NASP
NATIONAL ASSOCIATION OF
School Psychologists

School Psychology Forum:
RESEARCH IN PRACTICE

VOLUME 11 · ISSUE 3 · PAGES 77–90 · Fall 2017

The Role of Measurement Interval in Rate of Improvement Calculation

Timothy J. Runge
Indiana University of Pennsylvania

Caitlin F. Bennyhoff
Matthew R. Ferchalk
Eastern Lancaster County School District

Andrew E. McCrea
Lower Dauphin School District

ABSTRACT: Assessing a student's level and rate of improvement on an academic skill is a contemporary approach to the identification of specific learning disabilities. This approach, broadly categorized as responsiveness to intervention, however, does not obviate educators from scrutinizing the psychometric qualities of the data used to make important decisions about students. By way of illustration, we demonstrate that the measurement interval used to calculate rate of improvement can have substantial implications for evaluating the extent to which a student's rate of improvement is judged to be acceptable. Consequently, it is important for educators to fully appraise the method by which a student's rate of improvement is quantified. Suggestions for practical use are offered. Further, we pose a few questions that must be empirically tested if this contemporary approach to the identification of specific learning disabilities is to avoid the limitations of its predecessor ability–achievement discrepancy model.

It is often said that those who ignore lessons from the past are doomed to repeat them. The federal definitions of educational disabilities written in the 1970s characterized specific learning disabilities (SLDs) by a student's underperformance in an academic area compared to peers and intellectual ability (i.e., the 1975 Education for All Handicapped Children Act), which educators believed to mean a discrepancy between IQ/ability and achievement (Kovaleski, VanDerHeyden, & Shapiro, 2013). Several decades later, there is a wealth of research highlighting problems with operationalizing SLDs using the ability–achievement discrepancy and more contemporary cognitively based approaches (Sternberg & Grigorenko, 2002). This is quite alarming given that the ability–achievement discrepancy approach is still the primary method of identification in the majority of states (Maki, Floyd, & Roberson, 2015).

Fast forward to the reauthorization of the 2004 Individuals with Disabilities Education Improvement Act, and practitioners are again faced with interpreting a new federal regulatory allowance for determining students' eligibility for special education under SLD via a response-to-intervention (RTI) approach. Under this contemporary decision-making process, students whose response to empirically validated instruction and intervention is not commensurate with a priori standards are deemed to manifest an SLD.

Correspondence concerning this article should be directed to Timothy J. Runge, Indiana University of Pennsylvania, 1011 South Drive, Indiana, PA 15705; trunge@iup.edu

Within practice and the empirical literature, various operationalizations of RTI have been proposed. Van Norman and Christ (2016) noted that these operationalizations often include consideration of up to three criteria. The first criterion is evaluating rate of academic growth over time. The second criterion is evaluating the student's relative performance level compared to a grade or age standard. The third criteria typically utilized in these operationalizations is evidence of a dual discrepancy: a deficiency in both growth and level.

Fuchs (2003) proposed one of the earliest operationalizations of SLD by suggesting its classification should be predicated on postintervention level and rate of improvement (ROI) data. Termed the dual-discrepancy approach, this model conceptualizes an SLD as one in which a student has both deficient skills compared to peers and slower improvement than peers. Subsequent studies have validated this dual-discrepancy approach (e.g., Fuchs, Fuchs, McMaster, & Al Otaiba, 2003), although debate continues regarding precisely how to define atypical level and ROI (Burns, Scholin, Kosciolek, & Livingston, 2010). Further, some research suggests that these dual-discrepancy methods are no better than the IQ-achievement discrepancy approaches, resulting in similarly unreliable determination of an SLD (Francis et al., 2005; Hale et al., 2010). Ardoin, Christ, Morena, Cormier, and Klingbeil (2013) completed a review of existing literature to determine if empirical evidence exists for decision rules such as those embedded within dual-discrepancy approaches for determining special education eligibility. The authors concluded that decision rules such as trend line analysis are based on expert opinion and have not been empirically validated. Consequently, contemporary approaches to identifying SLD suffer from challenges just like traditional SLD identification models.

The contemporary decision-making process of evaluating responsiveness to intervention is often associated within a school reform effort known as multitiered system of support (MTSS). These two terms are often used interchangeably despite being unique but related constructs. MTSS is a service delivery approach in which students are provided varying frequencies and intensities of instruction and support based on their demonstrated needs (Fuchs & Deshler, 2007). Within the MTSS model, a student's responsiveness to intervention is used to determine whether an SLD is present. Lack of responsiveness to intervention is often operationalized as a dual discrepancy: atypical growth and level. Therefore, MTSS is used when referring to service delivery systems in schools whereas RTI refers to evaluating data to determine whether a student has an SLD.

Maki et al. (2015) found that only a small number of states ($N = 8$) requires RTI models as the method of determining an SLD. However, most other states permit RTI models among a few other methods for eligibility determination (e.g., pattern of strengths and weaknesses). For states that permit, but not require, RTI models as a method of determining an SLD, a problem-solving approach is typically recommended in which skill deficits are defined, interventions implemented, and progress-monitoring data collected to evaluate the efficacy of the intervention. Notwithstanding criticisms by some (Francis et al., 2005; Hale et al., 2010), the utilization of RTI as an eligibility decision-making approach warrants a careful understanding of various data computations and decisions made within that approach. Further, practical implications of data analysis and decision making within RTI must be understood regardless of whether or not a practitioner works in a state requiring such practices for SLD identification. RTI approaches can be used to systematically identify students in need of academic and behavior supports and provide evidence-based interventions regardless of whether or not such approaches are used for special education eligibility decision making. While it is important to further refine the operationalization of deficient level and ROI, our purpose is to raise awareness among practitioners and researchers that the measurement interval used in the computation of ROI may have substantial effects on how ROI is interpreted.

KNOW HOW ROI IS CALCULATED

For the purposes of this article, ROI is defined as the slope of the line generated from data collected on student performance (e.g., reading) that have been plotted on a time series graph. While there are a number of different methods used to calculate ROI, the consensus is that a regression formula using the ordinary least-squares method (OLS; Rao, 1973) provides the most psychometrically defensible indicator of growth (Hixson, Christ, & Bruni, 2014; Jenkins, Graff, & Miglioretti, 2009). The OLS method, however, is not without its limitations including an assumption that growth is linear and its vulnerability to outliers (Haupt, Lösel, & Stemmler, 2013). A regression line calculated using OLS minimizes the cumulative vertical distance between each data point and the regression line. Put another way, this method minimizes the difference between the observed data and those predicted by the line of best fit to the dataset (Rao, 1973).

