Journal of Research Methodology, Volume 21, Number 1 (January-April 2008)

วารสารวิธีวิทยาการวิจัย ปีที่ 21 ฉบับที่ 1 (มกราคม-เมษายน 2551)

คะแนนพัฒนาการสัมพัทธ์สะสมและคะแนนทัดเทียมตามลำดับเวลา

ธีระ อาชวเมธี

กมลวรรณ ตังธนกานนท์

บทคัดย่อ

บทความนี้มีจุดเน้นที่โค้งการเรียนรู้/พัฒนาการซึ่งสามารถแทนได้ด้วยสมการ $y=1-e^{-ax}$ โค้งการเรียนรู้/พัฒนาการดังกล่าวแสดงให้เห็นถึงการเพิ่มขึ้นในอัตราที่ลดลงอย่างต่อเนื่องไปตามเวลา ก่อนหน้านี้มีคะแนนสองชนิดที่ได้ถูกเสนอไว้เพื่อถ่วงคะแนนความแตกต่างใกล้ช่วงสุดท้ายของโค้งการเรียนรู้คือ (1) คะแนนพัฒนาการสัมพัทธ์ [ซึ่งมีค่าเท่ากับ "$(y_2 - y_1)/(F - y_1) \times 100$" โดยที่ y_2 คือคะแนนการประเมินครั้งหลัง และ y_1 คือคะแนนการประเมินครั้งก่อน และ F เป็นคะแนนเต็มในการประเมิน] และ (2) คะแนนพัฒนาการสัมพัทธ์สะสม บทความนี้แสดงให้เห็นว่าถ้าเรารู้สัมประสิทธิ์ "a" ของสมการโค้งการเรียนรู้ คะแนนพัฒนาการสัมพัทธ์สะสมก็จะสามารถแปลงไปเป็น "คะแนนทัดเทียมตามลำดับเวลา" ได้โดยสูตรง่าย ๆ ตัวอย่างการใช้คะแนนทัดเทียมตามลำดับเวลาในทางการศึกษาได้แก่ "คะแนนทัดเทียมเกรด" บทความนี้ยังได้พัฒนาวิธีการทางคณิตศาสตร์เพื่อกำหนดช่วงเวลาที่คะแนนพัฒนาการของคะแนนทัดเทียมเกรดเริ่มที่จะแตกต่างกันมากกว่าคะแนนพัฒนาการแบบดั้งเดิม นอกจากนี้ยังได้แสดงการประยุกต์ใช้คะแนนทัดเทียมตามลำดับเวลากับข้อมูลการวัดผลอิงหลักสูตรทางด้านการอ่านซึ่งเก็บรวบรวมได้จากโรงเรียนประถมขนาดเล็กแห่งหนึ่งในเขตชนบท

Cumulative Relative Gain Score vs. Chronological Equivalence Score

Teara Archwamety

Kamonwan Tangdhanakanond

ABSTRACT

This paper focused on a common type of learning/development curve represented by the mathematical equation: $y = 1 - e^{-ax}$. *This learning/development curve shows continuously decreasing gain as a function of time. In the past, (a) a Relative Gain Score (RGS)-defined as "*$(Y_2 - Y_1)/(F - Y_1) \times 100$*" where* Y_2 *is post-evaluation score,* Y_1 *is pre-evaluation score, and F is full score of the evaluation, and (b) a Cumulative Relative Gain Score (CRGS), were formulated to compensate for difficult gain near the end of the learning curve. In the present paper, it was shown that when the coefficient "a" of the learning curve equation is known the CRGS could be easily converted into pre-post Chronological Equivalence Scores (CES). A common CES used in education is the Grade Equivalent Score (GES). The present paper also formulates a mathematical method to determine the point in time where a grade equivalent gain score starts to show higher magnitude than its traditional gain score counterpart. Finally, the CES formulation was applied to a set of R-CBM data collected from a small rural elementary school.*

