not effectively serve the goal of improved learning for students.

One very common scenario involves parents pushing for a special education eligibility evaluation and special education services for their child who is struggling to learn to read, because the parents believe that special education will lead to a better outcome for their child. Unfortunately, this belief does not reflect the realities of special education in schools; special education has not

been shown to have a significant effect on the learning outcomes of students served under the category of Specific Learning Disability (Kavale & Forness, 1999). Another example involves communities advocating for smaller class size. On the surface, smaller class size sound good to everyone, but when you view the available resources as a single pie that must be divided for the greatest good, then something that seems desirable might not be worth the cost if it means not being able to implement another strategy that has been shown to improve achievement. When resources are allocated to one effort, they are not available for another effort.

Student outcome data can take the heat out of these debates about resource allocation, because any resource allocation decision simply becomes a hypothesis to be tested, and the action will be continued only if it returns the desired results. Using student data to inform resource allocation decisions has a number of important effects. First, it focuses decision makers on attaining improved learning outcomes. Second, it increases the probability that the decisions will favor actions that have been shown to successfully improve learning in the past or in other schools. Third, and perhaps most important, it creates an opportunity for decision makers to make midstream adjustments to implemented strategies to ensure that they return the desired effect. Selecting something that is likely to work is a good first step, but once something is implemented, the most important function of leadership is ensuring that desired outcomes are reached and sustained over time.

Student learning is the most fundamental outcome of schooling

Student learning is the outcome that schools and communities should prioritize. The purpose of the school is to ensure learning. This purpose is not at odds with big-picture questions that parents might care about: "Is my child happy at school?" "Does my child like learning?" "Is my child developing positive relationships with teachers and students and learning how to function well away from our home environment?" Rather, learning and growth of students is a powerful—perhaps the most powerful—indicator the school is a healthy, productive environment that supports students in engaging in learning tasks they can successfully complete (Hattie, 2009). Being successful at learning in school fosters a sense of well-being in the student and improves parent-school bonding. When resource allocation is based on data and the effective actions are emphasized, precious time is preserved and is thus available for socialization, recreation, and rest during the school day. Children have a right to effective instruction and a well-rounded schooling experience that fosters the development of the whole child. Many argue that effective instruction is the best path to that end (Barrett et al., 1991).

Focusing on a simple, measurable outcome like learning gives consumers, teachers, and other decision makers a compass. Thus, activities that promote learning become valued activities that warrant further investments of time and resources. Activities that do not promote learning receive less priority. Measuring process targets such as number of hours allocated to math instruction is much less meaningful and direct than tracking whether or not students are learning and growing in math proficiency. The value of the school's effort can be evaluated in terms of student skill proficiency, growth in achievement over time, and reduction of performance gaps between groups of students at baseline or when instruction begins. The yield of the effort can be computed as the positive effect on learning divided by the cost in per-pupil spending.

Selecting achievement as the fundamental outcome of schooling is logical and viable (Barrett et al., 1991; Hattie, 2009). Consensus for achievement as the primary outcome has emerged over the past three decades and is reflected in policy efforts promoting standard content expectations (National Mathematics Advisory Panel, 2008; National Reading Panel, 2000) and accountability legislation (No Child Left Behind Act [NCLB], 2001) that are intended to demonstrate that educational services enhance student outcomes over time. Research trends reflect a shift from correlational (where conclusions about causal relationships cannot be reached) to experimental (where conclusions about causal relationships can be reached) research, and syntheses of existing research studies (e.g., Hattie, 2009; Slavin & Lake, 2008) provide excellent direction for practitioners who wish to use educational strategies that will be of highest yield for students.

Bad decisions are not benign

When decisions are made to allocate educational resources in ways that do not yield achievement gains, the cost is greater than consumers might suspect. When a school chooses to use an ineffective strategy, it bears the tangible cost of materials and training for the new strategy. But the cost does not stop there. There is also the cost of lost opportunity to do something that would have better served the achievement goal; for example, lost instructional time, teacher absence from the classroom to participate in professional development for the new strategy, and substitution of the new strategy for an existing strategy that may have been higher yield. But perhaps the greatest cost comes in creating a legacy in the school that teachers will be asked to use unproven strategies, and when those strategies fail the program will be abandoned and replaced with a new initiative. This approach creates a culture of "attempt-attack-abandon" (D. Deshler, personal communication, August 23, 2008) that is highly detrimental to a productive program-improvement system that all schools and districts should use. All educators and community stakeholders must

understand that bad decisions (i.e., decisions to allocate resources in ways that do not return the desired effect) are not benign and can result in an apathetic teaching environment in which teachers just push through some new effort until it is replaced by some new mandate.