A regression line to a dataset takes the form $y = mx + b$, whereby b is the point at which the regression line intercepts the y-axis. The m value is interpreted as growth, or ROI, on the dependent variable, labeled x. What is important to fully appraise in the interpretation of m is the measurement interval used in its derivation. That is, m is interpreted based on the interval of time between the observed (i.e., collected) data points used to generate the regression formula. This understanding is no small matter and one that, if overlooked, may lead to very different interpretations of growth. For example, if m were calculated using data collected every day, then the interpretation should be of daily growth. If m were calculated using data collected once a month, then the growth estimate would be contextualized as monthly growth. As one can imagine, knowing the data collection interval has substantial implications for how growth is interpreted.

This awareness is particularly important for RTI in which a student's progress on some skill is monitored regularly and then decisions about adequacy of response are judged. In fact, contemporary approaches to identification of SLD using RTI (Kovaleski et al., 2013) are predicated on assessing the extent to which a student is making adequate progress when provided empirically validated instruction and intervention with fidelity. Data indicative of inadequate response to instruction and intervention are, in large part, used to establish atypical growth reflective of a disabling condition. With this criterion carrying considerable weight in the decision-making process in many states, it is important to be fully cognizant of how ROI is calculated and then interpreted.

Contemporary approaches to identifying students' academic and behavioral needs require that ROI decisions be made using psychometrically appropriate practices. To do otherwise would be repeating some of the very concerns that have plagued the traditional ability–achievement discrepancy models, including inconsistent application of state or federal eligibility criteria (Gresham, MacMillan, & Bocian, 1996), lack of uniform application across districts and states (Gottlieb, Alter, Gottlieb, & Wishner, 1994), and disagreements regarding the most statistically sound discrepancy model (Peterson & Shinn, 2002). So failure to establish consensus on best practice for interpreting ROI will likely result in the same ability–achievement discrepancy problems against which so many of us have railed.

Educators collecting progress-monitoring data must be aware of the measurement interval used to generate the regression formula and interpret its accompanying ROI. It is hoped that practitioners are fully aware of this. However, recent experiences gave us considerable pause because the ROI we calculated proved to be a function of the tool used to calculate the regression and the data collection interval utilized within these tools. Specifically, we plotted curriculum-based measurement (CBM) data for oral reading fluency (ORF) using Microsoft Excel and ChartDog (http://www.interventioncentral.org) and, depending on how the measure of time was specified within those tools, obtained different growth estimates. For example, an Excel chart was developed using ORF data and corresponding calendar weeks, while a ChartDog display was created from the same ORF data with corresponding actual dates of data collection. The time series graphs from Excel and ChartDog resulted in different ROI indices, even though the ORF data were identical. We quickly learned that the distinguishing feature between the two graphs was the measurement interval used to create the graphs.

A review of the extant literature did not reveal any publication regarding how growth estimates might be different depending on the software used and time interval designated in the dataset. Consequently, we saw a need to educate others so that ethically responsible decisions about students' ROI would be made. What follows is an illustration of how the interval of time designated within a set of progress-monitoring data can result in different ROI indices. We then conclude with a general review of how some of the more popular progress monitoring tools calculate ROI and offer recommendations for the field to improve the reliability and validity of decisions made within RTI approaches.

ILLUSTRATION

A case study using ORF data is provided to illustrate how the method by which ROI is computed may have implications for interpretation of student progress. Alex is a third-grade student with a history of slow acquisition of important literacy skills, namely phonemic awareness and basic decoding skills. Alex was benchmarked using a typical CBM ORF probe (i.e., DIBELS; Dynamic Measurement Group, 2008) in the late summer/early fall of third grade. Because Alex's reading challenges have been evident since first grade, teachers decided to regularly monitor progress in the beginning of third grade even before tiered instruction was implemented. Standardized administration and scoring procedures were employed for both the benchmark assessment (median of three probes) and progress monitoring (one probe per session).

Hixson and colleagues (2014) provided a comprehensive review of best practices regarding the frequency of data collection within MTSS, as well as guidelines for interpretation of those data. Hixson and colleagues (2014), and most other experts, suggest that weekly measurement of skills is necessary to ensure the data are reliable and valid for interpretation (Christ, 2006; Christ, Zopluoglu, Monaghen, & Van Norman, 2013) and to optimally inform instructional decisions (Ditkowsky, 2009). Using these guidelines, Alex's teachers intended to monitor ORF progress every Wednesday. Beginning in the second week of October, Alex was placed in a Tier 2 intervention and monitoring of progress continued on the same weekly time interval. Data were graphed using Microsoft Excel by teachers and ChartDog by consultants.

Measurement Intervals

The typical challenges of maintaining a weekly measurement interval occurred from the beginning of the school year, including student and teacher absences, changes in daily routines (e.g., assemblies, field trips), and holiday breaks. Breaks in the school calendar were particularly evident in November and December. Regarding November, Alex's school is located in a community with a large proportion of hunters, so the Thanksgiving break is lengthened for the beginning of deer season. Likewise, nearly 2 weeks off from school occurred in late December and early January for religious and secular holidays. Given all these disruptions in the progress-monitoring schedule, Alex's skills were monitored, on average, every 7.8 calendar days, not quite the originally planned weekly schedule.

Alex's ORF data, along with relative dates and calendar and school weeks, are provided in Table 1. The relative date is a measure of the number of calendar days from the initial fall benchmarking. Calendar and school weeks are indicated given, as described earlier, holiday breaks from school that disrupted Monday through Friday instruction across contiguous weeks. Therefore, each ORF datum corresponds to a particular relative date, calendar week, and instructional week based on a more precise indicator of measurement intervals.

Estimates of Growth

Visual displays of these data are presented in Figures 1–3. Notice that the measurement interval (i.e., relative date, calendar week, school week) is the only discriminating feature between the three graphs. Otherwise, the ORF correct words per minute (CWPM) data are all identical. As applicable, data paths are omitted for ORF data collected in a noncontinuous manner per the designated measurement interval.

