Data Management, Data Mining, and Data Utilization
with Curriculum-Based Measurement Systems

Gerald Tindal and Julie Alonzo
Behavioral Research and Teaching (BRT) – College of Education, UO

Perspective

Research on data utilization with curriculum-based measurement (CBM) began almost concurrent with the development of the measures used to inform the decisions. This focus was conceptually the purpose of the measurement system (to formatively evaluate instructional programs) as well as the key element in a validation argument. However, in the 35 years since this program of research began on CBM, a number of issues have developed that require attention. In this paper (and chapter), three issues are addressed that are critical in further validation efforts. Data management refers to the process of collecting and rendering data so they are accurate; data mining is an analytic process of investigating and documenting patterns in the data; data utilization involves teacher decision-making (primarily in reference to norms, standards, and individuals). With high-speed computers and networks, all three events should be possible to implement in real time. The presentation ends with a view of the future that involves integration of data file development and autocatalytic processes for validation through iterative use in virtually real time.

Data Management

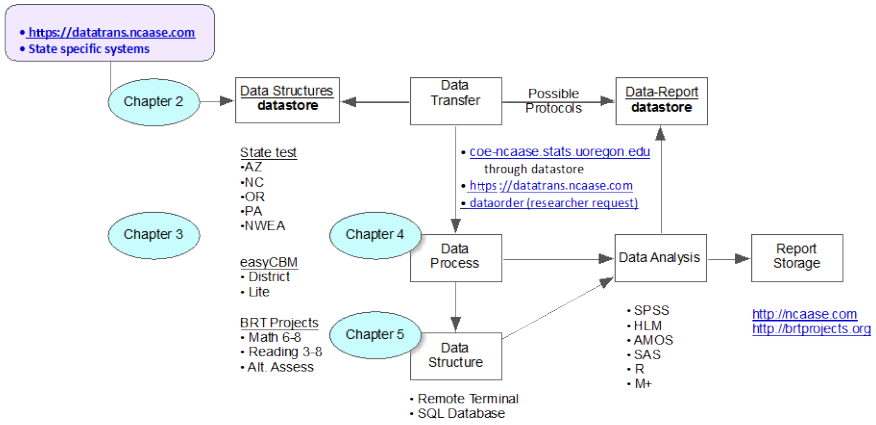
Building Large Data Sets

One of the problems with growth modeling has been the establishment of records that maintain consistency over time in the values being referenced by variables. For example, in the research conducted at the National Center on Assessment and Accountability for Special Education housed in Behavioral Research and Teaching (BRT) at the University of Oregon, annual state test data were used to build cohorts and in the process, a number of issues had to be resolved. For example, students with disabilities were not consistently classified into the same category over the cohort, a challenge similar to that reported by Puranik, Petscher, Al Otaiba, Catts, and Lonigan (2008); these latter researchers eventually used a hierarchical piecewise growth curve model (PGCM) for students whose speech impairment from grades one through three was persistent, was resolved, or became entangled in a lack of achievement and therefore warranted a re-labeling of the student as learning disabled. In a similar manner, students with disabilities have not always been identified consistently over an interval, with three possibilities present: (a) in special education at the beginning of the interval, (b) always in special education throughout the interval, and (c) in special education only at some points during the interval.

To resolve this issue, prior coding schemes need to be digitally embedded in the responses being recorded so data can be sliced at any time point and aggregated meaningfully at all possible points during the interval. Although large-scale state tests are unlikely to be managed more than annually, other systems such as one of the formative-interim assessments also being studied in NCAASE, easyCBM, have rosters structured so that districts can follow a consistent coding scheme while maintaining their own individualized one.

Traversing Institutional Boundaries While Maintaining Security

A critical issue in any large data set is the establishment of security and access in transfer and storage. The most significant issue is moving a file from a state department to a server with multiple access to researchers, all of whom are working with different software to render and analyze the data. For big data development, information needs to be fluidly transferred so that access is available quickly and efficiently for a number of researchers at different locations. As the following diagram illustrates, a complex network of servers and software packages was created for the NCAASE researchers as part of the prerequisite early development work required for data analysis to proceed (see http://ncaase.com).
The protocols for use of this technology infrastructure are addressed in a Data Management Manual but the actual components and functionality of this system are described below.



- For the team of researchers, the center supports four VMware instances set up by the UO information technology (IT) services and configured by the lead BRT technologist. Two instances provide data processing software and capacity to support educational scientists who are analyzing student test data. The virtualized nature of these machines allows resources to be allocated in a just-in-time strategy offering efficient utilization. An additional two servers deploy in-house built web applications that serve smaller populations. These machines are also managed for resource efficiency, security, and uptime.
- NCAASE and BRT have a collection of physical servers primarily used for backup and consisting of a combined total of 16 spindles and 22TB native storage. These machines currently hold around 9TB of actual data including video of best teaching practices, presentations, and student assessment performance. Additionally weekly .sql.gz archives from the two easyCBM systems (see below) and an offsite backup of the primary BRT file server is housed on these machines.
- NCAASE and BRT have a fileserver for the workgroup, consisting of 6 total spindles and configured for disk redundancy. In addition to the automated offsite HEDCO server room backup, complete copies are written to magnetic disk and housed offsite and offline quarterly. University of Oregon campus physical services modified a closet for secure storage and computer-specific power and cooling to support this server. There are just over 5TB of files on this device.
- Data collection includes a link to a well-visited website: http://easycbm.com that supports the easyCBM Lite application. There are approximately 250,000 teacher accounts located across the US with limited additional international accounts, and on any given weekday, around 500 new users sign-up to administer student tests (each "test" results in the neighborhood of

approximately 250 requests sent back and forth from the user to the server). This system is currently served off a dedicated machine at a commercial provider (Rackspace) with support that includes: (a) an operating system managed for security and minor software revisions, (b) managed backup and file recovery, (c) five-minute interval application monitoring with NOC response. The monitoring probe is designed to test the entire LAMP stack.

- NCAASE and BRT have established accounts with commercial providers Linode and Amazon AWS for 'as needed' temporary cloud computing instances. This system allows experimental coding (a common location not connected to the primary architecture) and evaluation and testing of new software. The center and BRT also support one long-standing VPS with Linode for collection of resource monitoring data from the other computing systems. Additional UP/DOWN monitoring and notification is provided by the commercial provider pingdom.

