Student Growth Models and Principal Evaluation

The Performance Evaluation Reform Act (PERA) was approved by the Illinois General Assembly and signed into law by the Governor of Illinois in January 2010. An important requirement of this act includes school districts incorporating student data and indicators of student growth as a *significant factor* of performance evaluations of principals (including assistant principals) and teachers. What PERA does not specify is… how. Specifically, at this point, how to incorporate student data and indicators of student growth is up to individual districts.

With PERA in mind, the purpose of setting up an evaluation system using student data is to separate the variability in student scores that can be explained by the effectiveness of the school (i.e., teacher effectiveness) versus growth that would occur naturally. In order for such a model to function properly, sufficient psychometric characteristics (i.e., reliability and validity) of the assessments used are a prerequisite. In addition, the system of accountability must be founded on the ability of the system to correctly distinguish between progress and lack thereof. Finally, both intended and unintended consequences of using any model of evaluation (or test) must be carefully considered, ideally prior to implementation of that model (Messick, 1989). According to Roland H. Good, the first commandment of using data for high-stakes, decision-making is "thou shalt not make capricious decisions about children" (Good, Kaminski, Cummings, Powell-Smith, MacConnel, 2009). We would be wise to consider this maxim, when we consider the idea of using student data to make decisions about teachers and principals as well. For a teacher or principal evaluation model there are several elements that must be in place in advance of the implementation of the system. The elements necessary for the successful implementation of such a system require that at a minimum, principals and teachers must recognize (a) what is expected, (b) the success with goals are within their control, (c) the goals are attainable, and (d) they have the resources to accomplish the goals.

It seems so easy to measure growth. For example, when we measure how tall a child is, we use a common metric, centimeters, inches, or a combination of inches and feet. When we measure weight, we use grams, ounces, or pounds. How tall or how much a student weighs follows an expected path (or trajectory). Height, weight and girth are characterized by being equal interval, and relatively easy to measure in a reliable and valid way, and we can see it. Although these three measures of physical growth each measure a different dimension of "growth," it is not difficult to conceptualize how these measures might be combined to show change. This is not necessarily the case with education.

Problems and requirements of models of measuring student growth. Growth in education is different than the physical growth measured in terms of height, weight and girth. Within the context of educational measurement, complex statistical models are run with the intent of being accurate. However, it is questionable whether educational measurements are actually equal interval (Ballou, 2008; Martineau 2006; Reckase, 2005; Yen, 1986). In addition, errors in the validity of educational measurement are not only present; in some cases error may be as large as or larger than gain / loss measuring growth (Krueger & Lindahl, 2001). At the individual student level, we see this when we see how often student performance on

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

a single test indicates a problem, when in reality, this result is due to something other than learning. Because of the degree of error in educational measurement, far more than two tests or two points in time are necessary to accurately model growth. Although there are several complex statistical models for measuring academic growth that already exist and are in use across the country (O'Malley, 2008), each has benefits and limitations. Overall these models are problematic for use in schools for several reasons including:

1. The focus on student growth should improve teaching and learning. Though over time, one might argue that complex statistical models may lead to improvement, at least in the short run, these models are too complex to be examined midway through the process, when principals and teachers can still focus on improvement. The result is a summative, "gotcha" model of evaluation.
   If the focus on student growth does not improve teaching and learning during the year, but serves only to reward or penalize teachers or principals as a summative rating, then the model is flawed in its' primary purpose.

2. Lack of connection to the standards-based focus. The majority of complex statistical models focus on growth that is referenced to factors other than our expectations or standards. When student scores are combined in this way, the detail in outcomes can cloud the answer to the question we are actually concerned with; Are students growing at a level that we have set as an expectation.
   If the success or failure of the model to distinguish between students who are increasingly meeting standards, then the model is flawed. The results of the model should not be a surprise to those being evaluated.