In Figure 3, an example of bad decision making is shown. In this particular school, a decision was made to implement a new mathematics program just as the school was experiencing a strong upward trend in mathematics achievement. Of course, the data below are not experimental and no causal conclusions should be reached about the efficacy of the new program, but the need for a new program can and should be rightfully questioned when achievement is trending upward. Similar mistakes happen with great frequency in systems where decision makers decide to adopt a new program without local evidence to show that it can work to serve the needs of students in the district or even that the new program fits the needs of the school.

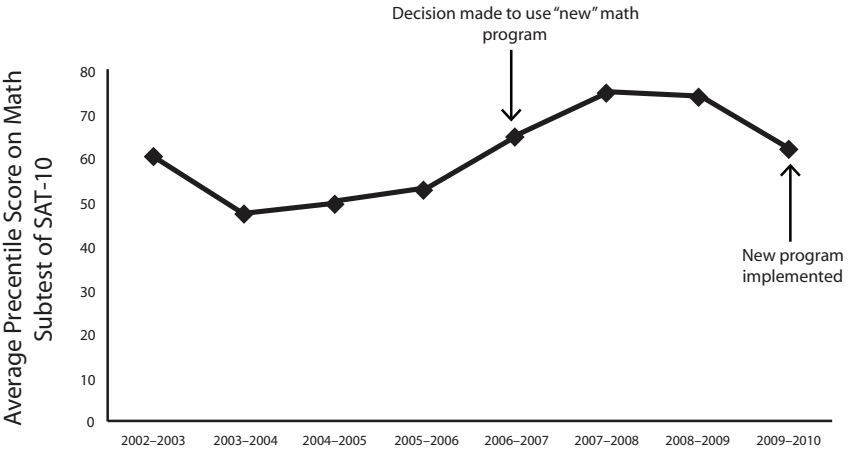

Figure 3. Third-grade math achievement.

Response to Intervention (RtI)

The use of student performance data, collected during the course of instruction, is an ideal basis for determining where resources are needed to improve learning outcomes. Systems of using student performance data to make resource allocation decisions that improve learning for the greatest number of students

are referred to as response to intervention systems. RtI is not a product. It cannot be purchased. It is a decision-making process that uses student performance data as the ever-present arbiter of all instructional decisions. Teaching can be like flying at night in poor visibility without navigational instruments to tell you how far you are from the ground, how far you are from your target, and whether or not you are moving toward or away from your target. When there is no easy way to monitor the effects of instruction and make adjustments, the likelihood is high that the instructor will miss the target altogether for many of the students.

The use of student performance data as a basis for evaluating instructional efforts, setting instructional targets, and planning instructional changes is also a very powerful way to solve the tension that can exist between stakeholders in the schools and those who must run the schools using a limited number of resources. The data are transparent for all to interpret, and any corrective action can be evaluated by all interested parties.

The questions that should guide instruction at the classroom, school, and district level are:

1. Are we making the differences that matter? Are we changing the odds of student success?

2. If we are not changing the odds of student success, what are we going to do about it?

These questions have quantifiable answers. I would like to return to the questions at the end of this article after first describing how student performance data can be used to improve learning through RtI and then describing two of the most common barriers to the effective use of RtI.

RtI has enjoyed widespread popularity as a framework for using student performance data to set system improvement targets and attain system improvements. Implementers must have data to determine risk; identify systemic problems; plan instructional changes systemwide; plan interventions for individuals, small groups, and whole classes as a supplement to core instruction; and evaluate intervention effects and inform future resource allocation decisions. RtI can be used to reduce unnecessary evaluations, initiate and sustain instructional changes that produce the desired improvements in learning, and improve learning outcomes for all students. Most states report partial or full implementation of RtI. However, certainly not all implementations return similar results. Unfortunately, the potential for improved results is often lost to implementation errors. In the section that follows, I will discuss two of the most common errors made during RtI implementation.

Weighing a cow does not make it fatter (or the prevalent problem of overassessment)

It is amazing to consider that 10 to 20 years ago assessment of student learning occurred only rarely. Now children participate in a great deal of assessment. There is no doubt that assessment is necessary to improve learning, and therefore it is not surprising that nearly all prominent policy documents related to improving outcomes in education feature routine student assessment as an essential recommendation (e.g., National Reading Panel, National Mathematics Advisory Panel). Installing technically adequate and well-implemented student assessments into schools is the first stage of RtI implementation. This effort has been speeded along in most school systems via requirements and funding provided through a statewide Reading First initiative and year-end accountability assessment. Many instructional products integrate student assessment into their materials and procedures. Hence, routine assessment of student performance is now commonplace in most schools.