Cleaning Data

Rather than import data directly into statistical software, it often is optimal to check values for errors using structured query language (SQL). This programming language provides fast and intuitive syntax for both structuring the queries in determining values as well as recoding values for eventual use in analysis. In the example below, an initial query of several hundred thousand records revealed the following problems with student gender.

```
SELECT COUNT( * ) AS `Rows` , `gender`
FROM `all_measures_1112_gr1r_WIDE`
GROUP BY `gender`
ORDER BY `gender`
LIMIT 0 , 30
```

Count	Values
4,381	
1	5
11,803	F
1	femaile
1,205	Female
12,717	M
1,201	Male

Using Excel and a series of formulae, these inconsistencies can be resolved in a very efficient manner by placing the initial values in column A, the new recode values in column B, the following formula in column C (which is then filled down).

```
="UPDATE all_measures_1112_gr3rev SET gender_recode = '" & B2 & "' WHERE gender = '" & A2 & "';"
```

FR	TO	SQL Statement
		UPDATE all_measures_1112_gr5rev SET gender_recode = '' WHERE gender = ''; UPDATE all_measures_1112_gr5rev SET gender_recode = '' WHERE gender = '5';
5		
F	Female	UPDATE all_measures_1112_gr5rev SET gender_recode = 'Female' WHERE gender = 'F';
femaile	Female	UPDATE all_measures_1112_gr5rev SET gender_recode = 'Female' WHERE gender = 'femaile';
Female	Female	UPDATE all_measures_1112_gr5rev SET gender_recode = 'Female' WHERE gender = 'Female';
M	Male	UPDATE all_measures_1112_gr5rev SET gender_recode = 'Male' WHERE gender = 'M';
Male	Male	UPDATE all_measures_1112_gr5rev SET gender_recode = 'Male' WHERE gender = 'Male';

The result is a clean recode of gender in a matter of minutes.

Rows	Gender
4,382	
13,009	Female
13,918	Male

The efficiency of this process can be best appreciated when the number of possible values becomes excessive, as was the case in a data file with the variable *ethnicity*.

1-10, 20, 30, 40,99, 639	American Indian/Alaska NativeWhite	Black/African AmericanWhite	MR	Non-Hispanic
A	AmerIndian/Alaskan Native	Caucasian or White	MU	Non-US Native American
A - Asian	Amr Indian/Alska Nat	D, E, H, H - Hispanic/Latino	Multi	Not Hispanic / Latino
A-Asian	AS	H-Hispanic	Multi Race - Two or More Races	Not Specified
African American	Asian	Hawaiian/Pacific Islander	Multi-Racial	NULL
Amer Indian/Alaska Native	Asian or Pacific Isl	HI	Multiple	O, P
AMER.INDIAN/ALASKA NATIVE	AsianWhite	Hisp	Multiple Races	P-Nat Hawaiian/Pac Islander
American Indian / Alaska Native	B, B - Black/African American	Hispanic	Multiracial	PI
American Indian or Alaska	B-Black/Not Hispanic	Hispanic - Any Race	N	T
American Indian or Alaska Native	BL	Hispanic / Latino	Nat American	Two or More
American Indian or Alaskan Native	Black	Hispanic Ethnicity	Native Alaskan/American Indian	Two or More Races
American Indian/Alaska Native	Black or African American	Hispanic or Latino	Native Hawaiian or Other	Unknown
	Black, non-Hispanic	HP	Native Hawaiian	W
	Black/African American	I	Pacific Islander	W - White
		I-American Indian	Native Hawaiian/Other	W-White
		IN	Pacific Islander	WH
		M	Native Hawaiian/Pacific Islander	White
		M - Multiracial		White or Caucasian
		M-Multi Race	No	White, non-Hispanic
				Y
				Yes

Rows	ethnicity_recode
13,880	
321	American Indian or Alaskan Native
473	Asian
1,054	Black or African American
4,077	Hispanic or Latino
698	Multi
39	Native Hawaiian or Other Pacific Islander
1,615	Not Hispanic
9,152	White

The last step in cleaning data involves values for various dependent variables. In SPSS, this can be done efficiently using the following syntax to document minimum and maximum score values for oral reading fluency.

```
DESCRIPTIVES VARIABLES=score1 score2 score3 score4 score5 score6 score7 score8 score9 score10 score11
score12 score13 score14 score15 score16 score17 score18 score19 score20 score21 score22 score23 score24 score25
score26 score27 score28 score29 score30 score31 score=32 score33
  /STATISTICS=MEAN STDDEV MIN MAX.
```

For example, in a data file used in a later section of data mining, we discovered the following values for oral reading fluency: (a) grade 3 values of -6 in score 1, -108 in score 3, -84 in score 3, 11107 in score 7, 741 in score 10, (b) grade 4 values of 853 in score 5, (c) grade 5 values of -447 in score 6, and (d) grade 7 values of 708. Needless to say, none of these values were viable scores, and if their presence had not been detected during the data cleaning process, the integrity of further analysis would have been jeopardized.

Creating Individual Records and Files of Data Through SQL

In building files with the easyCBM data, we began with files downloaded from the website with multiple records per student. Every record is flagged with district and student demographic fields: district, district_studentid, internal_student_id, state, building_name, student_grade, gender, gender_recode, disability, disability_recode, ethnicity, ethnicity_recode, ell, ell_recode, district_data_1, district_data_2, district_data_3, district_data_4, district_data_5. In addition to these district and student demographic fields, a number of measurement fields appear for each student:

used_for…….	benchmark or progress monitoring
measure_grade…...	1 to 8
measure_form……	1-13 or 1-17
score……..	integer
date_given…….	month/date/year

Meaning in Missing Data

Missing data may lead to qualifications in interpretations or generalizability, depending upon the reason for data to be missing. Little and Rubin (1987) classify missing data into three types: (a) missing completely at random – MCAR (an unlikely condition and often difficult to verify), (b) missing at random – MAR (a condition that can easily occur and still allow analyses if all of the

data are used with a fully efficient estimation procedure), and (c) non-missing at random – NMAR that cannot be ignored (which may still allow interpretation as long as only a fraction of the information is missing). MCAR is the most restrictive with the assumption of no relation between observed and unobserved values. With MAR, missing values may be related to the observed values of other variables in the data. And finally, with NMAR, a relation exists between the values that are present and those that are missing.

Three methods can used to account for missing data (Byrne, 2010). List wise deletion of data can be used to resolve the issue of missing data. In this process, records with any missing values for any of the variables in the data set are removed for any computations; the net effect is that the sample includes only records with complete data (although the effect can be a significant loss of data from a reduced sample size and consequently a decrease in statistical power). Another strategy, therefore, to avoid this problem is pairwise deletion in which case records are retained for all variables being analyzed; the net effect is that the sample size varies as a function of the particular analysis. Finally, imputation procedures can be used to replace unobserved values with an estimated value (e.g., the mean can be used as the substitute of the missing values; a regression analysis can be used to postulate missing values from complete data, or pattern matching can be used with cases having similar data patterns to estimate the missing values). Note that in hierarchical linear modeling, missing outcome scores do not prevent the study of growth as long as all participants have at least one data value and the data are missing at random.