3. Violations of basic assumptions of models that can lead to imprecise interpretations. Traditional (i.e., parametric) statistical models treat growth at any point along the continuum as equal interval data, i.e., "ten points is ten points", regardless of whether the ten points of growth are coming from the top scoring, or the lowest scoring students. As educators attempt to apply statistical models to demonstrate student growth, this assumption has been increasingly questioned (Ballou, 2008; Martineau 2006; Yen, 1986). *If the model is based on assumptions that are not met, then the model is flawed in it's design. At a minimum a model should meet the assumptions on which it is based.*

4. Focus on the details rather than the big ideas. Statisticians often argue about issues that are far removed from the classroom, including several methods of manipulation and examination of measurement error. Unfortunately, a seemingly small decision about how to treat statistical assumptions can result in very different outcomes (Briggs & Weeks, 2009). *If the outcome of the model is based on factors that are so difficult to explain that the factors cannot be questioned, then the model is flawed in its' implementation. The results of the model should not be based on factors that are so far removed from those being evaluated that the person being evaluated does not feel in control of the factors on which they are being evaluated.*

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

5. The overly complex statistical models have not been shown to be substantially more effective than less complex models; however, more complex models are far more difficult to explain to principals, teachers and the community as a whole (Reckase & Martineau, 2004). *The law of parsimony states that all else being equal, the simplest model is the best model.*

Accountability should be reasonable. For a system of accountability to work, it should provide **trustworthy**, **useable** and **accessible** (Carnine, 1997) information to evaluate principals and subsequently teachers as part of the requirements of PERA. We don't want to make high - stakes decisions based on only one type of data. Of course, this is not a problem because we have a lot of academic data demonstrating performance over time for students. We do not, however, have a complete set of academic data for all students across the district, or even within a particular grade level. Some students may be missing one or more scores, for a variety of reasons. Further, the varied measures are not scaled in the same way, meaning that expected growth on one measure may differ quantitatively from growth on another. Methods are available to standardize scores across measures. However, as mentioned above, the details of the standardization and combination process may render the overall meaning of improvement, difficult to explain, if not altogether meaningless. We need to maintain our focus on the question at hand, rather than wait for the exact answer, or in the immortal words of John Tukey (1962), "*Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise*."

A proposed solution. Although it may seem like the information provided thus far is leading to a conclusion of an unworkable system, I believe there is a solution. The solution requires us to return to the question. Specifically, within the context of using student data for teacher and principal evaluation, the question that we are asking is: ***are students growing academically?*** The answer that we are attempting to supply fits into one of four categories, namely: Unsatisfactory, Needs Improvement, Proficient, and Excellent. With this in mind, it does not even make sense to attempt to apply complex statistical models to the question; rather, our model should evaluate academic growth, not in a single dimension, but across dimensions, i.e., multiple measures that cross subject areas. Next, our model should result in a summary statement that is clear, and defensible within the constraint of the four categories of growth.

The solution is based on two premises:

1. We know what success and failure look like.
2. We have meaningful standards in-place.

The support for these premises includes:

- The Illinois Standards Achievement Test provides not only scaled scores, but also useable performance categories.
- The WIDA Consortium issued a document delineating expected rates of growth for English Language Learners, co-varied on the grade and language proficiency level of the student
- Cut Scores for Curriculum Based Measures have been developed delineating mastery of basic skills for each measure based on the grade level

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

- Procedures for determining cut scores are well documented
- For the past several years, many districts have been engaged with standards based grading.

In addition, we have developed a procedure to make decisions about students as part of our process of identification of students in need of additional academic support, as well as identification of students for our advanced, accelerated and gifted programming. We have piloted and refined this procedure of *convergence and magnitude* for decision-making. This decision-making heuristic is based on the idea that **if most of the data we have about a student indicates a problem, then there is a problem**. Alternatively, if most of the data we have about a student indicates the student's performance is exceptional, then it is likely that the student may require additional challenge. When there are not enough data to draw a conclusion, we collect more information. For the purposes of teacher and principal evaluation, the convergence *and magnitude* heuristic is altered slightly in the following manner:

1. Normative references are replaced with ordinal criterion references including: Academic Warning, Below Expectations, Meets Expectations, and Exceeds Expectations.
2. Instead of examining a compilation of single tests, and assessments, the focus is on the categorical change from one time to the next, then categorized as (unsatisfactory growth (-2 or -1), growth that is neutral or needs improvement (0), growth that is proficient (1) or growth that is excellent (2).