Next, we calculated the daily and weekly slope of the regression (i.e., ROI) using the OLS regression data from each of the aforementioned figures (Figures 1–3). To calculate weekly ROI from the relative date

Table 1. Alex's Baseline and Intervention Correct Words per Minute Data

Phase	Relative Date	Calendar Week	Instructional Week	CWPM
Baseline	1	1	1	28
	9	2	2	30
	14	3	3	31
	22	4	4	29
	30	5	5	27
	35	6	6	30
Intervention	44	7	7	31
	52	8	8	34
	56	9	9	28
	65	10	10	34
	72	11	11	25
	78	12	12	30
	92	14	13	39
	99	15	14	44
	107	16	15	47
	126	19	16	32
	133	20	17	36
	141	21	18	41
	149	22	19	50

Note. Relative date: the number of calendar days from the fall benchmark assessment; calendar week: the number of weeks from the fall benchmark; instructional week: the number of weeks of instruction per the school calendar; CWPM = correct words per minute.

chart (–0.009), for example, we multiplied the obtained ROI by 7, resulting in a weekly ROI of –0.063. Similarly, we took the calendar week ROI in baseline (–0.029) and divided that by 7 to obtain a daily ROI based on calendar week ROI, resulting in –0.004. These same conversions were made for all baseline and intervention data derived from either relative date, calendar week, and instructional week ROI measurement intervals. These results are presented in Table 2.

As evident in the obtained slopes reported in Table 2, the measurement interval by which CWPM were monitored results in different ROIs. The weekly ROI ranges from –0.029 to –0.063 during baseline and 0.933 to 1.280 during intervention depending on the measurement interval used. The magnitude of variance in ROI depending on the measurement intervals has the potential for substantially different interpretations and, perhaps, educational recommendations. During baseline, the ROI calculated using relative date was 117% larger than the ROI calculated using either calendar or instructional weeks. Similarly, intervention ROI calculated using instructional week was 37% larger than the ROI calculated using calendar week.

Christ (2006) reminds us that all measurement, including ROI indices, has error. Therefore, confidence intervals should be calculated around an obtained slope. Confidence intervals take into consideration the standard error of the slope and, when used for interpreting ROI, may lead to different interpretations of growth had such measurement error not been acknowledged. In general, increasing calls to consider such sources of error when evaluating growth over time appear in the literature (e.g., Fuchs, Mock, Morgan, & Young, 2003; Mercer & Keller-Margulis, 2015; Parker, Dickey, Burns, & McMaster, 2012). Consequently, 95% confidence intervals were built around Alex's obtained slopes for the three different measurement intervals and are presented in Table 3.