Attending to Data Files or Data Foils

As researchers with NCAASE have acquired data sets that are in various states of completeness (considering fields and missing data) and the years represented to be longitudinal, a number of issues have gone into rendering them to be clean and accurate. Examples include:

- General difficulty of receiving full demographic information (some data sets from school districts are limited in the completeness of information related to status of special education eligibility or demographics)
- Licensure of software (costs and limitations) and sheer type of software and programs that researchers not only need to have but also be fluent in using (SQL, HLM, SPSS, AMOS, M+, R, SAS, Excel)
- Mapping data into decisions (e.g., changes in using AYP for participation and AYP for performance)
- Rules for rendering data (see state code books and Data Management Manual)
- File naming, assignment, and transfer (storage) among multiple researchers (and need for rules of the road like ‘sharp curve ahead’)
- Coding of variables and consistency and completeness of data within an SEA and across an SEA to an LEA (e.g., best score within a year for Oregon state test data, change of disability across years)
- Findings that may either require conditional interpretations or stop the show (e.g., edition effects in a state, changing standards in another state for 09-10 math and 10-11 reading scores, and different tests in different years in yet a third state).
- Difficulty converting test score files intended for annual accountability into longitudinal files (multiple instances of same id’s used for more than one student)
- Unique data quality issues for special education students and exacerbation of common data quality issues (e.g., changes between alternate and general assessment across years, grade

retention). The smaller *n* for special education population means that all of these problems will have greater impact on robustness of results and confidence in inferences made.
- Data quality issues such as discrepant disability classification for a student depending on source of information (e.g., using disability classification from annual test file versus from data file used for federal reporting).
- Mystery data appear (e.g., disability classification missing from largest district in a state for one year only with no explanation for missing data, required seeking new data sources and more file processing to get key variable).

If this list of issues cannot be resolved *a priori*, then a big data system designed to mine data later is likely to be fraught with problems. The key to resolving these data management problems is to build an interface into a web site so that upload and download procedures address them automatically.

Data Mining – Measurement Sufficiency

Thirty years of research on curriculum-based measurement (CBM) have documented its technical adequacy and potential formative instructional benefits for students with disabilities (Deno, Lembke, & Reschly, 2003; Fuchs, Deno, & Mirkin, 1982; Jenkins, Deno, & Mirkin, 1979; Tindal, 2013). A substantial evidence base exists to support the appropriate use of CBM for improving the instruction and educational outcomes of students with disabilities by providing general and special educators meaningful information about the progress students are making in mastering a year's worth of curriculum. To facilitate the research-to-practice translation of this evidence-based practice, researchers have developed technology-based platforms to support teachers with the administration and scoring of CBM.

Unfortunately, researchers have noted that the actual practice-in-use of CBM is undisciplined with insufficient data on the intersection of measurement sufficiency, instructional integrity, and data-based decision-making (Tindal, Nese, Saez, & ALonzo, 2012, February). Despite widespread use of technology-based CBM, many teachers lack the resources and knowledge to effectively implement CBM and systematically use the collected data for selecting appropriate goals, analyzing student skills, and managing data (Tindal, 2013). This practice-in-use problem significantly undermines the effectiveness of CBM to improve instruction and outcomes of students with disabilities (SWD).

In this next section, we analyze data from easyCBM, a specific form of curriculum-based measurement developed at the University of Oregon (UO), Behavioral Research and Teaching (BRT). The existing user base of easyCBM includes nearly 1.5 million students in easyCBM, five million measures by January 2012, with a total of nearly 250,000 registered educator users on the free teacher site. Teachers access easyCBM by going to the website (http://easyCBM.com). With the district version, approximately 75,000 registered users tested over 1.1 million students (with over 8 million tests administered).

In either site, a number of features are available for teachers to use in their instructional intervention program: A student tab is used for teachers to create groups. The system provides teachers multiple CBM measures (the district version provided both benchmark (fall, winter, and spring forms) and progress monitoring measures (13-17 forms), while the Lite Edition provided

9 progress monitoring measures per measure type per grade). The measures tab is selected to access these forms. Teachers also have the option to receive training on all aspects of administration of measures and interpretation of results, and both systems also provide reports of student performance.

Timing in Measurement Occasions

When teachers measure progress, the timing (when in the year and how often) is critical in establishing an appropriate data base from which to make decisions. If teachers begin measurement too late in the year or measure too infrequently, it is likely to be difficult to use the data appropriately for evaluating instructional outcomes with time enough to adjust them. Yet, as can be seen below, teachers often began progress measurement late in the year.

First Month of Measurement for Grade Three Students

Month	*Count*	*Percent*
9	354	15.2
10	994	42.6
11	376	16.1
12	104	4.5
1	202	8.7
2	123	5.3
3	127	5.4
4	42	1.8
5	10	.4
Total	2332	100.0

First Month of Measurement for Grade Four Students

Month	*Count*	*Percent*
9	381	14.6
10	1048	40.3
11	525	20.2
12	129	5.0
1	163	6.3
2	107	4.1
3	174	6.7
4	63	2.4
5	11	.4
Total	2601	100.0

First Month of Measurement for Grade Five Students

Month	*Count*	*Percent*
9	257	13.5
10	797	41.7
11	346	18.1
12	155	8.1
1	87	4.6
2	133	7.0
3	83	4.3
4	38	2.0
5	13	.7
Total	1909	100.0

Level of Measurement

One of the problems with progress measurement is establishing appropriate (sensitive) long-range goal measures. Many students who are just learning to read are unlikely to show progress if they are measured in grade-level material. For example, a third grade student may need passages from grades one or two so that the difficulty of the passage can reflect fluency, an important construct in learning to read and comprehend (Wang, Algozzine, Ma, & Porfeli, 2011). Yet, as can be seen in the sample tables below, teachers in grades 3-5 are measuring in a number of grade levels.

Grade Level of Measures used with Grade Three students on the First Occasion

	Count	Percent	Valid %
Grade 1	128	5.5	5.5
Grade 2	200	8.6	8.6
Grade 3	1979	84.9	84.9
Grade 4	22	.9	.9
Grade 5	3	.1	.1
Total	2332	100.0	100.0

Grade Level of Measures used with Grade Four students on the First Occasion

	Count	Percent	Valid %
Grade 1	21	.8	.8
Grade 2	149	5.7	5.7
Grade 3	163	6.3	6.3
Grade 4	2232	85.8	85.8
Grade 5	36	1.4	1.4
Total	2601	100.0	100.0

Grade Level of Measures used with Grade Five students on the First Occasion

	Count	Percent	Valid %
Grade 1	11	.6	.6
Grade 2	57	3.0	3.0
Grade 3	111	5.8	5.8
Grade 4	122	6.4	6.4
Grade 5	1599	83.8	83.8
Grade 6	9	.5	.5
Total	1909	100.0	100.0

Concurrent Measurement Providing Convergent and Divergent Criterion Validity

As teachers make decisions, they need to be sure that the measures fit together in both a convergent and divergent manner. Some measures should fit together while other measures should not. For example, as a student becomes proficient in early literacy skills, the skills measured by tests of letter names, letter sounds, and phoneme segmentation may begin to hit a ceiling and the more sensitive measures become word reading and passage reading fluency. Note: This table displays the number of records, which is not the same as the number of students; rather it is the number of times a benchmark or progress monitoring measure was entered into the system.