The specific application of the *convergence and magnitude* heuristic to a fully standards based approach is straight forward. The specific assessments used in the model only require two assumptions beyond traditionally required psychometric characteristics are met to be included in the model. First, students must have data from two scores for each test, i.e., time 1 and time 2. All students do not have to have data for all assessments used in the model. Missing data is not estimated in the model, rather missing data are simply not considered for the individual student. Second, each test must result in a score that equates to one of four ordinal performance levels, e.g., Academic Warning, Below Expectations, Meeting Expectations, or Exceeding Expectations.

Value tables place a numerical value on the social value of growth. The convergence and magnitude heuristic works well to evaluate assessments at a point in time, but to evaluate growth an additional procedure is needed. The procedure, referred to as "Value Tables" (Hill, et.al., 2005) sets a numerical value to the worth (or social value) of growth based on the movement of student scores through ordinal bands of performance categories from one time to another. There are six steps to determining student growth using value tables to summarize the convergence of data from multiple sources and types of assessment to evaluate student growth. The steps are presented below:

1. Classify the level of performance on any measure as Academic Warning, Below Expectations, Meeting Expectations, or Exceeding Expectations. This performance category rating will be used for subsequent analyses.
2. For each score-pair (i.e., a measure in which students have been assessed more than once) , assign a numeric value to each growth designation (i.e., -2, -1, 0, 1, 2) as proposed in Table 1.

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

3. For each student, determine a score for each score-pair, using the performance matrix presented in Table 1. A hypothetical example is presented at the end of this proposal.
4. Because the growth designation values are ordinal, the appropriate measure of central tendency (average) is the median; thus take the median value for the growth determination scores as an indication of typical growth for each student. The precise equation for the calculation of the median is presented in the appendix.
5. The overall summary score for each unit, i.e., class, grade level, school, is the median of that unit. As such, the median growth determination for all students within a given unit is the overall growth value to be applied to the evaluation system. A hypothetical example of five classrooms each with different overall summary indices is example is presented at the end of this proposal.
6. Finally, the overall summary score for the unit is converted into a performance category label. An example of how performance categories might be constructed is presented in Table2, and illustrated in the hypothetical example.

The values summarized in Table 1 are an initial numerical solution to the social value of student growth. In the development of final value tables, key stake holders including at a minimum, those evaluating and those being evaluated should be surveyed to place the final numerical values into the matrix.

Table 1 Summary of growth designation values

Time 1 Performance	Time 2 Performance Level			
	A. Warning	Below	Meets	Exceeds
Academic Warning	-2	1	2	2
Below Expectations	-2	0	1	2
Meets Expectations	-2	-1	1	2
Exceeds Expectations	-2	-2	-1	2

Note. Numerical values presented in Table 1 represent the relative worth of growth from one assessment time period to the next (i.e., time 1 to time 2). The values reflect a judgment and may vary depending on the focus for improvement.

Examination of the numerical values in Table 1 reveals the following social values:

1. Exceptional Growth (2) as well as maintenance of exceptional performance is highly valued, thus any score-pair that moves up two or more categories or ending in the highest performance category is worth two points.
2. Proficient Growth i.e., Meets (1) is defined as any score-pair that moves up one category or is maintained i.e., scores growth at a rate commensurate with the increasing expectations of the Meeting Expectations category.
3. Inadequate Growth (-1) is defined as a score-pair in which the performance category at time 1 is higher than the performance category at time 2, or growth from time 1 to time 2 is not sufficient to move a student from the below expectations category into the meeting expectations category.
4. Unsatisfactory Growth (-2) is defined as a score-pair in which the performance category drops by two categories from time 1 to time 2 or ends in the Academic Warning Category.