Introduction

Tracing students' physical, cognitive (intellectual or academic), and social development is probably one of the most important tasks of educators. Archwamety, Tangdhanakanond, and Pitiyanuwat (2005) proposed three fundamental forms of these learning/development curves. One of these three forms was the focus of a subsequent paper by Archwamety and Tangdhanakanond (2007). This learning/development curve *(see Figure 1)* has the form:

$$Y = 1 - e^{-aX} \qquad (1)$$

where Y is a measure of learning or development

X is a measure of time

a indicates the speed the curve approaches maximum

This form of learning/development curve is applicable to many situations. Examples appear in Brown and Saks (1985, p. 124), in Hulse, Deese and Egeth (1975, pp. 24-26), in Haber and Fried (1975, pp. 172 & 185), and in LeFrancois (1995, p. 39). Its application in the area of Curriculum-Base Measurement in Reading (R-CBM) is shown in Figure 2. The non-smooth curve shows the "national norm" growth of reading measured in number of words read correctly per minute from first grade to eight grade (AIMSweb, 2006). The smooth curve superimposed on it is the result of using Equation (1) as the theoretical model of growth.

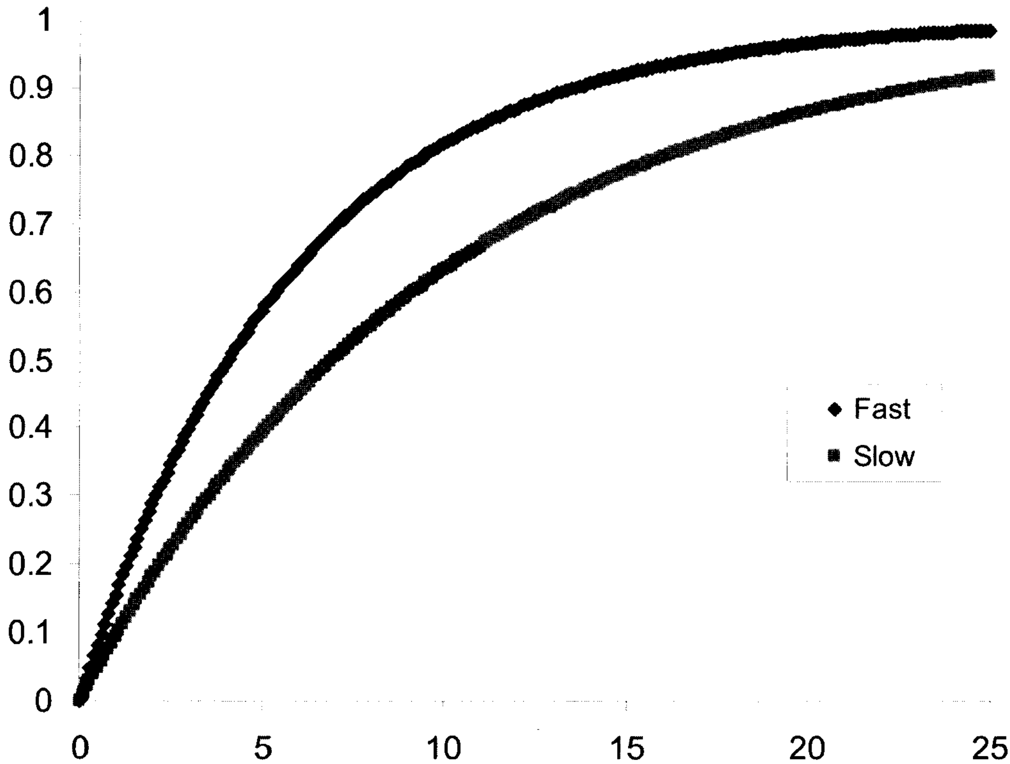

Figure 1 Learning curves modeled by $Y = 1 - e^{-aX}$ In the upper curve (fast), a = 0.17 and in the lower curve (slow), a = 0.01.