Figure 1. Relative Date as the Measurement Interval

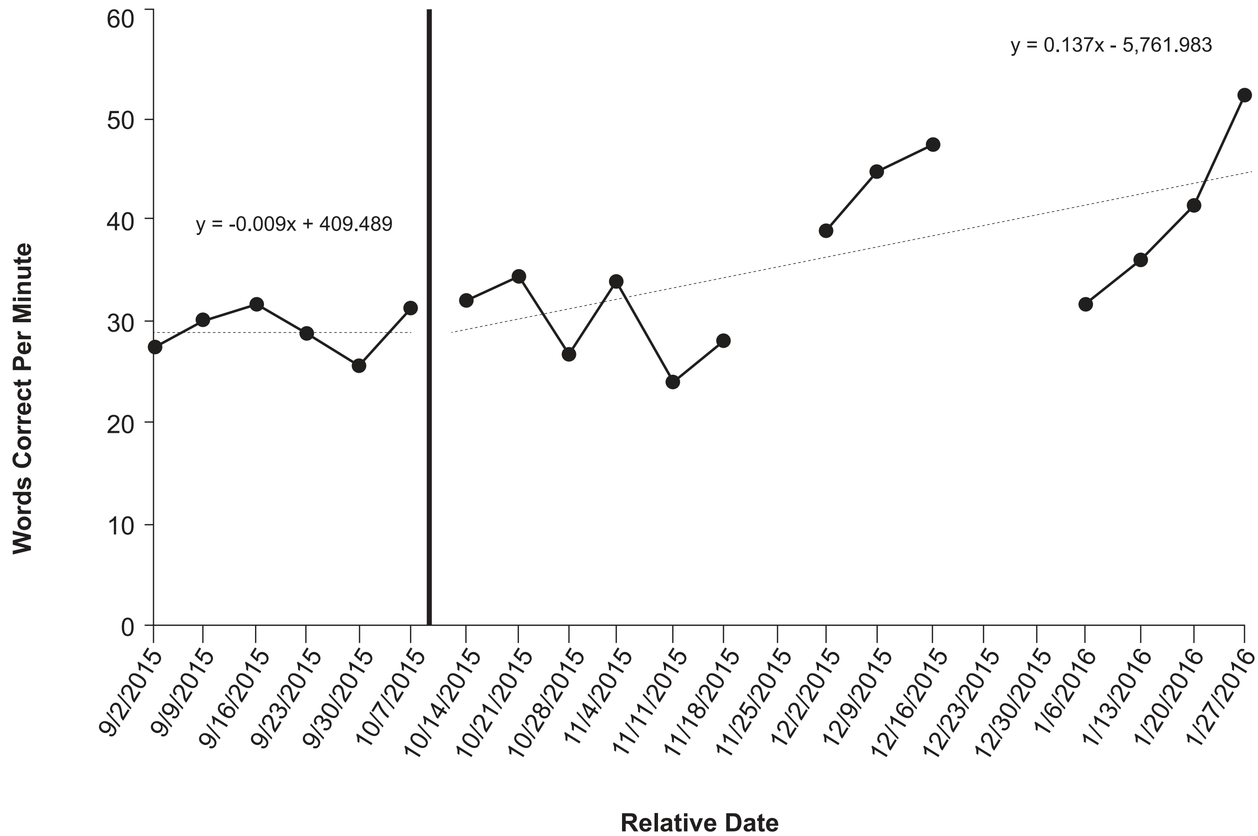

Figure 2. Calendar Week as the Measurement Interval

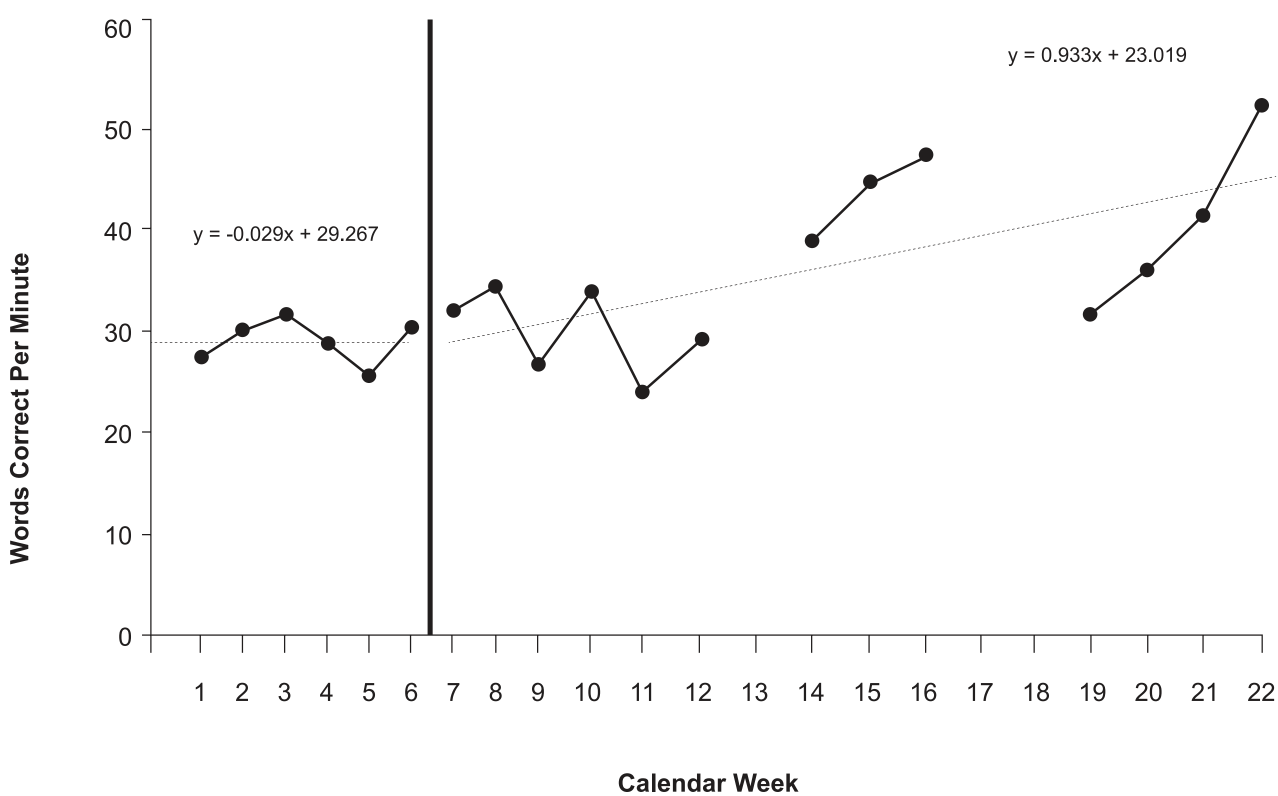

Figure 3. Instructional Week as the Measurement Interval

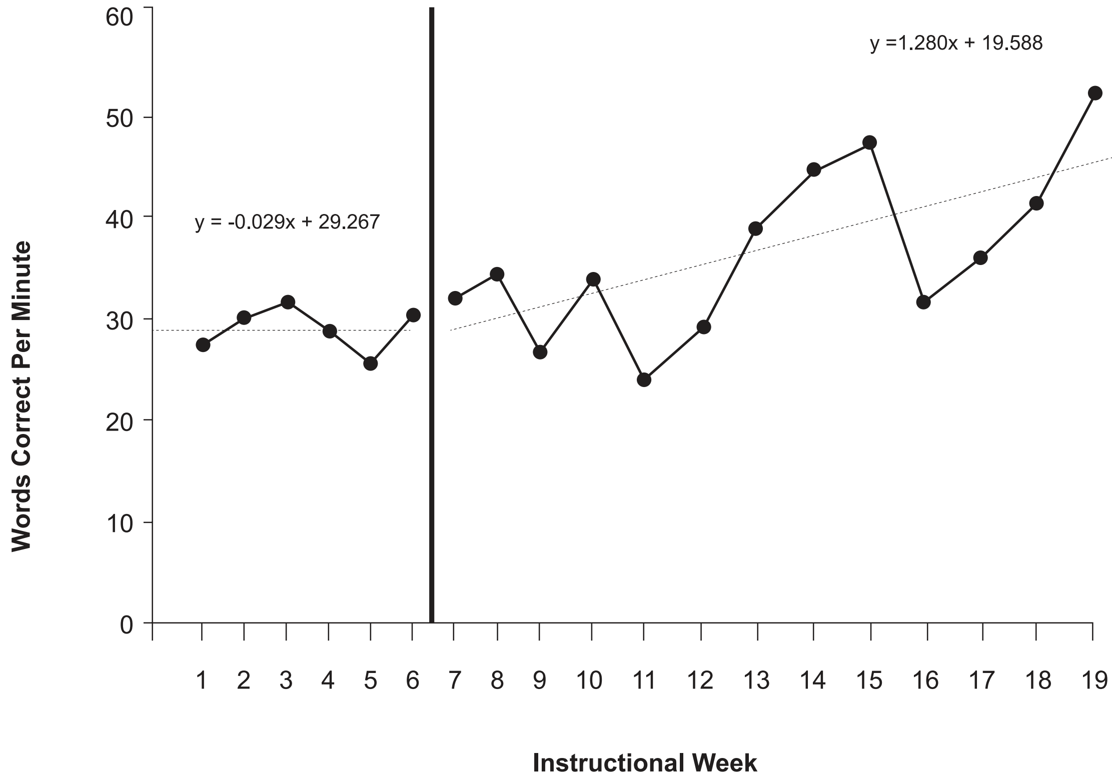

Table 2. Calculated Rate of Improvement Using Different Measurement Intervals

	Baseline		Intervention	
Measurement Interval	Daily ROI	Weekly ROI	Daily ROI	Weekly ROI
Relative date	–0.009	–0.063	0.137	0.959
Calendar week	–0.004	–0.029	0.133	0.933
Instructional week	–0.004	–0.029	0.183	1.280

Note. ROI = rate of improvement. ROI data reported as correct words per minute.