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

Table 2. Potential Indicators of Growth Values

Growth		
Value	Performance Category	Description
> 0	Unsatisfactory	The performance category drops by two categories from time 1 to time 2 or ends in the Academic Warning Category
0	Needs to Improve	The performance category at time 1 is higher than the performance category at time 2, or growth from time 1 to time 2 is not sufficient to move a student from the below expectations category into the meeting expectations category.
1	Proficient	The performance category moves up a single category from time 1 to time 2 or growth is at a rate commensurate with the increasing expectations of the meeting expectations category
2	Excellent	The performance category moves up two categories from time 1 to time 2 or growth is at a rate commensurate with the increasing expectations of the exceeds expectations category

No system of evaluation is perfect. It is highly questionable whether complex applications statistical models of growth are **trustworthy** (Reckase, 2004). These models require assumptions that are not tenable for educational data. When the assumptions of models are not met, they are as likely to obscure interpretation as they are to clarify. Furthermore, the complexities of the models are **inaccessible** to the public. Without advanced training in statistics, the decisions made in the process of the application of complex statistical models are not likely to be questioned; however, these decisions, more often than not result in the success or failure of the model to find differences in scores. Because of the complexity of the models, the information is not **useable** at a time when something can be done about it, resulting in a "gotcha" environment. There is no doubt that the precise measurement of student growth is far more complex than it would appear at first glance.

Contrary to complex models, the model of *convergence and magnitude* for decision-making presented here is *flexible*, yet *transparent*. The question it addresses is general, **Is there sufficient evidence that students are growing?** Because the question is general, the analysis is *conservative*. In addition, the simplicity of the model means that the results are more likely to be used *formatively*, in-process, leading to *on-going improvement* of teaching and learning though a self -correcting model of data collection and instructional decision – making.

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

A hypothetical example

Classify the Growth of a Single Student. The question of whether or not a school is demonstrating sufficient growth begins with the ability of a model to describe the growth of a single student across time, on multiple measures. Once the scores of a single student have been summarized, the next task is to summarize the growth for many students. Finally, the model should aggregate and describe the growth of the school. To illustrate this process, the results of five tests, each measured at two points in time for a single student, are summarized and described in Table 3.

Table 3. An example of a individual student performance on five tests, each administered two times

	Time 1	Time 2	Growth index	Description of growth
Test 1	Warning	Below	1	Though below expectations, performance is increasing
Test 2	Below	Below	0	Performance is below expectations and not changing substantially.
Test 3	Meets	Meets	1	Performance is meeting expectations, though not changing substantially.
Test 4	Below	Meets	1	Most current test indicates the student is currently meeting expectations, performance is increasing
Test 5	Meets	Below	-1	Most current test indicates the student is currently below expectations, performance is decreasing

The median growth index is calculated as the middle value in a group of scores (the precise statistical calculation is presented at the end of the appendix). In this case the median growth index for this student is simply the middle score. This is easily visualized by arranging the growth indices in order from lowest to highest, then canceling out high and low values (i.e., ~~-1, 0~~, ~~-~~1, ~~1, 1~~) . The median score of 1, proficient growth is apparent. This means that on average, this student's performance is sufficient, though not exceptional.

While this works quantitatively, (i.e., the numbers work) the question of whether this works qualitatively must be examined. A qualitative description of the data indicate that performance on test 1 cancels performance on test 2, and performance on test 4 cancels performance on test 5. This leaves the median growth index / description of growth for Test 3 as most representative of student growth. Using the growth table 2 (or Table 4 below), this performance would be designated as proficient growth

Table 4. Potential Indicators of Growth Values

Growth	
Value	Performance Category
< 0	Unsatisfactory
0	Needs to Improve
1	Proficient

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

2 Excellent

Classify the Growth of a Group of Students. Following the calculation of growth values for a particular student, is the calculation of growth values for all students within and across classrooms. For example in Table 5, growth values for five classrooms each representing a different overall growth performance level are listed.