Complete Data Collection System

File	**No. Records**
Kindergarten	225,659
Grade 1	318,547
Grade 2	402,726
Grade 3	379,524
Grade 4	381,360
Grade 5	359,512
Grade 6	276,076
Grade 7	218,683
Grade 8	200,103
TOTAL	**2,762,190**

Yet, as can be seen (indirectly) with the number of different measures being given concurrently, the number of records varies considerably within grades and across measures as well as across successive grades. In the table below, only grades Kindergarten and 8 are displayed. In fact, some measures are being administered from out-of-grade-level (as indicated by the values that are bold and italicized). Of course, the value of this practice needs to be ascertained (e.g., a teacher may be using an out-of-grade-level measure appropriately, to verify the presence or lack of skill in relation to another measure).

Complete Measurement System

GRADE K		GRADE 8	
cbm_type	**COUNT**	**cbm_type**	**COUNT**
ln	27,012	***ln***	***16***
ls	58,354	***ls***	***2***
math	39,069	math	45,084
math_alg	***1***	math_alg	6,702
math_danoa	***15***	math_danoa	5,979
math_geo	4,987	***math_geo***	***180***
math_msmt	3,619	math_geomsmt	3,723
math_noag	***4***	***math_gma***	***246***
math_numop	8,591	***math_mda***	***143***
math_numopalg	***260***	***math_mga***	***450***
mcrc	***20***	***math_msmt***	***57***
prf	***62***	***math_noag***	***637***
ps	50,279	math_numop	3,019
vocab	***37***	math_numopalg	1,969
wrf	33,349	***math_numoprat***	***205***
TOTAL	225,659	mcrc	52,609
		prf	31,862
		ps	***4***
		vocab	46,997
		wrf	***219***
		TOTAL	200,103

Data Utilization

In any assessment system, three types of references can be used in making decisions (Tindal & Marston, 1981): (a) norm-referenced in which performance for a student is compared to (age) appropriate peers; (b) standards-referenced in which a specific domain (usually standards) is used to establish proficiency levels, and (c) individual-referenced with an emphasis on within student progress (change over time).

Norm-Referenced Decisions

With an interim-formative assessment system, norm references are usually made using benchmark levels of performance from three times of the year that correspond to a traditional school year of beginning of the year in the fall, mid-year in the winter, and end-of-year in the spring. With the data being used often for risk screening (at least in the fall), the most appropriate analysis would be based on Risk and Receiver Operating Characteristics (ROC).

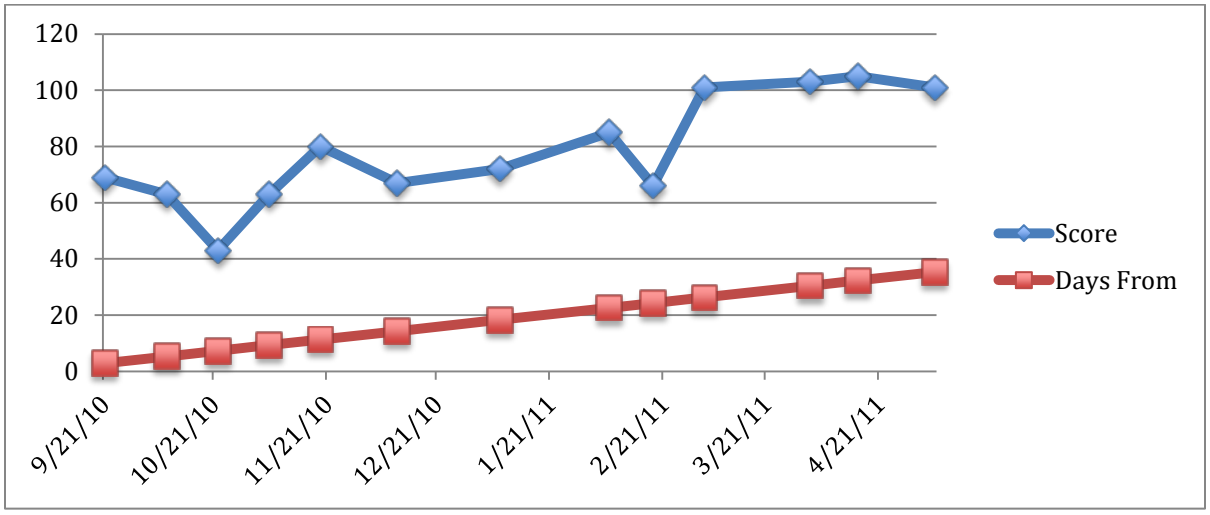
Standards-Based (Criterion-Referenced) Decisions

With the move to individual state standards under No Child Left Behind (2001) and now the common core state standards to be adopted and implemented in the next few years, the reference is standards-based. Rather than comparing a students' performance to others, the standards are used to make interpretations (usually in reference to a pre-defined level of proficiency). With data being used in this manner, the alignment of (test) items to standards is a critical issue; if items are aligned then both the screening (norm-referenced) and mastery-based (standards-referenced) performance can be used to identify students at risk as well as provide appropriate remediation. Two unique considerations in this reference are the degree to which the items represent a range of difficulty (are appropriately scaled) and the degree to which they provide useful information for grouping students, one of the first steps needing to be addressed.