Table 3. Alex's Weekly Slope Estimates With 95% Confidence Intervals

	Baseline		Intervention	
	Weekly ROI	95% CI	Weekly ROI	95% CI
Relative date	–0.009	–1.169 – (–1.043)	0.959	0.210–1.780
Calendar week	–0.029	–1.120 – (–1.063)	0.933	0.182–1.685
Instructional week	–0.029	–1.120 – (–1.063)	1.280	0.314–2.247

Note. ROI = rate of improvement; CI = confidence interval. The 95% confidence interval was calculated using the obtained of the slope (SE*b*). All data are reported as correct words per minute.

As noted in Table 3, there is some overlap among the confidence intervals. However, they do not overlap entirely creating the potential for weekly ROI that may be different from calendar or instructional week ROI depending on the precision of entering the appropriate measurement interval. Despite this call for consideration of confidence intervals built around estimates of ROI, Christ (2006) noted that "research has yet to provide useful estimates of [the standard error of the slope] for CBM-R" (p. 129). Further, he noted that standard error of the slope decreased as more data points were used to calculate the regression line. The 13 data points during intervention used to calculate Alex's ROI, however, fall within general recommendations of at least 8–14 data points to reliably estimate slope (Christ et al., 2013; Thornbald & Christ, 2014). So it is likely that Alex's ROI may be substantially different depending on the measurement interval used in the calculations. Clearly this is an area in need of further exploration so that ethically responsible data analysis can occur regarding measuring ROI.

Some may argue that these differences in ROI as a function of measurement interval are trivial and might be comparable if figures are rounded to the nearest whole number. While that may be the case in some situations, even small differences can be compounded rather extensively if educators heed Kovaleski and colleagues' (2013) advice to consider growth trajectories when interpreting the adequacy of a student's ROI. Specifically, Kovaleski and colleagues (2013) intimate that an important question educators must wrestle with when reviewing progress monitoring data is this: How long will it take for a student to catch up to typically developing peers given the present ROI and current intensity of intervention?

Data presented in Table 4 provide estimates of how long it would take Alex to achieve the targeted criterion in subsequent triannual benchmarks for the current academic year using the ROIs calculated from the three different measurement intervals. Assuming instruction and intervention do not change, these data provide an estimate of the number of weeks it would take for Alex to reach the benchmark criterion in fall, winter, and spring. Using ROI calculated by instructional week and an average of 34 instructional weeks in an academic year, it is reasonable to predict that Alex could catch up to typically developing peers in approximately 1.5 years. However, it would take more than 2 years for Alex to reach benchmark if using relative date or calendar week to calculate ROI. It is clear, then, that fully appraising the measurement interval has implications for evaluating adequacy of ROI and predicting when students will close achievement gaps with peers.

Kovaleski and colleagues (2013) further suggest that practitioners evaluate the extent to which the student's ROI is comparable to peers' ROI and the ROI needed to achieve a particular goal in the future. In these comparisons, the student's observed ROI is referred to as the attained ROI. The ROI between two consecutive benchmarks is referred to as typical ROI. The ROI calculated from the student's current performance to the next benchmark is referred to as the targeted ROI and is the slope necessary for the student to catch up to typical peers. Kovaleski and colleagues (2013) then recommended conducting an analysis of the gap between attained ROI and both typical ROI and

Table 4. Number of Weeks to Meet Third-Grade Benchmark Criterion Using Rates of Improvement From Three Different Measurement Intervals

	ROI/week (Typical = 1.14–1.33)	Fall (Benchmark = 77)	Winter (Benchmark = 92)	Spring (Benchmark = 110)
Relative date	0.959	48	64	82
Calendar week	0.933	49	65	82
Instructional week	1.280	36	48	62

Note. ROI = rate of improvement. ROI/week data are reported as correct words per minute. Correct words per minute typical ROI provided is a range for the 45th percentile rank in fall of third grade. The number of weeks are rounded to the nearest whole number.

Table 5. Gap Analysis for Alex's Progress Monitoring Data

	Relative Date ROI	Calendar Week ROI	Instructional Week ROI
Typical ROI (A)	0.709	0.714	0.833
Targeted ROI (B)	3.067	3.067	3.833
Attained ROI (C)	0.959	0.933	1.280
Attained ROI gap analysis against typical ROI (C/A × 100)	135% of typical ROI	130% of typical ROI	154% of typical ROI
Attained ROI gap analysis against targeted ROI (C/B × 100)	31% of targeted ROI	31% of targeted ROI	33% of targeted ROI

Note. ROI = rate of improvement. ROI data are reported as correct words per minute.

targeted ROI as another set of data to consider when evaluating whether the intervention should/can be sustained or changed. If a struggling student's ROI plots a steep trajectory in which the gap with typically developing peers can be eliminated within a few months or a year, a decision may be made to continue with existing supports. A modest ROI trajectory may predict closing the gap with typically developing peers in 1–2 years, so educators have to decide whether that period of time is acceptable. If a student's ROI trajectory were relatively flat or plateaued, then a change in supports may be warranted much sooner because closing the gap with typically developing peers may not occur for 3 or 4 years if at all (Fuchs, Fuchs, et al., 2003; Kovaleski et al., 2013).

A sample gap analysis for Alex is found in Table 5. As is reflected in Table 5, the ROI and resultant gap analyses are influenced by the measurement interval. Once again, precision in calculating the ROI based on the appropriate measurement interval is necessary for the gap analysis to be valid. Consideration of confidence intervals around the attained ROI further complicates these discussions when projecting the likelihood of closing the gap many months or even a year into the future. Consequently, gauging whether there are sufficient resources to maintain or increase intervention efforts to close the achievement gap with peers (Kovaleski et al., 2013) is contingent on calculations performed using the correct measurement interval but further complicated when considering measurement error. Minimally, and until the science regarding psychometrically sound derivation and interpretation of the standard error of the slope is elucidated, we still must be cognizant of ROI differences as a matter of measurement interval. Slight differences in ROI might have substantial implications for decision making when discussing the adequacy of growth trajectories and conducting gap analyses.