Table 5. Example of 5 classrooms, each demonstrating a different level of growth performance

Example Student	Class 1	Class 2	Class 3	Class 4	Class 5
Each row within each column represents a different student indicating there are 26 students in a class, not that that each student is in each class.	-2	-1	0	2	0
	-2	-1	0	2	1
	-2	-1	-1	1	-1
	-2	-2	-1	0	2
	-2	-1	-1	1	2
	-1	-1	-1	2	2
	-1	-1	0	2	2
	-1	1	0	1	1
	-1	-1	0	0	-1
	-2	-1	-1	-1	2
	-2	0	0	-1	2
	2	-1	-1	-1	2
	-2	0	-1	-1	2
	-2	0	-1	2	2
	-2	0	-1	0	2
	-2	0	2	2	2
	-2	0	0	2	1
	-2	-1	-1	2	1
	0	0	-1	1	1
	-2	-1	1	-1	0
	0	-1	0	0	1
	-1	-1	0	1	2
	0	0	2	1	0
	1	2	1	2	2
	2	2	2	1	2
	1	0	0	1	1

Examination of the average (median) growth performance levels across class illustrates how student performance may be summarized and categorized by the next larger unit of analysis, the classroom. Of course, using this information for teacher evaluation would assume the individual teacher was the sole contributor to the growth or lack thereof; this assumption is not tenable in today's schools, but this is the subject of the application of this model to teacher evaluation.

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

Table 6. Example performance of five classes

Overall	Class 1	Class 2	Class 3	Class 4	Class 5
Growth	-2	-1	0	1	2
Performance Category	Unsatisfactory		Needs to Improve	Proficient	Excellent

Summarize Growth. Taken together, the median value for these 130 students across five classrooms is 0 (i.e., ~~-2, -1,~~ 0, ~~1, 2~~). This means that in this demonstration example, although two classrooms have demonstrated growth in the proficient or excellent category, overall, the grade level or principal of the school if this represented a school, would be *Needs to improve.*

Illinois law requires that for the 2012 -13 school year, 25% of principal or assistant principal evaluations are constituted of student growth, where growth is defined as progress measured over *two or more points in time* (23 Il Admin. Code 50.30) with assessments that ***best*** *measure the impact of the principal, school and school district with regards to student growth* (23 Il Admin. Code 50.310).

Goal Setting. The above example will be used to demonstrate how a goal for performance could be written, of course with this in mind, it is important to state, ***the example data are fictitious and designed to illustrate groups of students performing at different designations of growth performance.*** So for setting goals with example the above, the student growth index (i.e., -2, -1, 0, 1, 2) is the median growth index of any number of relevant measures given a two points in time, where growth values have then been substituted via growth tables. Although an overall growth summary may be sufficient for some goals, a principal may want to target a particular portion of the distribution of scores, i.e., the percentage of students demonstrating adequate growth (i.e., combined proficient and excellent growth). The flexibility of this model not only allows for a summative rating but also targeted goal setting. The percentage of students demonstrating growth across the four performance categories is presented in Table 7.

Table 7. Summary of students demonstrating growth across performance categories

Growth		Class 1		Class 2		Class 3		Class 4		Class 5		Overall
Value	Category	count	(%)	count	(%)	count	(%)	count	(%)	count	(%)	(%)
> 0	Unsatisfactory	20	*77%*	14	*54%*	11	*42%*	5	*19%*	2	*8%*	***40%***
0	Needs to Improve	4	*15%*	9	*35%*	12	*46%*	4	*15%*	3	*12%*	***25%***
1	Proficient	1	*4%*	2	*8%*	1	*4%*	8	*31%*	7	*27%*	***15%***
2	Excellent	1	*4%*	1	*4%*	2	*8%*	9	*35%*	14	*54%*	***21%***
Adequate Growth			8%		12%		12%		65%		81%	**35%**

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

Alternatively, goals may be set based on a characteristic (i.e., subgroup) using the same strategy. For example, the last five students listed in each class may make up a cohort of students who are specifically targeted for improvement. While the overall assessment of growth performance of the students in the building may require improvement, the evaluation of the specific subgroup may result in a different outcome i.e., 68% of students in the cohort demonstrating adequate progress versus 35% in the entire group.