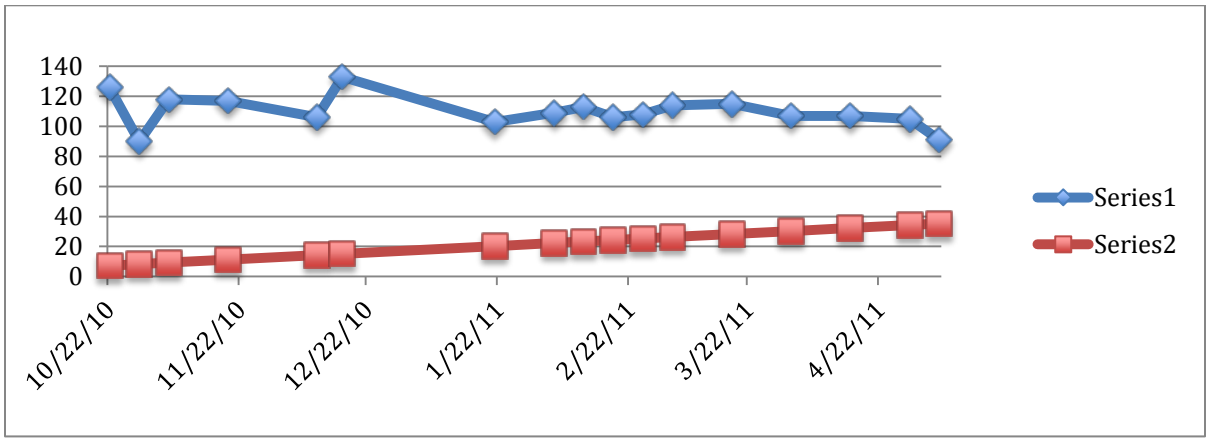
Individually Referenced Decisions

The final reference for making decisions is to use performance from a student over time to determine how much progress has been made. This type of reference is relatively new to educational settings and is the foundation for any curriculum-based measurement (CBM) system. Begun in the late 1970s with *Data-Based Program Modification* (Deno & Mirkin, 1977) and most recently summarized by Tindal (2013), CBM provides a time series display of data for teachers to monitor students' progress and use this information to evaluate instruction. Three critical features of time series data are used to make decisions: (a) slope, or the 'run over rise' in showing the rate of change; (b) variability, or the amount of variation in the trend of change over time; (c) and change in level of performance, as evaluated across each instructional phase. In the graphs below, data are displayed from the 2010-2011 easyCBM oral reading fluency data files. All three graphs can be interpreted with the three indices above.

Improvement in Time Series of Oral Reading Fluency Improvement Over the Year



No Improvement in Time Series of Oral Reading Fluency Improvement Over the Year



The purpose behind formative assessments, however, is to improve instruction. Usually, this is accomplished by using the data noted above along with some decision rules. Generally, two types of rules have been codified over the three decades in which this type of research has been conducted: (a) empirical rules, and (b) goal (aim line) referenced rules. In the empirical-rule based decision-making system, a baseline phase is used to reference performance and progress without an intervention and establish a pattern that can be used to compare post-intervention trajectory (Gast, 2010; Horner et al., 2005). With the goal (aim line) approach to making decisions, a final level of performance is established at the end of the interval (usually the end of the year in the spring) and then changes in instructional programs make reference to this goal (Barnett et al., 2006; Burns, Scholin, Kosciolek, & Livingston, 2010). In either type of decision-making system, interventions are maintained or changed in accordance with visual-analysis guidelines to evaluate data of individuals or small groups (Gast, 2010).

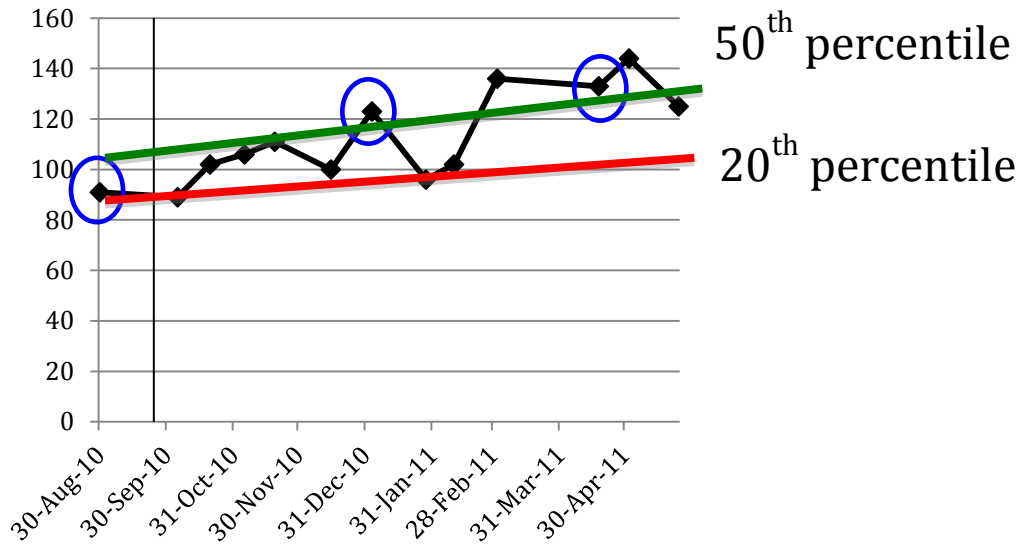
Some researchers have provided explicit guidelines for decision-making, such as monitoring progress weekly or biweekly for 6 weeks, and evaluating the adequacy of progress according to the rule that instructional changes should be made if three consecutive data points fall below the goal line (Deno, Lembke, & Reschly, n.d.) as cited in Jenkins and Terjeson (2011) Recommendations also have included daily measurement (Deno, 1985), at least three times per week (Mirkin, Deno, Tindal, & Kuehnle, 1982), twice per week (Fuchs, Fuchs, & Hamlett, 1989), weekly or biweekly (Deno et al., n.d.), or every 3 or 4 weeks (Jenkins, Graff, & Miglioretti, 2009). Christ (2006) suggests that to account for measurement error, biweekly data need to be collected across 10 weeks. Other researchers defined sufficient progress as three to five consecutive data points above the aim line, and suggested that either a more ambitious goal was needed or intervention termination should be considered, and defined insufficient progress as three to five consecutive data points below the aim line, and suggested that the intervention needed to be modified (Burns et al., 2010; Jenkins & Terjeson, 2011; Mirkin et al., 1982). Jenkins et al. (2009) compared growth slopes based on measurements taken every 1-4 and 9 weeks and found that the frequency of progress monitoring could be significantly reduced without detracting from the validity of growth estimates. One study included decision rules that the number of data points needed to make a reasonably valid estimate of a student's progress was closer to 20 data points collected across 2.5-3 months, and for a decision regarding a student's

eligibility for special education, a total of 40 data points collected across 5-6 months (Ardoin, 2004)

This next section with six example graphs is based on a presentation at the *Pacific Coast Research Conference Data-based Decision Making: Practice to Research* (Alonzo, 2012, February). The graphs, pulled from actual easyCBM 2010-2011 assessment and intervention data, illustrate some of the patterns we see in teacher use (and mis-use) of CBM data. They are provided here to highlight some of the issues that must be addressed if CBM performance data are to be interpreted meaningfully.

Example 1 illustrates the potential for mis-allocation of resources when an intervention is provided based on performance on a Fall Benchmark assessment, the intervention apparently has a positive effect, resulting in the student nearing and then exceeding grade-level expectations on the measure being used for progress monitoring, but no change noted in the intervention even when the student's performance indicates that no additional intervention is warranted.