CONTEXTUALIZATION IN DATA WAREHOUSE SYSTEMS

As illustrated with Alex, school teams are encouraged to be aware of the measurement interval used to calculate ROI and use that knowledge when interpreting the adequacy of responsiveness to intervention. Teams using Microsoft Excel, ChartDog, or any other tool wherein the measurement interval can be defined locally to chart data and calculate ROI must be very clear about how the data are entered into the database. Our illustration of Alex highlights how the same similar software (e.g., Microsoft Excel, Google Sheets, ChartDog, Numbers [Mac]) can produce very different ROI indices.

Many schools likely use a commercially available product for measuring and monitoring student progress. Interestingly, the measurement intervals for the majority of products calculates ROI based on a calendar week. This is true for the following products: AIMSweb (Pearson, 2012), STAR (Renaissance Learning, 2012), FAST Progress Monitoring (FastBridge Learning, 2017), Measures of Academic Progress

(Northwest Evaluation Association, 2012), Yearly ProgressPro (Stecker, 2009), and easyCBM (Anderson, Alonzo, & Tindal, 2014). DIBELS Next (Dynamic Measurement Group, 2012) does not provide a measure of ROI either through their own materials or the widely used University of Oregon DIBELS Data System. Practitioners have taken it upon themselves to remediate this gap by computing a DIBELS ROI through secondary programs like those previously mentioned (e.g., Microsoft Excel, ChartDog). Assessment scores, however, are reported in the DIBELS data system by calendar weeks not by the actual date of administration and scores downloaded from the DIBELS data system mimic the calendar week format. It is very likely that researchers and school personnel use calendar weeks when self-calculating ROI for DIBELS assessments due not to a best-practices decision but to the ease with which the scores can be obtained. Teams that use these commercially available products likely need not worry what measurement interval is used but rather should ensure that the interval remains the same over time and any growth norm used also derives from a similar interval. This may be of particular interest in locations where multiple measurement tools are used. Comparing growth on one measure to that on another—even when similar converted standard scores or lexiles are provided—would be inappropriate if the measurement intervals differed.

APPLICATION TO PRACTICE

Others have demonstrated the importance of using a regression approach to calculate growth when using CBM data (Hixson et al., 2014), especially given the likelihood of incorrect interpretations of ROI using visual analysis alone (Van Norman, Nelson, Shin, & Christ, 2013). Despite a general call for more research to refine our calculation and interpretation of ROI (Christ, 2006; Christ et al., 2013), work in this area has only recently emerged (e.g., Van Norman & Christ, 2016). Specifically, the accuracy of decisions made using ROI data is improved when educators use well-developed commercial probes that are administered and scored per standardization procedures. Robust amounts of data (e.g., 8–14), testing that occurs under ideal environmental conditions, and consideration of ROI concurrent with goal and aim lines also lead to more accurate decisions. These studies indicate the importance of using psychometrically sound CBM probes and adherence to best practices in administration, scoring, and interpretation of ROI.

Notwithstanding the work that must continue to improve decision making using ROI data, it is important that educators remain acutely aware of how the ROI data are actually computed. The current work considers one component of the OLS regression formula and demonstrates through example how changing the measurement interval from relative date to calendar week to instructional week resulted in different calculated ROI indices. Even when error was considered in the calculation of growth, the 95% confidence intervals around each ROI did not completely overlap.

The difference in results between measurement intervals likely would become clinically meaningful when the ROI is used for decision making. This would likely be the case regardless of whether those decisions are low-stakes decisions such as changing interventions or increasing the time spent in support or high-stakes decisions such as contributing to a disability classification determination. This is especially important given the sensitivity to change found in CBMs such as ORF and digits correct (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). Given that this sensitivity to change is one of the appealing characteristics of CBM, it is important to be fully appraised of how that change, or ROI, is computed.

If we return to the case of Alex, it is possible that very different decisions about intervention efficacy and/or whether the resources necessary to sustain the attained ROI are practical depending on whether relative date, calendar week, or instructional week are used to calculate ROI. Data in Table 4 suggest that Alex is unlikely to close the gap for more than 2 years if ROI were calculated by relative date or instructional week. In this situation, the team would likely decide to change the intervention completely. The same team, however, may consider sustaining the current intervention or increase its intensity if the ROI were calculated using instructional week. Given that the latency until Alex closes the gap is much less for the ROI calculated by instructional week compared to the other two ROI calculations, teams may reach different conclusions. Again, this underscores the need to be mindful of the measurement interval used to calculate ROI.

Teams computing ROI themselves (through Microsoft Excel, ChartDog, or other) should be mindful to take a consistent approach when setting up their spreadsheets, always using the same measurement interval and/or carefully importing the correct measurement interval into the database. Which interval is ultimately chosen might be dictated by any set of growth norms a team will use to make decisions, as it would be inappropriate to compute local growth with one measurement interval and compare results with norms derived from a different measurement interval. Teams using products that compute ROI (e.g., AIMSweb, STAR) must be careful not to compare growth on one instrument to that on another, even when scores are converted to common metrics such as standard scores or lexiles, especially if the measurement interval on each instrument is not the same. Further, given the illustrated differences in ROI in the Alex example, teams using commercial products that calculate ROI should ensure that the actual data collection intervals are as close to weekly as possible. Again, even modest deviations in intended weekly assessments can affect the computed ROI leading to potentially faulty decisions about responsiveness to intervention.

Given the frequent and sometimes lengthy disruptions in instructional days illustrated by the Alex case study, it would seem that using relative date would be the most valid measurement interval. In that scenario, there were 13 calendar days lost to deer hunting season and holidays in a 6-week period that would, in other months, have counted as instructional days. This equates to nearly three instructional weeks lost. So using relative date appears to be most appropriate during the months of November and December given that school's academic calendar. Contrarily, the beginning of the academic year and January through March have few disruptions, so it is possible that ROI calculated using relative date, calendar week, and instructional week may be comparable for intervals that largely cover those time periods in the academic calendar.