Assessment is absolutely essential to make instructional decisions that improve instructional targets, but *too much assessment is detrimental to instructional systems because assessment alone will never improve achievement*. Frequent assessment is useful when it leads to a different instructional action the next day as a result of the assessment. That is, when the data are used formatively, it is reasonable to expect achievement gains (Fuchs, Fuchs, Mathes, & Simmons, 1997; Yeh, 2007). Yeh computed effect sizes for student achievement under the following conditions: frequent student assessment (two to five times per week), 10% increase in per-pupil spending, voucher programs, charter schools, and increased accountability. He then computed the cost for each approach. Frequent student assessment was 4 times as effective as increased spending per pupil, 6 times as effective as vouchers, 64 times as effective as charter schools, and 6 times as effective as increased accountability, even after accounting for the increased costs associated with conducting frequent assessments. Hattie (2009) found that formative evaluation is one of the most reliable and powerful ways to improve student achievement, yielding an average effect of $d = 0.90$ among the 30 studies included in his analysis.

From the list of actions that must be performed well to use RtI, screening and progress monitoring are tasks that most schools do not struggle to implement (Burns, Peters, & Noell, 2008). In other words, most schools implementing RtI are able to accurately collect screening and progress-monitoring data. Teams struggle to "do something" with the data. That is, they struggle to interpret, plan, and deploy corrective actions, and to evaluate and troubleshoot those actions (Burns et al., 2008). Perhaps the relative ease and competence with which schools collect assessment data contribute to the error of overassessment. There is a tendency among school leaders to think if some is good, then more must be better. When I work with schools and districts to build an action plan

to improve achievement, often the first suggestion from the leadership team is to obtain more assessment data. This suggestion is so prevalent that I routinely use a slide entitled, "Schools are drowning in data and the same students still cannot read."

I recently worked with a kindergarten center and calculated for the teachers how much time they were allocating to assessment. I wish all school and district leadership teams would undertake this exercise. I find that what administrators and teachers say they will do instructionally often is not possible given the available hours of instruction. At the kindergarten center, for example, children attended about 180 days of school. If 6 hours were used solely for instruction in all of those 180 days, then teachers had about 1,080 hours of usable instructional time for the year. Teachers reported spending about 120 hours assessing skills over four reporting periods (two report cards and two midterm reports), 10 hours per year screening, 15 hours per year monitoring progress for low-performing students, and 6 hours per year on end-of-unit tests. Hence, teachers were spending a total of 151 hours per class per year on assessment activities. If we assume that teachers were using 100% of the balance of available time for instruction (which is not possible because teachers must leave some time for transitions, non-instructional routines, and enrichment), then they were spending 14% of available instructional time on assessment. Whether or not this allocation of resources to assessment is an investment that is well spent is a question for which there is a definitive answer, but few schools seem to raise the question or look at their data in this way.

All schools should list all assessments used in the school, identify the decision that will be made from each assessment, and determine which assessments are redundant and which are not actually contributing data needed to inform instructional actions. Overassessment is a costly error that comes at a direct and substantial cost to instruction.

Implementation failures are sentinel events but usually go undetected in education

In medicine, the term "sentinel event" is defined as "an unexpected occurrence involving death or serious physical or psychological injury, or the risk thereof. Serious injury specifically includes loss of limb or function. The phrase, 'or the risk thereof' includes any process variation for which a recurrence would carry a significant chance of a serious adverse outcome. Such events are called 'sentinel' because they signal the need for immediate investigation and response." (The Joint Commission, 2011). The aviation community closely examines failures with the explicit goal of preventing those failures from occurring in the future. Defining and attending to events that come at a high

cost to the stated goals of a profession (e.g., death or injury that could have been prevented in medicine, where the goal is to promote health and well-being) is a testament to the commitment of a profession to attain its stated goals. It is not pleasant to acknowledge, let alone study, our failures, but education would do well to follow the examples of medicine and aviation.

One of the most common RtI failures involves overemphasizing intervention selection and underemphasizing intervention management (VanDerHeyden & Tilly, 2010). In RtI, every decision and action leading up to intervention may occur perfectly, but if the intervention is not implemented correctly for a consistent period of time, the intervention will fail and student learning will not improve. Research tells us that intervention failures should be exceedingly rare events (Torgesen et al., 2001; VanDerHeyden, Witt, & Gilbertson, 2007). A high rate of intervention failure is a sure sign of intervention implementation error. Many research teams have highlighted the persistent challenge of deploying interventions accurately and effectively outside of research settings (Fixsen & Blase, 1993; Witt, Noell, LaFleur, & Mortenson, 1997). Despite these data, research teams have also documented the careless disregard most interventionists and intervention researchers pay to monitoring the degree to which an intervention was correctly implemented when reaching a conclusion about the intervention's effectiveness (Gresham, Gansle, & Noell, 1993; McIntyre, Gresham, DiGennaro, & Reed, 2007).