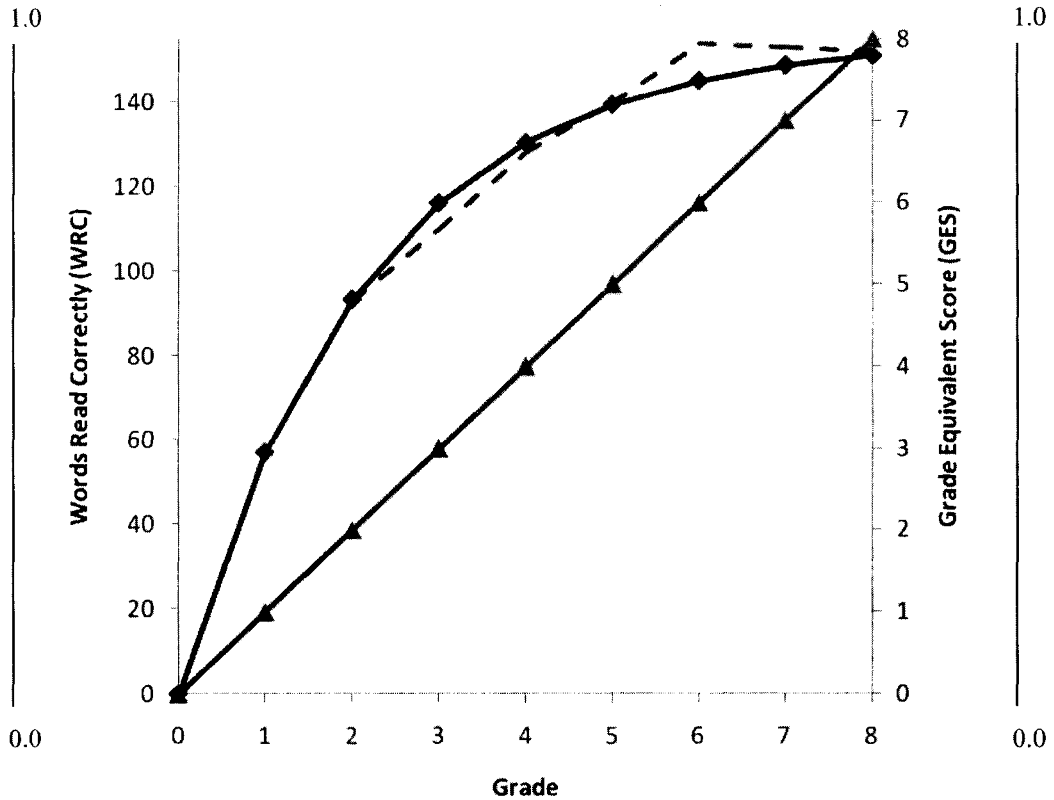

Figure 2 AIMS web data in raw WRC score as well as unit length (left vertical axis) together with corresponding GES and its unit length (right vertical axis) as a function of grade. Least Square method was used in fitting the learning curve " $Y = 1 - e^{-aX}$ " (a = 0.46) on the WRC data.

One drawback of this form of learning/development curve is that it shows very little improvement as students move towards the end of learning or development. Kanjanawasee (1989) proposed the concept of "Relative Gain Score" (RGS) to correct this shortcoming (also see Ruengtrakul, 2002). The RGS is defined as:

$$\text{Relative Gain Score} = \frac{Y_2 - Y_1}{F - Y_1} \times 100 \qquad (2)$$

where Y_2 = Score of post-evaluation

Y_1 = Score of pre-evaluation

F = full score of the evaluation

Note that $F-Y_1$ becomes smaller and smaller as the gain score is taken towards the end of the learning curve-thus giving more and more weight to Y_2-Y_1. This is an improvement over the traditional gain score considering the nature of the learning curve. The RGS places progressively higher weights for the gains made towards the end of learning or development. However, Archwamety and Tangdhanakanond (2007) pointed out that the RGS index is not additive, and thus proposed the concept of "Cumulative Relative Gain Score" (CRGS) which has the desirable property of additivity. The CRGS that corresponds to the form of leaning/development curve shown in Equation (1) is defined as:

$$\Delta y_1/(1-y_1) + \Delta y_2/(1-y_2) + \ldots + \Delta y_n/(1-y_n) = \int_a^b \frac{1}{1-Y}\, dY$$

$$= -\ln(1-Y)\ \Big]_a^b$$

$$= \ln(1-Y_a) - \ln(1-Y_b) \qquad (3)$$

where n goes from 1 to infinity, *a* is the beginning of *Y* (pre-evaluation), b is the end of Y (post-evaluation), and the full score of evaluation is unity. Note that this "*a*" is not the same as "*a*" in Equation (1).