Example 1 – One Intervention Followed by Increasing Growth and No Change to Intervention

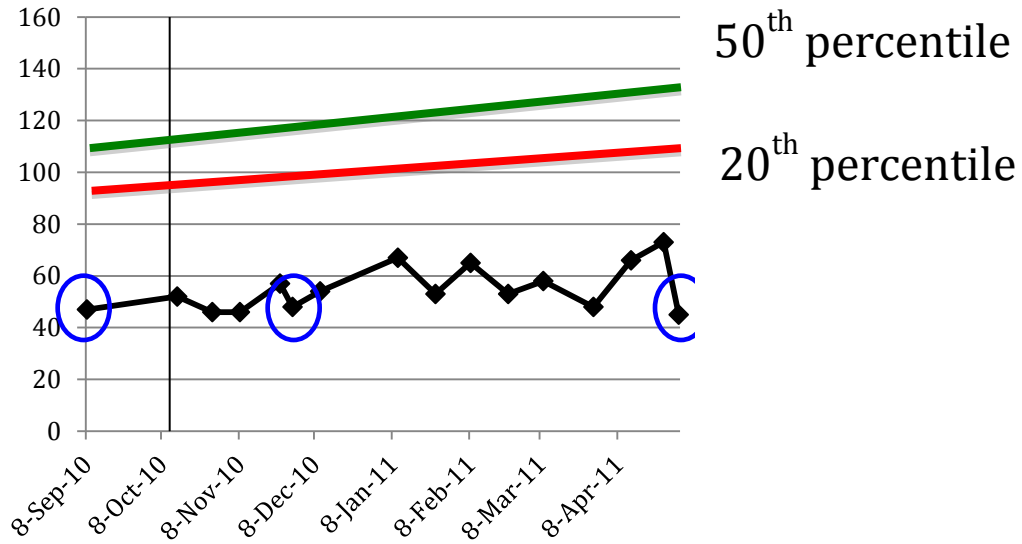


Note. Benchmark assessments are circled. Vertical line notes date of intervention.

Intervention 1	Rewards 2x/week for 20_x000D_ Strategic and Intensive kits for comprehension and vocabulary 2 x/week for 20 min_x000D_ Read Naturally 20 min 3x/week_x000D_

As in the first example, Example 2 shows a single intervention being provided following low performance on the Fall Benchmark assessment. Unlike the earlier example, however, here we see a student for whom there is no evidence of significant growth. What Example 2 illustrates most clearly is a failure for student performance data to prompt a change in instructional approach on the part of the teacher. Sadly, this example provides evidence of squandered resources, with progress measures being administered faithfully throughout the year, no improvement in student performance, yet no documented change in instruction.

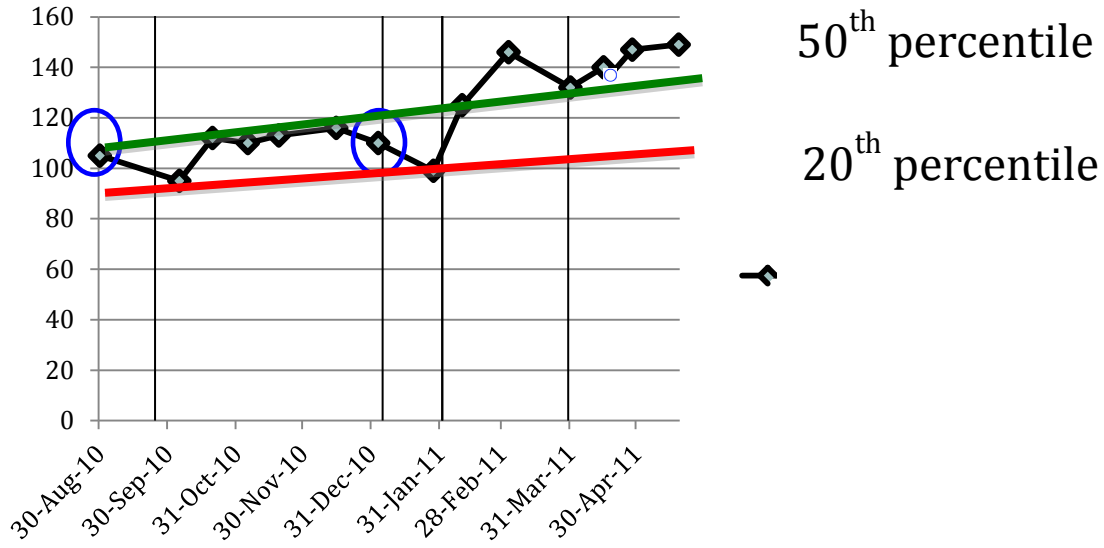
Example 2 – One Intervention Followed by Flat Growth and No Further Interventions



Intervention 1	Tier 3b (BEST): Skill focus: Phonics & Fluency. Using 4th grade Triumphs Materials & the Rewards program. 5 students taught by an IA. 30m/day, 4x a week.

In contrast, Example 3 provides evidence of student performance data impacting instruction and changes in instruction subsequently impacting student performance data. This pattern of CBM assessment, followed by instructional change, follow-up assessments showing no improvement, prompting further instructional change, resulting in accelerated learning could be labeled the "Poster Child" for appropriate CBM use.

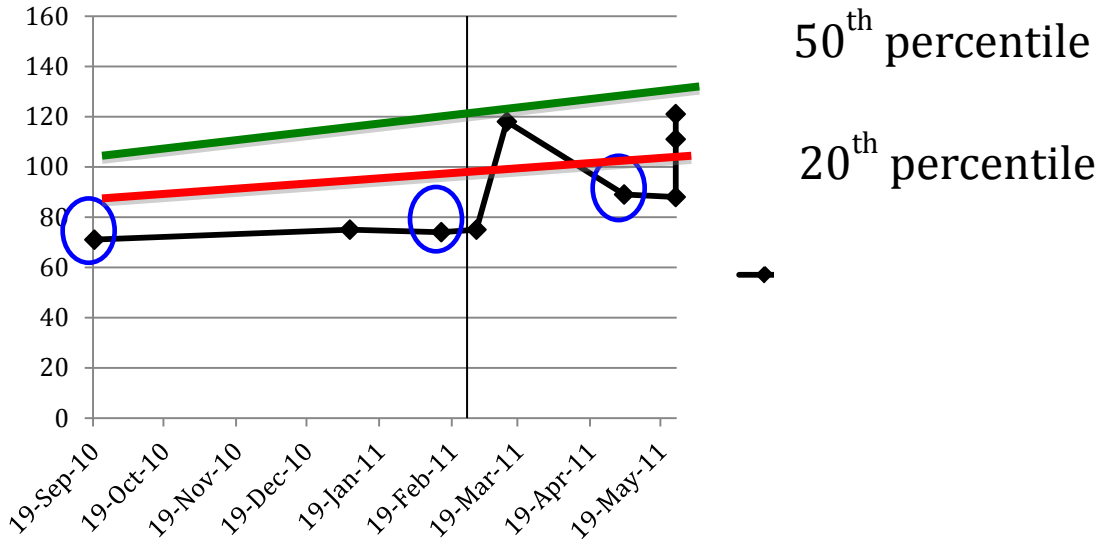
Example 3 – One Intervention Followed by Flat Growth and Additional Interventions Followed by Some Growth