The lack of attention to implementation accuracy is puzzling given the likelihood that poor intervention integrity will threaten or weaken intervention results and lead to inaccurate conclusions about an intervention's value in improving learning for a student or a class of students. Integrity failures are sentinel events in education. It is a sentinel event because the decision errors lead directly to the allocation of unneeded additional resources, the abandonment of a strategy that might have worked had it been implemented correctly, the use of more costly and probably more restrictive interventions for the student, and an inaccurate belief about a child's capability for learning.

One important lesson from implementation research is that often implementations fail for seemingly simple reasons that would be relatively easy to address if only someone were paying attention to the indicators. Common causes of implementation failure include not having ongoing access to a person who knows how to implement an intervention, the child not being available for intervention sessions due to scheduling problems, intervention error (e.g., modeling too rapidly, failing to give corrective feedback to the student), not having the right materials available, a belief on the part of the implementer that implementation is not being tracked and is not important, and no one tracking and troubleshooting intervention effects. It is important to remember that intervention failures should be rare events. Hence, a very simple approach to monitoring integrity is to track student learning outcomes. Where student

learning outcomes are not improving, implementation error should be investigated and ruled out or repaired before changing the intervention (Gilbertson, Witt, Singletary, & VanDerHeyden, 2008; Witt, VanDerHeyden, & Gilbertson, 2004).

CONCLUDING COMMENTS: ARE WE MAKING A DIFFERENCE?

We know a great deal about how to improve instruction and learning (Hattie, 2009). When we know what works to improve achievement, why do so many school systems struggle to put these strategies into practice in classrooms? I believe our failures have had little to do with measurement or pedagogy or many of the other causes we tend to focus on and discuss. I believe we have consistently failed to use data to guide instruction and then deliver that instruction well. When children fail to learn the skills we expect them to learn, our strong tendency, historically and persisting today, is to attribute those failures to them (Ysseldyke, Algozzine, & Epps, 1983).

To attain improved learning outcomes, implementers should use student performance data to guide resource allocation decisions. RtI systems provide an excellent framework for doing so, but the results obtained depend entirely on how well the system is followed. Implementers must minimize assessment requirements, collecting only the data needed to make the instructional change that will move the students and school system closer to its targets. Implementation must be monitored closely to ensure that the decisions made are high yield. For implementers, smarter decision making will allow them to work with greater quality, intensity, and consistency because they can discontinue unnecessary and unfruitful efforts.

In Figure 4, the progress of a whole class of students can be tracked to ensure that learning gains are being made toward the instructional goal. Students who lag behind once the class as a whole reaches mastery can be provided with small-group or individual intervention.

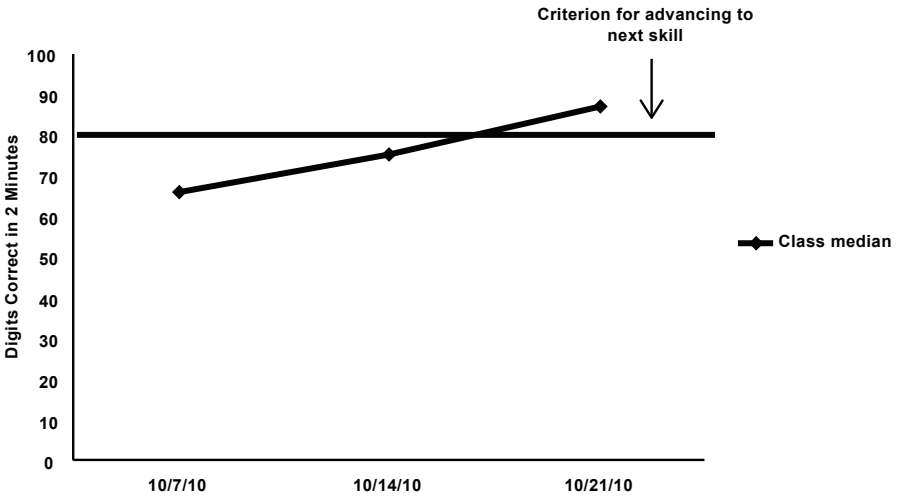

Figure 4. Digits correct with mixed addition and subtraction.

Where progress is monitored consistently across classrooms (e.g., where several classwide learning problems have been detected), the rate of mastery of skills can be tracked across classes to identify classes whose scores are lagging behind other classes participating in similar instruction or whole-class supplemental intervention. An on-site support person (e.g., coach, resource teacher) can go to those classes that are lagging, observe instruction, coach the teacher, and provide performance feedback to improve the efficacy of instruction. In this example (Figure 5), classes 9 to 11 should receive in-class support to improve the efficacy of the intervention.