The purposes of the present paper are :

1. To show the relationship between CRGS and Chronological Equivalence Score (CES) such as the well-known "Grade Equivalent Score" (GES).

2. To propose a mathematical technique of calculating at what point in learning/development would Grade Equivalent Score show higher gain than Raw Score.

3. To present the application of the above concepts to R-CBM data collected from an elementary school in the U.S.

CRGS, CES, and GES: Theory and Data

Relationship between Cumulative Relative Gain Score (CRGS) and Chronological Equivalence Score (CES)

Starting with the basic learning/development curve equation, $Y = 1 - e^{-aX}$ let's express X (chronological score) in terms of Y (unit-based score) instead, as follows:

$$Y = 1 - e^{-aX}$$

$$e^{-aX} = (1 - Y)$$

$$-aX = \ln(1 - Y)$$

$$X = \frac{\ln(1 - Y)}{-a} \qquad (4)$$

Next, let Y_1 be pre-evaluation score and Y_2 be post. Then the gain score in X can be expressed as

$$X_2 - X_1 = \frac{\ln(1 - Y_2)}{-a} - \frac{\ln(1 - Y_1)}{-a}$$

$$X_2 - X_1 = \frac{1}{a}\left[\ln(1 - Y_1) - \ln(1 - Y_2)\right] \qquad (5)$$

Thus, comparing Equation (5) with Equation (3), it is clear that the gain in Chronological Equivalent Score (CES) "X_2-X_1" is equal to CRGS divided by the coefficient "a" of the basic learning/development curve equation.

When does Grade Equivalent Score show higher gain than Raw Score?

As seen in *Figure 1*, as well as the primary (left-hand side) vertical axis in *Figure 2*, which is a graph of Equation (1), a raw score shows a "decreasing" rate of growth from one point in time to the next. By contrast, a chronological equivalence score shown as the secondary (right-hand side) vertical axis in *Figure 2*, which is the same as the time line (horizontal axis), shows a "constant" rate of growth from one point in time to the next.

In Equation (1), *Y* represents any measure of learning or development rescaled to a maximum of 1.0 *X* could be taken to be any measure of chronological units such as clock time, day, week, month, year, age, grade, and so on. According to the form of Equation (1), raw score has a maximum of 1 while chronological equivalence score (CES) has a theoretical maximum of infinity. However, in practice, the "practical maximum of chronological equivalence score" (maxC) should be at about the point where the learning/development curve has reached the plateau approaching the maximum. This "practical" maximum ("maxC") could be used to rescale any CES to have a maximum of 1.0 by dividing it into the CES (x/maxC). Let "z" stand for this unitized CES ("z = x/maxC"). This would allow us to compare raw score (*Y*)'s rate of growth and CES (Z)'s rate of growth with fairness since both have a maximum of 1.0. For example, if the learning curve of a particular measure reaches the plateau at the eighth grade and a student has a Grade Equivalent Score of fourth grade, the student's "unitized" (or "standard") Grade Equivalent Score (*Z*) would be 0.5 (or 50%). *Figure 2* also shows the graph of raw score *Y* and chronological score *Z* as a function of time with both maxima set to one unit length each. With both types of score scaled to have a maximum of one unit, let us ask an interesting question. At what point on the time line will the rate of gain in chronological equivalence score be larger than the rate of gain in raw score?

To answer this question, we need to find the value of *X* in Equation (1) when

$$\Delta y \,/\, \Delta z = 1 \qquad (6)$$

where Δy is the rate of gain for *Y*

Δz is rate of gain for unitized CES or, at the infinitesimal level:

$$\frac{dY}{dZ} = 1 \qquad (7)$$

Because z = x/maxC, we have from Equation (1)

$$Y = 1 - e^{-a.\max C.X/\max C} = 1 - e^{-a.\max C.Z}$$

We next differentiate Y with respect to Z:

$$\frac{dY}{dZ} = a.\max C.\, e^{-a.\max C.Z}$$

But $a.\max C.e^{-a.\max C.Z} = 1$ [from Equation (7)]