Table 8. Summary of a cohort of students demonstrating growth across performance categories

Growth		Class 1		Class 2		Class 3		Class 4		Class 5		Overall
Value	Category	count	(%)	count	(%)	count	(%)	count	(%)	count	(%)	(%)
> 0	Unsatisfactory	1	*20%*	1	*20%*	0	*0%*	0	*0%*	0	*0%*	***8%***
0	Needs to Improve	1	*20%*	2	*40%*	2	*40%*	0	*0%*	1	*20%*	***24%***
1	Proficient	2	*40%*	0	*0%*	1	*20%*	4	*80%*	1	*20%*	***32%***
2	Excellent	1	*20%*	2	*40%*	2	*40%*	1	*20%*	3	*60%*	***36%***
Adequate Growth			60%		40%		60%		100%		80%	**68%**

According to the Illinois Principals Association (IPA, 2012) five steps guiding school districts in developing student growth components that are now required to be a part of principal and assistant principal evaluations include:

1. Identify growth targets
2. Identify relevant cohort
3. Determine which measurements to use
4. Determine data points
5. Determine what constitutes adequate students growth (IPA, 2012)

Growth targets and cohorts are determined by stakeholders at the building in alignment with the district plan. The purpose of this document thus far has been to elucidate a process that may be used to include all relevant data to maximize the reliability and subsequent consequential validity of principal evaluations. Further, the use of the heuristic of convergence and magnitude within the context of value tables simplifies the use of multiple sources of data. The final step in the process, d*etermining what constitutes adequate students growth* is a judgment that is made based on the status quo of the district (i.e., how much growth students have made in the past) combined with a desire for continuous improvement.

Three sets of potential goals are presented below using this strategy. The first two sets of goals specify that the degree of progress demonstrated by students will increase, with a simultaneous decrease in students making unacceptable progress. The strength of this type of goal is that it defines a change in overall performance in terms of growth. This type of goal is appropriate when the amount of progress students are making is not adequate. The third set of goals is based on using the summary function of the model described here.

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

Goal Set 1 - All school goal setting (recall hypothetical data are lower than actual data for illustrative purposes)

1. By February 2013, given available Type I and Type II assessments administered at two points in time school wide, students at school name will demonstrate an **increase in the proportion making adequate progress** from 35% to 40%.
2. By February 2013, given available Type I and Type II assessments administered at two points in time school wide, students at school name will demonstrate a **decrease in the proportion making unsatisfactory progress** from 40% to 30%.

Goal Set 2 – Cohort goal setting

1. By February 2013, given available Type I and Type II assessments administered at two points in time school wide, students identified as define cohort at school name will demonstrate an **increase in the proportion making adequate progress** from 68% to 80%.
2. By February 2013, given available Type I and Type II assessments administered at two points in time school wide, students identified as define cohort at school name will demonstrate a **decrease in the proportion making unsatisfactory progress** from 8% to 4%.

Goal Set 3 – Overall School / Grade Level / Cohort Goals

1. By February 2013, given available Type I and Type II assessments administered at two points in time school wide, students at School Name will demonstrate **excellent** growth as defined by the convergence and magnitude of data classified with district define value tables.
2. By February 2013, given available Type I and Type II assessments administered at two points in time school wide, students at 4th grade will demonstrate **excellent** growth as defined by the convergence and magnitude of data classified with district define value tables.
3. By February 2013, given available Type I and Type II assessments administered at two points in time school wide, students Identified as English Language Learners at School Name will demonstrate **excellent** growth as defined by the convergence and magnitude of data classified with district define value tables.

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

Appendix

The precise formulation of the predicted rating is specified in Equation 1 as the median rank of the median of the available student test vectors. Specifically, the median score for each student is calculated, then the median of calculated scores in the larger unit (i.e., grade level) is obtained and converted via growth values specified in Table A2.

$$\widehat{Rating} = M_{unit}\left(\overrightarrow{M_{st.scores}\left[\overrightarrow{s_1}, \quad \overrightarrow{s_2}, \quad \dots, \overrightarrow{s_n}\right]}\right)_j \qquad \text{(Equation 1)}$$

Where *M* is the median score, and $\overrightarrow{s_1}$, $\overrightarrow{s_2}$, and $\overrightarrow{s_n}$ are the vectors representing the median growth value for students within a unit (i.e., grade), and $\overrightarrow{(\dots)_j}$ is the vector of units (i.e., grade levels) within a larger unit (i.e., schools). Medians are calculated as the value representing the midpoint in the length of the available vector of scores. The specific statistical formula used in the calculation of the median for a group of scores is shown below:

$$M = l + \left[\frac{\frac{N}{2} + \Sigma f_{below}}{\Sigma f_{within}}\right] c.i \qquad \text{(Equation 2)}$$