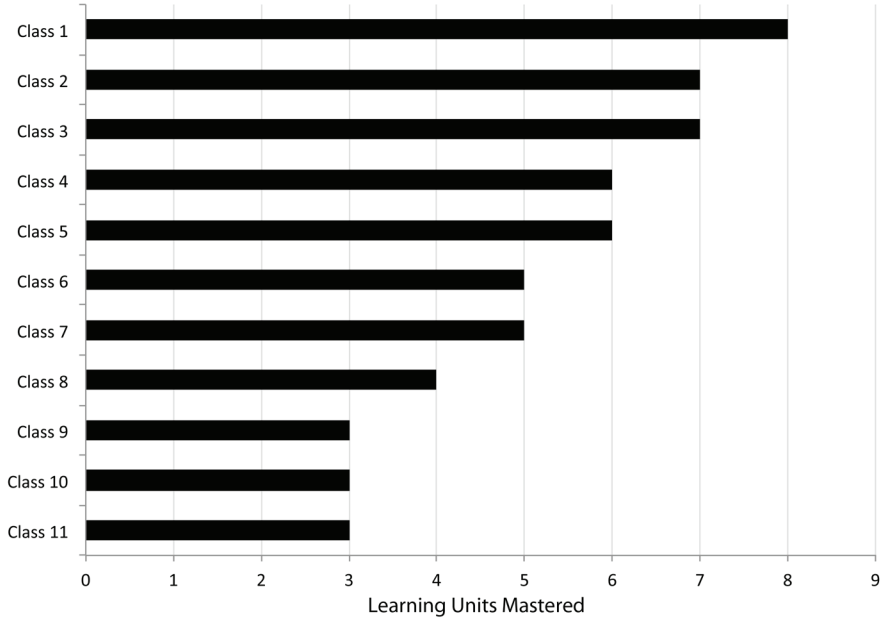

Figure 5. Learning units mastered across classes.

Follow-up screening data can be used to verify that over time instructional efforts are reducing the number of students at risk (Shapiro & Clemens, 2009). In Figure 6, each pair of bars shows the fall screening and winter screening for each teacher at first grade. The dark gray portion of the bar shows the percentage of students performing in the frustrational range, the white portion the percentage of students in the instructional range, and the light gray portion the percentage of students in the mastery range. This type of graph is highly useful to school and district leaders in generating an action plan for improvements (e.g., providing whole-class intervention versus small-group) and for evaluating and troubleshooting the improvement efforts that are already underway (e.g., giving Teacher 6 in-class support because his class is the only one that did not show a marked reduction in the number of students scoring in the frustrational range from fall to winter screening).