Intervention 1	Rewards 2x/week for 20_x000D_ Strategic and Intensive kits for comprehension and vocabulary 2 x/week for 20 min_x000D_ Read Naturally 20 min 3x/week_x000D_
Intervention 2	Group Change: Moved to ___'s group(smaller size). Doing Harcourt Intensive materials, Read Naturally 2x/week and Study Island 2x/week to practice fluency and comprehension.
Intervention 3	Study Skills: 20 min. of test taking practice and strategies
Intervention 4	Concerns noted with teacher about the lack of comprehension. Decided to wait a few more weeks to see if extra class helped out.

Lest we begin to celebrate the success of CBM to meaningfully impact instruction and consequently student learning, Example 4 returns us to the reality of insufficient measurement sufficiency and lack of instructional relevance (no intervention provided until mid-February despite Fall and Winter Benchmark and a progress monitoring score showing consistent performance below the 20th percentile rank) and inappropriate measurement (notice the three progress monitoring measures administered on the same day at the end of the school year, with a change in performance that might well be attributed to a practice effect rather than any real improvement in the underlying construct being assessed).

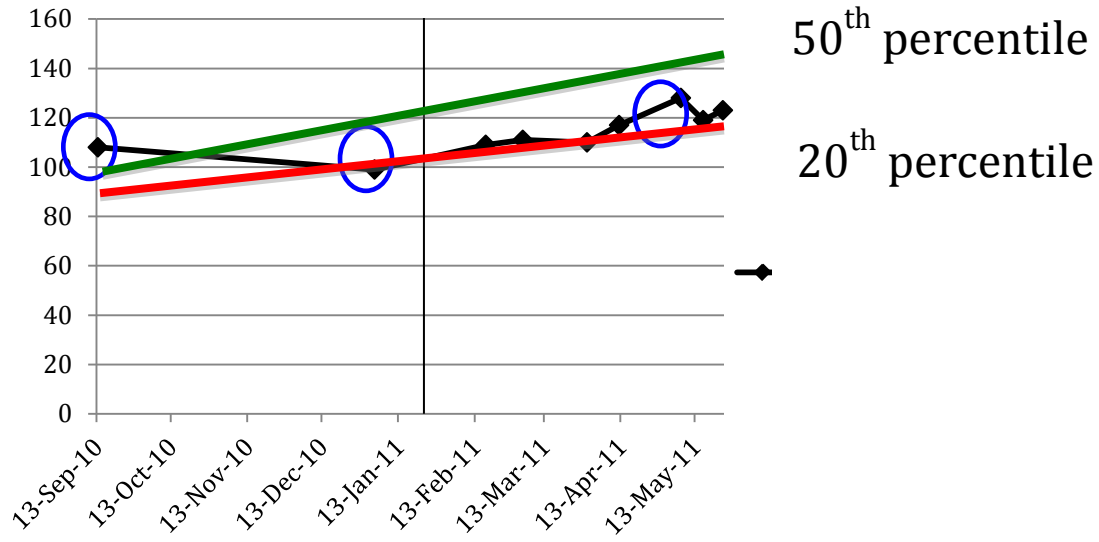
Example 4 – Intervention Inappropriately Late in the Year and Some Growth After Intervention



Intervention 1	Corrective Reading B-2: 40 min.4x/week;group of 6

In contrast, Example 5 offers slightly more hope for the student performance data impacting instructional practice. In this example, the Fall Benchmark score placed the student above the 50th percentile on the measure being administered. As a result, the student was not provided any additional interventions at the start of the year. The mid-year benchmark, though, administered in the Winter, show that the student's performance has now slipped below grade-level expectations, with a score corresponding to the 20th percentile rank. Based on the Winter Benchmark assessment, the student was provided an instructional intervention and progress monitoring measures were administered approximately every two weeks. Visual inspection of these data points suggests slow but steady improvement post-intervention.

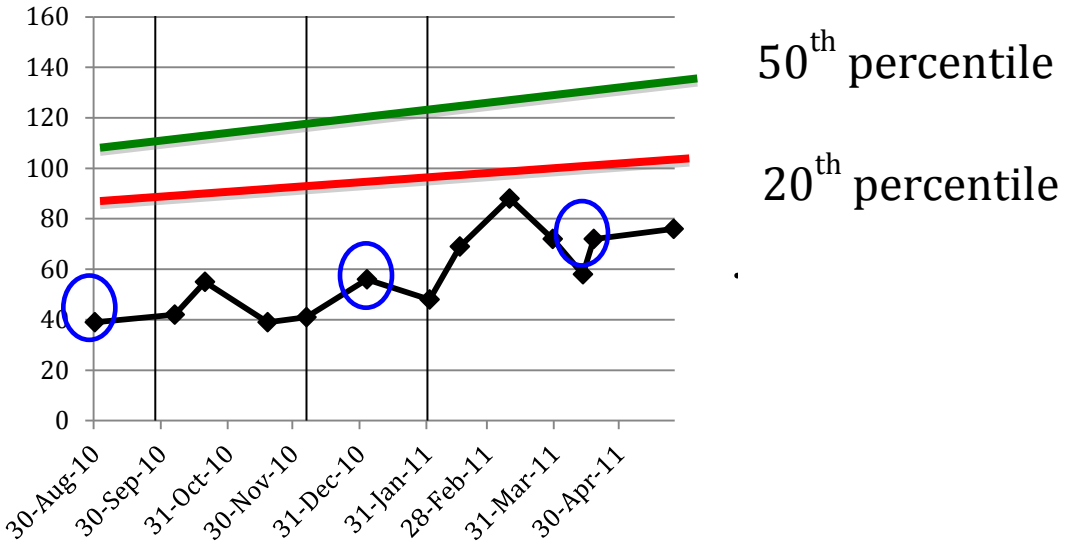
Example 5 – Intervention Appropriately Late (Given High Fall Benchmark Score) and Some Growth After Intervention



Intervention 1	Read Natural: Student goes to read naturally each morning during Tiger Time (30 mins).