Where l = exact lower limit of the class interval upon which the median lies
$N / 2$ = one – half the total number of scores
$\sum f_{below}$ = sum of the scores on all intervals below or
$\sum f_{within}$ = frequency within the interval upon which the median falls or
c.i. = Length of the class interval (*if >1*)

For example, if 140 students were assessed on a particular test, the following frequency distribution might be obtained.

l = 177.5

$\frac{N}{2} = \frac{140}{2}$ = 70

$\sum f_{below}$ = 68

$\sum f_{within}$ = 20

Score	*f*	*cum f*
208	2	140
185	8	138
184	16	130
183	10	114
179	16	104
178	20	88
177	19	68
176	18	49

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

c.i. = 1	**165** 12 31 **164** 10 19 **156** 7 9 **154** 2 2
	= 177.5 +[(70 - 68)/20]*1 = 178

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com

References

Ballou, D. (2008, April). Test scaling and value-added measurement. Paper presented at the National Conference on Value-Added Modeling, Madison, WI.

Betebenner, D. W. (2008). Toward a normative understanding of student growth. In K. E. Ryan & L. A. Shepard (Eds.), The future of test-based educational accountability (pp. 155–170). New York, NY: Taylor & Francis.

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. Education Measurement: Issues and Practice, 28(4), 42–51.

Braun Henry. 2005. Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models. Princeton, N.J.: Educational Testing Service. http://www.ets.org/Media/Research/pdf/PICVAM.pdf

Briggs, D. C., & Weeks, J. P. (2009). The impact of vertical scaling decisions on growth interpretations. Education Measurement: Issues and Practice, 28(4), 3–14.

Carnine, D. (1997). Bridging the research-to-practice gap. Exceptional Children, 63, 513-521.

Good, R. H. (1998). Using the Outcomes-Driven Model and DIBELS for Response to Intervention. http://www.ospaonline.org/pdf/presentations/Good_handouts_2009.pdf

Hill, R, Gong, B., Marion, S., DePascale, C., Dunn, J., Simpson, M. (2005). Using Value Tables To Explicitly Value Student Growth, NCIEA, November 2005, www.nciea.org/publications/MARCES_RH07.pdf.

Illinois Principals Association. (2012). Illinois principal evaluation plan identifying student growth criteria. http://www.ilprincipals.org/resources/resource-documents/principal-evaluation/ipep/copy_of_IdentifyingStudentGrowthCriteria.pdf

Krueger, A & Lindahl, M. (2001), .Education for growth: Why and for whom?., Journal of Economic Literature 39(4), 1101.1136.

Martineau, J. A. (2006) Distorting value added: the use of longitudinal, vertically scaled student achievement data for growth-based value-added accountability. Journal of Educational and Behavioral Statistics, 31(1), 35-62.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. Educational Researcher, 18(2), 5-11.

O'Malley, K. (2008, June). Review of student growth models used by states. Retrieved from Pearson Web site: http://www.pearsonassessments.com/NR/rdonlyres/E2F72E40-5CBD-48B9-9033-2830295A8DA8/0/Student_Growth_Models.pdf?WT.mcn_id=TMRS_Review_of_Student_Growth_Models

Reckase, M. D. (2004). The real world is more complicated than we would like. Journal of Educational and Behavioral Statistics, 29, 117-120.

Reckase, M. D., & Martineau, J. (2004). The vertical scaling of science achievement tests. Paper commissioned by the Committee on Test Design for K-12 Science Achievement. Center for Education: National Research Council.

Tukey, J. W. (1962). The future of data analysis. Annals of Mathematical Statistics 33(1), 1–67.

Yen, W. M. (1986). The choice of scale for educational measurement: An IRT perspective. Journal of Educational Measurement, 23(4), 299–325.

written by Ben Ditkowsky, Ph.D.
Ben@MeasuredEffects.com