Another consideration is the annual window of time each spring (in most states) when high-stakes accountability assessments occur. In our state of Pennsylvania, this 4–6 week period is often filled with numerous school days spent completing various state assessments while little actual instruction occurs. We often ask ourselves if these days should count as instructional days when interpreting ROI. Therefore, we recommend using relative date as the measurement interval given disruptions in the original plan to weekly progress monitor, especially when such disruptions are magnified over the 3–5 months typically needed to obtain an adequate amount of data to estimate growth (Christ et al., 2013; Thornbald & Christ, 2014). Despite this recommendation, we acknowledge the multiple complexities of decision making with an RTI framework (Fuchs, Mock, et al., 2003) and hope the science can begin to minimize at least some of the errors we continue to make.

Perhaps in the distant future we will be calculating ROI based on actual time spent in core instruction and tiered interventions and protocols with a more thorough understanding and utilization of measurement error associated with slope. While this may seem farfetched, some computer-assisted interventions (e.g., Study Island; iReady) provide measures of intervention fidelity and/or duration. For example, if Alex were to complete 53 lessons in a computer-assisted intervention for a total of 20 hours across 13 weeks, we could calculate an ROI based on the number of lessons completed and hours of intervention. Perhaps this is too complex for practical use right now. However, such may be the future given increasing adoption of computer-assisted intervention and assessment (e.g., DeWitte, Haelermans, & Rogge, 2015) and demands for precise measurement of growth within RTI approaches, especially when making critical high-stakes decisions such as eligibility for special education. Our ethics require this level of scrutiny and care when interpreting students' data, although the science is not there just yet.

Fundamentally, our illustration underscores the importance of understanding how ROI is calculated and interpreted within decision-making frameworks. Ignorance of this could easily result in a repetition of ability–achievement discrepancy criticisms highlighted for many years (Sternberg & Grigorenko, 2002). Obviously we all want to avoid that potential recurrence. To that end, our illustration also emphasizes the need for researchers to further validate the most appropriate measurement interval and sources of error when calculating ROI and how those data are interpreted and utilized within RTI models. Until that occurs, we recommend calculating ROI based on relative date and consideration of confidence intervals to account for measurement error of the slope of the regression. Failure to do so will merely repeat, albeit in a repackaged manner, the same errors that have plagued our discipline for decades.

NEED FOR FUTURE RESEARCH

What still remains unclear, however, is empirically supported guidance on which measurement interval is the most reliable method to calculate ROI and most meaningful for interpretation. Our illustration demonstrates that the measurement interval selected can have an effect on either low- or high-stakes decisions even when confidence intervals are taking into consideration. Consequently, future research using large datasets must elucidate which measurement interval is most psychometrically sound and leads to the most valid conclusions about ROI. Echoing Christ (2006), more research is needed to understand the magnitude of the standard error of the slope and its implications on interpretation of growth.

CONCLUSION

School psychologists and teams of educators reviewing progress monitoring data, whether it be for evaluating responsiveness to intervention or special education eligibility determination, must be fully aware of the measurement interval used to create graphs for data analysis. ROI data generated from these graphs will likely be used for important decision making, and failure to understand measurement intervals in these data may result in faulty decisions about students. Despite this call for educators to be fully appraised of the measurement interval used to calculate ROI, considerable research must still be completed to improve decisions about and outcomes for students.

ADDITIONAL RESOURCES

There are a large number of resources available to assist with charting progress monitoring data and interpreting ROI. General information about response to intervention and the application of the dual discrepancy model can be found at http://www.rtinetwork.org. Various forums, videos, and practice guides are linked from this website. Similar resources can be found at http://www.rti4success.org.

Readers are also encouraged to visit http://www.interventioncentral.org for access to and instructions for ChartDog Graph Maker. We use this tool in our practice and consultation because of its relative ease of use and various statistics that are automatically computed, including ROI, phase means, and effect sizes. This website is an excellent resource for practitioners who are averse to or have limited experience using spreadsheets such as Microsoft Excel. Access is free and data for multiple students can be securely saved.

Those who use Microsoft Excel but struggle with creating appropriate graphs and generating means and regression formulas might find Barton and Reichow's (2012) task analytic set of instructions helpful for entering, plotting, and analyzing data. Cole and Witts (2015) and Dixon et al. (2009) offer task analytic instructions for users of older versions of Microsoft Excel.

Another set of fantastic resources can be located at http://www.rateofimprovement.org. This website has a number of Excel templates for elementary-grade progress monitoring, how-to guides for entering data and calculating slopes, and PowerPoints from numerous presentations at the National Association of School Psychologists conferences. Further, the website also has videos, answers to commonly asked questions, and a forum to pose questions seeking advice from other school psychologists.

In all, these resources provide practitioners with a wide range of tools to appropriately monitor student progress and interpret ROI.

REFERENCES

Anderson, D., Alonzo, J., & Tindal, G. (2014). *Technical manual: easyCBM*. Eugene, OR: Behavioral Research and Teaching.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology*, *51*, 1–18. doi:10.1016/j.jsp.2012.09.004

Barton, E. E., & Reichow, B. (2012). Guidelines for graphing data with Microsoft Office 2007, Office 2010, and Office for Mac 2008 and 2011. *Journal of Early Intervention*, *34*, 129–150. doi:10.1177/1053815112456601

Burns, M. K., Scholin, S. E., Kosciolek, S., & Livingston, J. (2010). Reliability of decision-making frameworks for response to intervention for reading. *Journal of Psychoeducational Assessment*, *28*, 102–114. doi:10.1177/0734282909342374

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: estimating standard error of the slope to construct confidence intervals. *School Psychology Review*, *35*, 128–133.

Christ, T. J., Zopluoglu, C., Monaghen, B. D., & Van Norma, E. R. (2013). Curriculum-based measurement of oral reading: multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology*, *51*, 19–57. doi:10.1016/j.jsp.2012.11.001

Cole, D. M., & Witts, B. N. (2015). Formative graphing with a Microsoft Excel 2013 template. *Behavior Analysis: Research and Practice*, *15*, 171–186. doi:10.1037/bar0000021

DeWitte, K., Haelermans, C., & Rogge, N. (2015). The effectiveness of a computer-assisted math learning program. *Journal of Computer Assisted Learning*, *31*, 314–329. doi:10.1111/jcal.12090

Ditkowsky, B. (2009). *Progress monitoring with confidence*. Retrieved from http://www.measuredeffects.com

Dixon, M. R., Jackson, J. W., Small, S. L., Horner-King, M. J., Lik, N. M. K., Garcia, Y., & Rosales, R. (2009). Creating single-subject design graphs in Microsoft Excel 2007. *Journal of Applied Behavior Analysis*, *42*, 277–293. doi:10.1901/jaba.2009.42-277

Dynamic Measurement Group. (2008). *DIBELS* (6th ed.). Eugene, OR: Author.