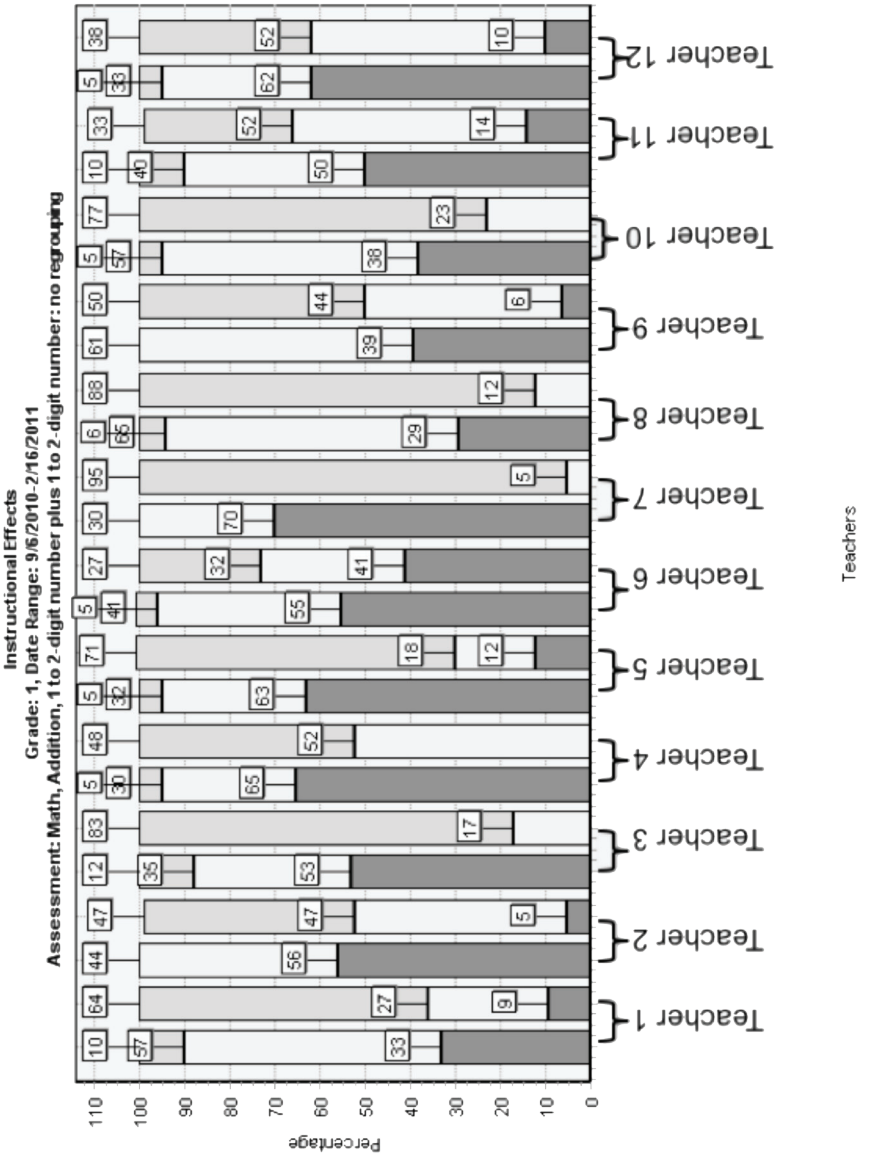

Figure 6. All first grade classes fall and winter screening for reading fluency.

We must have data to answer the two key questions raised earlier in this paper: Are we making differences that matter? And if we are not making a difference, what are we going to do about it? Figures 4 to 6 illustrate the type of data that can be collected to plan corrective actions and to evaluate and ensure the success of those actions over time. With data, any strategy can be tested

and those data can be shared with stakeholders in ways that are transparent and help everyone understand the basis for future actions and resource allocation decisions.

If we consider education as a good or service for which cost and effect can be quantified, we can track the yield of our efforts over time. In Figure 7, it is easy to compare the probability of reading success (dark gray area) and failure (white area) with supplemental intervention (left-hand column) and without supplemental intervention (right-hand column). The top row shows the probabilities of outcomes in a high-achieving school, and the bottom row the probabilities of outcomes in a low-achieving school. The probability of reading proficiency in schools with intervention is greater in both high and low achieving schools. However, decision makers must also consider the cost of providing intervention. Intervention value can be examined by computing yield per cost in each school, with and without supplemental intervention.

In high-achieving schools, the probability of passing the year-end test is 0.80 without supplemental intervention. When supplemental intervention is added (at a cost of 1.5 times the cost of general instruction or 30 instructional minutes added to 60 minutes provided during core instruction), the yield per cost is computed as 0.80 (0.80/[(1 x 100)/100]) probability of passing the year-end test in the school without supplemental intervention and 0.82 (0.90 / [(1.5 x 20) + (1 x 80)/100]) probability of passing the year-end test in the school with supplemental intervention. These data help decision makers understand that the added cost of intervention may be worthwhile since it increases the probability of reading proficiency even after accounting for the cost of providing the intervention.

In the low-achieving schools, the yield per cost of intervention analysis makes the decision very straightforward. If intervention is provided to 50% of students in the low-achieving school, then the yield per cost (expressed as the probability of reading proficiency) is 0.68 (0.85/[1.5 x 50) + (1 x 50)/100] which is superior to the probability of reading proficiency when no intervention is provided (0.60), even after accounting for the added cost of intervention. When only 20% of students are provided with the supplemental intervention (for example, when the system makes efforts to improve the efficacy of core instruction prior to beginning supplemental intervention), then the yield per cost analysis provides stronger evidence of value (0.77 with intervention versus 0.60 without intervention).