Therefore $e^{-a.\max C.Z} = \dfrac{1}{a.\max C}$

$$-a.\max C.Z = \ln 1 - \ln[a.\max C]$$

$$Z = \frac{\ln 1 - \ln[a.\max C]}{-a.\max C}$$

$$Z = \frac{\ln[a.\max C]}{a.\max C} \qquad (8)$$

Thus, beyond the point where, $Z = \dfrac{\ln[a.\max C]}{a.\max C}$, the rate of gain in chronological equivalence score will be larger than the rate of gain in raw score. For example, if a = 0.46 and maxC = 8, then z = 0.35 and x = 2.83 grade equivalent. That is, rate of growth in terms of GES will be larger than rate of growth in raw score only from about the third grade on.

Application to R-CBM Data

R-CBM Data From the U.S. National Norm. One popular measure of reading progress in children in the literature of Curriculum-Based Measurement for Reading (R-CBM) is the WRC (number of Words Read Correctly). Shown in *Figure* 2 is the graph of the normative data on WRCs of first to eighth graders obtained from AIMSweb (2006). The smooth curve superimposed on it is the result of a Least Square curve fitting (a = 0.46) using the learning/development curve represented by Equation (1): $Y = 1 - e^{-aX}$. The original scale on the Y axis in number of words read correctly (WRC) with its maximum set at 155, as well as its being rescaled to unity, is shown on the primary (left-hand side) vertical axis. Similarly, the original scale of Grade Equivalent Score, as well as its being rescaled to unity, is shown on the secondary (right-hand side) vertical axis as a dependent variable. For normative data, the GES is simply the grade level from the X axis. Equation (4) : $X = \frac{\ln(1-Y)}{-a}$ in the previous section suggests an approach to convert a unit-length-max raw score into its corresponding Grade Equivalent Score. For example, a student whose raw WRC score is 140 would be assigned a GES of ln (1-140/155) divided by -0.46 = 5.08.

Unfortunately, the conversion shown above does not work with a student whose score is equal to or exceeds the set maximum of 155 because it will create a natural log of zero or negative which is undefined. The maximum of 155 might work fine for an average or below average student but not for those above. To make provision for virtually all students who are above average, one could use as a maximum the score at about two standard deviations above the average WRC at the 8th grade point on the norm graph. AIMSweb (2006) provided the average and standard deviation at this point as 152 and 45 respectively. Thus, setting the maximum WRC at 152+45+45=242 or approximately 240 would accommodate virtually all students for the purpose of converting raw WRC score to GES using the method described in the previous paragraph. With this new maximum WRC, the new least-square curve fitting using Equation (1): $Y = 1 - e^{-aX}$ yielded "a = 0.17" and resulted in a learning/development curve shown in *Figure* 3. Note that the X axis has been extended to about the 16th grade-an equivalence of senior year in college. This extension to the 16th grade can only be theoretical at the present time because no data have been collected for R-CBM beyond the 8th grade. This extended graph with an "*a*" coefficient of 0.17 and a WRC maximum of 240 can now be used to calculate a Grade Equivalent Score for virtually any WRC score. For example, a student whose raw WRC score is 140 would be assigned a GES of ln(1-140/240) divided by -0.17 = 5.15. A student whose raw WRC score is 200 would be assigned a GES of ln(1-200/240) divided by -0.17 = 10.54.