Dynamic Measurement Group. (2012). *Progress monitoring with DIBELS next*. Eugene, OR: Author.

FastBridge Learning. (2017). *FAST progress monitoring tools*. Retrieved from https://www.fastbridge.org

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of learning disabilities: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities*, *38*, 98–108. doi:10.1177/00222194050380020101

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice*, *22*, 129–136. doi:10.1111/j.1540-5826.2007.00237.x

Fuchs, D., Fuchs, L. S., McMaster, K. N., & Al Otaiba, S. (2003). Identifying children at risk for reading failure. Curriculum-based measurement and dual discrepancy approach. In H. L. Swanson, & K. R. Harris, (Eds.), *Handbook of learning disabilities* (pp. 431–449). New York, NY: Guilford Press.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, *18*, 157–171. doi:10.1111/1540-5826.00072

Fuchs, L. S. (2003). Assessing intervention responsiveness: conceptual and technical issues. *Learning Disabilities Research & Practice*, *18*, 172–186. doi:10.1111/1540-5826.00073

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: how much growth can we expect? *School Psychology Review*, *22*, 27–48.

Gottlieb, J., Alter, M., Gottlieb, B., & Wishner, J. (1994). Special education in urban America: it's not justifiable for many. *Journal of Special Education*, *27*, 453–465.

Gresham, F. M., MacMillan, D. L., & Bocian, K. M. (1996). Learning disabilities, low achievement, and mild mental retardation: more alike than different? *Journal of Learning Disabilities*, *29*, 570–581.

Hale, J. B., Alfonso, V., Berninger, B., Bracken, B., Christo, C., Clark, E., … & Yalof, J. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: an expert white paper consensus. *Learning Disability Quarterly*, *33*, 223–236.

Haupt, H., Lösel, F., & Stemmler, M. (2013). Quantile regression analysis and other alternatives to ordinary least-squares regression: a methodological comparison on corporal punishment. *Methodology*, *10*, 81–91. doi:10.1027/1614-2241/a000077

Hixson, M. D., Christ, T. J., & Bruni, T. (2014). Best practices in the analysis of progress monitoring data and decision making. In P. L. Harrison & A. Thomas, (Eds.), *Best practices in school psychology: Foundations* (pp. 343–354). Bethesda, MD: National Association of School Psychologists.

Jenkins, J. R., Graff, J. J., & Miglioretti, D. L. (2009). Estimating reading growth using intermittent CBM progress monitoring. *Exceptional Children*, *75*, 151–163.

Kovaleski, J. F., VanDerHeyden, A., & Shapiro, E. S. (2013). *The RTI approach to evaluating learning disabilities*. New York, NY: Guilford Press.

Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: a comprehensive review. *School Psychology Quarterly*, *30*, 457–469. doi:10.1037/spq0000109

Mercer, S. H., & Keller-Margulis, M. A. (2015). Consistency and magnitude of differences in reading curriculum-based measurement slopes in benchmark versus strategic monitoring. *Psychology in the Schools*, *52*, 316–324. doi:10.1002/pits.21823

Northwest Evaluation Association. (2012). *Technical manual for use with measures of academic performance*. Portland, OR: Author.

Parker, D. C., Dickey, B. N., Burns, M. K., & McMaster, K. L. (2012). An application of brief experimental analysis with early reading. *Journal of Behavior Education*, *21*, 329–349. doi:10.1007/s10864-012-9151-3

Pearson. (2012). *AIMSweb ROI growth norms guide*. Bloomington, MN: Author.

Peterson, K. M. H., & Shinn, M. R. (2002). Severe discrepancy models: which best explains school identification practices for learning disabilities? *School Psychology Review*, *31*, 459–476.

Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). Hoboken, NJ: Wiley.

Renaissance Learning. (2012). *STAR reading technical manual*. Wisconsin Rapids, WI: Author.

Stecker, P. M. (2009). *Technical report: Technical adequacy, instructional utility, and efficacy of McGraw-Hill's yearly progress pro*. Monterey, CA: CTB/McGraw-Hill.

Sternberg, R. J., & Grigorenko, E. L. (2002). Difference scores in the identification of children with learning disabilities: it's time to use a different method. *Journal of School Psychology*, *40*, 65–83. doi:10.1016/S0022-4405(01)00094-2

Thornbald, S. C., & Christ, T. J. (2014). Curriculum-based measurement of reading: is 6 weeks of daily progress monitoring enough? *School Psychology Review*, *43*, 19–29.

Van Norman, E. R., & Christ, T. J. (2016). How accurate are interpretations of curriculum-based measurement progress-monitoring data? Visual analysis versus decision rules. *Journal of School Psychology*, *58*, 41–55. doi:10.1016/j.jsp.2016.07.003

Van Norman, E. R., Nelson, P. M., Shin, J., & Christ, T. J. (2013). An evaluation of the effects of graphic aids in improving decision accuracy in a continuous treatment design. *Journal of Behavioral Education*, *22*, 283–301. doi:10.1007/s10864-013-9176-2

AUTHOR BIOGRAPHICAL STATEMENTS

Timothy J. Runge, PhD, NCSP, is a Professor of Educational and School Psychology at Indiana University of Pennsylvania. His professional interests include assessment practices within multitiered systems of support.

Caitlin F. Bennyhoff, DEd, NCSP, is a school psychologist at Eastern Lancaster County School District. Her professional interests include graphing, calculating, and interpreting rate of improvement; data analysis teams; and system-level change.

Matthew R. Ferchalk, DEd, NCSP, is a school psychologist at Eastern Lancaster County School District in Pennsylvania. His professional interests include behavior consultation, curriculum-based measurement, and data analysis teaming.

Andrew E. McCrea, PhD, NCSP, was recently appointed as a Program Supervisor at the Capital Area Intermediate Unit in Harrisburg, Pennsylvania. His professional interests include measurement, program development, and organizational change.