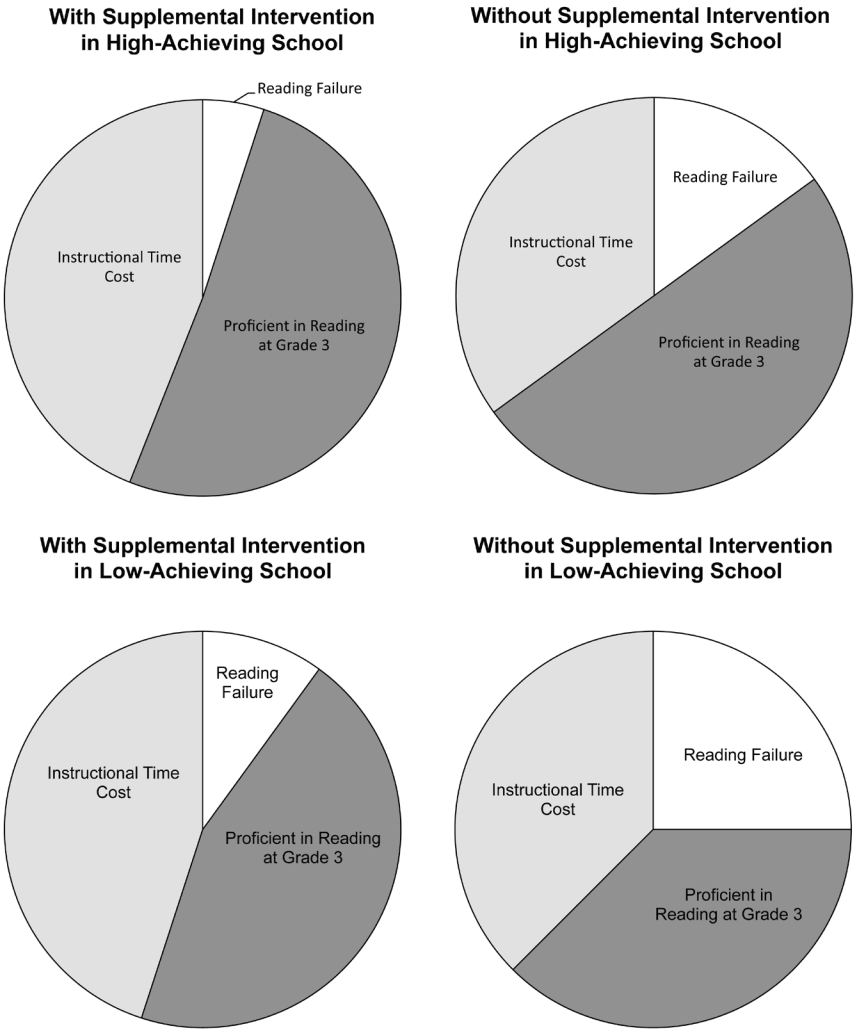

Figure 7. The probability of Grade 3 reading proficiency with and without intervention.

RtI data can be used to advance student outcomes if decision makers collect only the data that are needed to make instructional adjustments, make those adjustments with fidelity, and track their implementation to avoid common implementation errors. A controversial article appeared in the *New York Times* under the title "Can Cancer Ever Be Ignored?" (Brownlee & Lenzer, 2011). It was written in response to an expert medical panel's opinion to not recommend routine prostate screening. At the heart of this issue is the near-universal

belief that if some is good, more must be better—and consequently America's demand for more medical screening and treatment. In medicine, this belief is so prevalent that anything less than patients getting the list of diagnostic tests that they want has been described as "un-American" and compared to rationed health care (Brownlee & Lenzer, 2011).

Consumer-driven assessment and intervention in medicine actually makes for bad medicine because it is equivalent to overassessment and overtreatment. To wit, the rate of false positive errors associated with prostate screening is so high that being exposed to the screening can do more harm than good. Why? Because a positive screening leads to a more invasive medical procedure that can cause impotence, incontinence, or death. And the chances of actually dying of prostate cancer are very low among those who have prostate cancer. Overall, there is a greater risk of harm in having the screening than in not having the screening. Of course, for the individual with an aggressive prostate cancer, early detection matters, but looking for this individual among symptom-free adults causes more overall harm than good. Americans don't need more diagnosis and intervention. We need smarter diagnosis and intervention.

Smart diagnosis and intervention must be guided by four types of data:

1. The prevalence of a condition. This prevalence is the basis for computing the odds of a person having or not having a condition before any assessment or intervention is begun. These odds can be adjusted to reflect increases or decreases in odds given certain symptom profiles with the logic that if a person has a symptom, then the odds of having the condition may be higher, thus changing the utility of various assessment and treatment options. So in the case of prostate screening, determining when screening is likely to be a risk worth taking is a highly individualized decision that probably should be made only after an individual has experienced symptoms that increase that person's probability of having the condition.
2. The associated false positive and false negative error rates of screening measures.
3. The probability of negative outcomes if the condition is not diagnosed and no treatment is provided.
4. The probability of negative outcomes if the condition is diagnosed and treated.

The same scenario has been playing out in education for about 30 years. When a child struggled to learn to read, parents advocated for and sought a battery of psychoeducational assessments and an ultimate diagnosis of learning disability. This diagnosis skyrocketed 260% between 1977 and 2001, hitting a peak in 2001 when 6.1% of American students were identified as having

a learning disability. There were many reasons to question the validity of making a learning disability diagnosis (see VanDerHeyden & Burns, 2010, for a review), but in the end the proliferation of the diagnosis reflected the public's demand for more diagnosis and intervention, and the failure of the diagnosis to change student outcomes has caused people to reconsider the value of making the diagnosis in the first place.