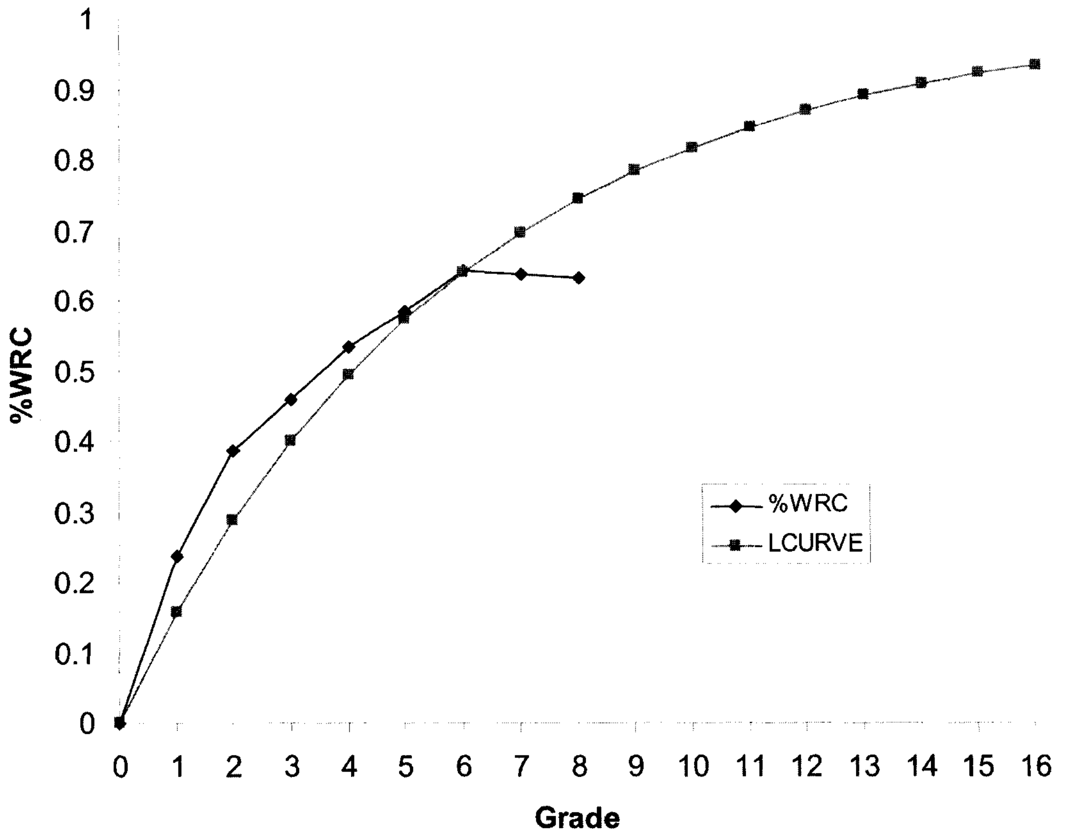

Figure 3 The smooth curve is the theoretical growth curve of a typical learner modeled by $Y = 1 - e^{-aX}$ projected beyond the eighth grade; The "*a*" coefficient = 0.17 by the Least Square curve fitting method; Estimated maximum WRC = 240 is rescaled to 1.0.

R-CBM Data from a Rural Elementary School in the U.S. Anderson and Christensen (2006) collected longitudinal WRC data from 69 sixth-grade regular students in a U.S. Midwestern state rural school district. Most of these students were tested for WRC when they were in the fifth and third grade. For each of the third and fifth-grade years, they were tested (or "probed") three times-Fall, Winter, and Spring. For the sixth-grade year, however, only Spring data were available. Thus, there were seven data points (seven WRC averages) that could be plotted to show the development trend. Because some of the students missed one or more of the tests, the seven averages did not have equal sample sizes. The students' raw WRC data were used in the present study to demonstrate the conversion from raw WRC score to GES score using the methods described in the previous sections. *Figure 4* shows the trend of the seven raw WRC averages as well as the trend of their corresponding GES averages. Note how closely the seven raw WRC averages fit the idealized curvilinear curve derived from AIMSweb national data. Also note how closely the seven GES averages fit the idealized linear trend.