Our final example, Example 6, illustrates the opposite extreme from some of the earlier examples. Here, we see a student whose benchmark and progress monitoring scores consistently place him/her well below grade-level expected performance, with scores well below the 20th percentile. Plotted on the graph are three different interventions. The first two interventions appeared to have resulted in little improvement in student performance, but following the third intervention the student's performance began a definite trend upward, such that by the Spring Benchmark assessment this student almost reached the 20th percentile rank. On the surface, this graph appears to show appropriate use of CBM data to guide instructional decision-making. Closer inspection, however, reveals too few data points between interventions to provide robust interpretation of student growth in skill. Of particular note, in between Intervention 2 and Intervention 3, a single data point was collected.

Example 6 – Many Interventions and Not Enough Measures Between Each Intervention



Intervention 1	Read Naturally 20 minute sessions four days a week. Harcourt Strategic comprehension skills 20 minute sessions 2 days a week.
Intervention 2	Phonics for Reading Level 3. 20 min. session four days a week.
Intervention 3	New Skill Block: Moving to __'s Skill Block. Sp. Ed.

These six examples from the easyCBM dataset illustrate some of the insights into actual classroom practice and its resulting impact on student performance data that are possible to pull from a dataset with robust student numbers and from which instructional intervention data are combined with student performance data. Rich datasets such as the one from which these data were pulled, can give researchers a window into the intersection of student and teacher academic behavior.

Moving Forward

Next steps in recursive research and development. The first and most important step in moving to big data systems is to more actively structure the data so they can be moved seamlessly across secure sites for open access to multiple researchers. This step will require the development of data base tools that can be used to bring consistency in variable coding (values), field placement, and record maintenance.

Moving to smart apps: setting occasions and prompted systems. The next step in movement to big data is to develop smart applications that respond within pre-established boundaries to prompt users. As teachers collect formative assessments on students, this prompting can be in the form of suggestions for grade level to measure, when to measure, consistencies in errors, and interpretive guidelines for values. These prompts can come through smart calendars, text messages, or emails.

Differentiated reports. As smarter data collection systems become built, with data management and mining reflecting consistent (and expected) data values, reports need to be developed that allow teachers multiple interpretations for students and incorporate all three references: (a) norm-referenced to determine a student's standing in a group, (b) standards-referenced to identify skill and knowledge strengths and weaknesses, and (c) individual-referenced to show the impact of instructional programs.

Just in time professional development. Although improvements can be made in making data collection more consistent and with better measurement sufficiency, teachers need more professional development in how to interpret data. This kind of professional development needs to be situated to the problems being encountered by individual teachers and students, accessible when it is important, and efficiently delivered as part of the intervention development process.

Integration of issues in moving forward. All of these improvements need to take place in a cloud environment that provides both practitioners and researchers open access and analysis that is, if not in real time, within minutes of any transactions. All three issues of data management, data mining, and data use can then be used concurrently to structure change in an iterative manner.

References

Alonzo, J. (2012, February). *Data-based Decision Making: Practice to Research* Paper presented at the Pacific Coast Research Conference, Coronado, CA.

Ardoin, S. P. (2004). The response in response to intervention: evaluating the utility of assessing maintenance of intervention effects. *Psychology in the Schools, 43,* , 713-725. doi: 10.1002/pits.20181

Barnett, D., Elliott, N., Graden, J., Ihlo, T., Macmann, G., Nantais, M., & Prasse, D. (2006). Technical adequacy for response to intervention practices. *Assessment For Effective Intervention, 32*(1), 20-31.

Burns, M., Scholin, S., Kosciolek, S., & Livingston, J. (2010). Reliability of decision-making frameworks for response to intervention for reading. *Journal of Psychoeducational Assessment, 28*(2), 102-114.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applicatoins, and programming (2nd edition)*. New York: Routledge.

Christ, T. (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*(1), 128-133.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Deno, S. L., Lembke, E., & Reschly, A. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention, 28*(3 & 4), 3-12. doi: 10.1177/073724770302800302

Deno, S. L., Lembke, E., & Reschly, A. (n.d.). Progress monitoring: study group odule. . Minneapolis, MN: University of Minnesota: Department of Special Education, .

Deno, S. L., & Mirkin, P. K. (1977). *Data based program modification: A manual*. Minneapolis, MN: Leadership Training Institute for Special Education.

Fuchs, L., Deno, S., & Mirkin, P. (1982). Direct and frequent measurement and evaluation: Effects on instruction and estimates of student progress *Research Report*. Minneapolis, MN.

Fuchs, L., Fuchs, D., & Hamlett, C. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55*(5), 429-438.

Gast, D. L. (2010). *Single-Subject Research Methodology in Behavioral Sciences*. New York, NY: Routledge.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179.

Jenkins, J., Deno, S., & Mirkin, P. (1979). Measuring pupil progress toward the least restrictive environment *Monograph*. Minneapolis, MN: Institute for Research on Learning Disabilities (IRLD)-University of Minnesota.

Jenkins, J., Graff, J. J., & Miglioretti, D. (2009). Estimating reading growth using intermittent CBM progress monitoring. *Exceptional Children, 75*(2), 151-163.

Jenkins, J., & Terjeson, K. (2011). Monitoring reading growth: Goal setting, measurement frequency, and methods of evaluation. *Learning Disabiities Research & Practice, 26*(1), 28-35.

Little, R., & Rubin, D. (1987). *Statistical analysis with missing data*. New York: John Wiley.

Mirkin, P. K., Deno, S., Tindal, G., & Kuehnle, K. (1982). Frequency of measurement and data utilization as factors in standardized behavioral assessment of academic skill. *Journal of Behavioral Assessment, 4*, 361-370. doi: 10.1007/BF01341230

No Child Left Behind. (2001) *Committee on Education and Labor* (First ed., pp. 1-95). Washington, DC: U. S. Government Printing Office.

Puranik, C. S., Petscher, Y., Al Otaiba, S., Catts, H. W., & Lonigan, C. J. (2008). Development of Oral Reading Fluency in Children With Speech or Language Impairments : A Growth Curve Analysis. *Journal of Learning Disabilities, 41*, 545-560.

Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. *ISRN Education (International Scholarly Research Network), 29*. doi: 10.1155/2013/958530

Tindal, G., & Marston, D. (1981). *Classroom-based assessment: Evaluating instructional outcomes* Columbus, OH: CHarles Merrill.

Tindal, G., Nese, J. F. T., Saez, L., & ALonzo, J. (2012, February). *Validating Progress Monitoring in the Context of RTI*. Paper presented at the Pacific Coast Research Conference, Coronado, CA.

Wang, C., Algozzine, B., Ma, W., & Porfeli, E. (2011). Oral Reading Rates of Second-Grade Students. *Journal of Educational Psychology, 103*(2), 442–454.