RtI has been touted as an alternative to a learning disability diagnosis that carries the potential for making appreciably positive changes in student learning outcomes over time. When implemented well, RtI can lower false positive errors and reduce the risk of long-term learning failures. Yet, smart decision making is required or RtI may go the way of prostate screening. Implementers must understand that more is not always better and that all decisions carry errors that can and should be quantified to guide future decision making. In RtI, screening should be used only if its use increases the odds of accurate identification of learning problems above those obtained by chance (or prevalence alone).

Universal screening measures commonly used in RtI (e.g., reading curriculum–based measurement and the Dynamic Indicators of Basic Skills) often carry high false positive error rates. Follow-up assessment procedures that can be used in RtI implementations offer a low-risk and practical way to reduce the rate of false positive screening errors. More assessment of *all* students does not improve the accuracy of screening decisions. Rather, gated screening procedures are supported where the sample is filtered and subsets of the original sample participate in additional assessment. Schools implementing RtI can and should examine whether the use of assessments and interventions reduces risk of learning failures over time for all students and for students who are thought to be especially vulnerable. Given the historically great investments that have been made in education and the current economic climate pushing for spending reductions, policy makers and local decision makers must avoid the "more is better" logic and instead seek information about which investments (i.e., assessments, interventions) yield the greatest return in student learning.

REFERENCES

Barrett, B. H., Beck, R., Binder, C., Cook, D. A., Engelmann, S., Greer, R. D., …Watins, C.L. (1991). The right to effective education. *The Behavior Analyst, 14*, 79–82.

Brownlee, S., & Lenzer, J. (2011, October 5). Can cancer ever be ignored? *New York Times.* Retrieved from http://www.nytimes.com/2011/10/09/magazine/can-cancer-ever-be-ignored.html?_r=1

Burns, M. K., Peters, R., & Noell, G. H. (2008). Using performance feedback to enhance implementation integrity of the problem-solving team process. *Journal of School Psychology, 46*, 537–550.

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

Fixsen, D. L., & Blase, 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.

Gilbertson, D., Witt, J. C., Singletary, L., & VanDerHeyden, A. M. (2008). Improving teacher use of interventions: Effects of response dependent performance feedback on teacher implementation of a peer tutoring intervention. *Journal of Behavioral Education, 16*, 311–326. doi:10.1007/s10864-007-9043-0

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257–263.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

The Joint Commission, (2011), *Sentinel Event*, Retrieved from http://www.jointcommission.org/sentinel_event.aspx

Kavale, K. A., & Forness, S. R. (1999). Effectiveness of special education. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 984–1024). New York: Wiley.

McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the *Journal of Applied Behavior Analysis* 1991–2005. *Journal of Applied Behavior Analysis, 40*, 659–672.

National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*, U.S. Department of Education: Washington, DC.

National Reading Panel (2000). Report of the National Reading Panel. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

No Child Left Behind Act of 2001. Public Law 107–110, 107th Cong., 115 Stat. 1425.(2002) (enacted).

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

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

Snyder, T. D., & Dillow, S. A., (2010). *Digest of Education Statistics 2010*. U.S. Department of Education: Washington, DC. Retrieved from http://nces.ed.gov/pubs2011/2011015.pdf

Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*, 33–58. doi:10.1177/002221940103400104

VanDerHeyden, A. M., & Burns, M. K. (2010). Essentials of response to intervention. In A. S. Kaufman & N. L. Kaufman (Series Eds.), *Essentials of psychological assessment*. Hoboken, NJ: Wiley.

VanDerHeyden, A. M. & Tilly, W. (2010). *Keeping RtI on track: How to identify, repair, and prevent mistakes that derail implementation*. Horsham, PA: LRP Publications.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*, 225–256. doi:10.1016/j.jsp.2006.11.004

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

Witt, J. C., VanDerHeyden, A. M., Gilbertson, D. (2004). Troubleshooting behavioral interventions: A systematic process for finding and eliminating problems. *School Psychology Review, 33,* 363–383.

Yeh, S. S. (2007). The cost effectiveness of five policies for improving student achievement. *Journal of Evaluation, 28*, 416–436. doi:10.1177/1098214007307928

Ysseldyke, J. E., Algozzine, B., & Epps, S. (1983). A logical and empirical analysis of current practice in classifying students as handicapped. *Exceptional Children, 50,* 160–166.