Does the rate of growth in terms of GES larger than the rate of growth in raw

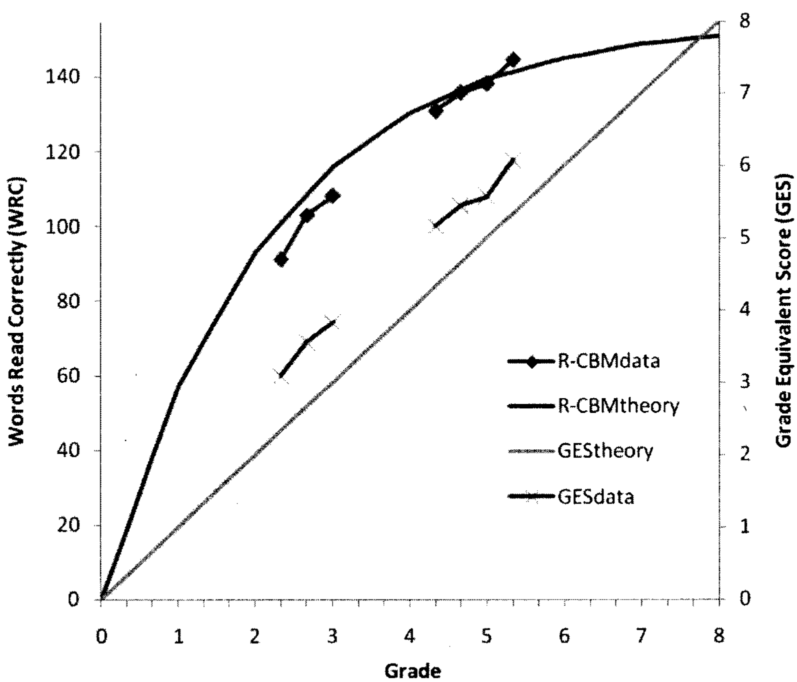

Figure 4 How the elementary school data in WRC fit the theoretical learning curve "$Y = 1 - e^{-aX}$" with a = 0.46, and how the same data in GES fit the theoretical linear line.

score from the third grade on? In the paragraph following Equation (8), it was suggested that, if the "*a*" coefficient in the idealized curve represented by Equation (1) is equal to 0.46 and maxC = 8, then z = 0.35 and x = 2.83 grade equivalent. That is, the rate of growth in terms of GES will be larger than the rate of growth in raw score only from the third grade on. The AIMSweb national data precisely fit this description (see *Figure* 2) and our data from the rural elementary school fit the national data (see *Figure 4*). Therefore, we could hypothesize that *"the rate of WRC growth for children in this school in terms of GES will be larger than the rate of growth in raw score from the third grade on."*

To test the above hypothesis, one could perform a t-test for related measures between Fall and Winter WRCs as well as between Winter and Spring WRCs, for both third grade and fifth grade, in terms of raw scores as well as grade equivalent scores. One should expect to see progressively higher level of significance. The results of these t-tests for related measures were as shown in Table 1.

Table 1

Means of Words Read Correctly (WRC) and of Grade Equivalent Scores (GES) by Fall(F), Winter(W), and Spring(S) Periods Within Grade, and t-statistics Testing Significant Difference Between Two Successive Test Points within Grade

Grade	*N*	**WRC**		**t**	**p**	**GES**		**t**	**p**
		M	***(SD)***			***M***	***(SD)***		
3F	69	91.10	(41.82)			3.09	(1.93)		
3W	67	103.10	(38.93)	3.56	0.00	3.56	(1.78)	2.73	0.00
3S	54	108.17	(38.97)	2.42	0.01	3.84	(2.04)	2.32	0.01
5F	65	130.86	(42.63)			5.17	(2.65)		
5W	65	135.78	(41.24)	2.41	0.01	5.46	(2.69)	2.19	0.02
5S	65	138.14	(40.14)	1.01	0.16	5.56	(2.60)	0.66	0.26
6F	65	144.54	(40.80)			6.07	(2.95)		

Contrary to the expectation, the several t-tests for related measures performed showed progressively "lower" level of significance for Grade Equivalent Scores as we moved from Fall vs. Winter, and Winter vs. Spring in the third grade, to Fall vs. Winter, and Winter vs. Spring in the fifth grade. This unexpected finding seemed to be the result of increasingly larger standard deviation as a function of time when Grade Equivalent Scores were used (see the standard deviations for GES vs. those for WRC in Table 1). The larger standard deviation seemed to offset the advantage of larger difference between means in a t-test.

The above explanation makes sense when we consider the theoretical situation of using raw scores vs. grade equivalent scores. *Figure 1* shows the theoretical curves when raw scores are used. *Figure 5* shows the theoretical linear lines when grade equivalent scores are used. In each case, the top line represents the fastest learner while the bottom line represents the slowest learner. One could imagine other lines in between representing the other learners. It is clear from the graphs that for raw scores the standard deviation grows larger at first and then becomes smaller later while for grade equivalent scores the standard deviation grows ever-increasingly larger.

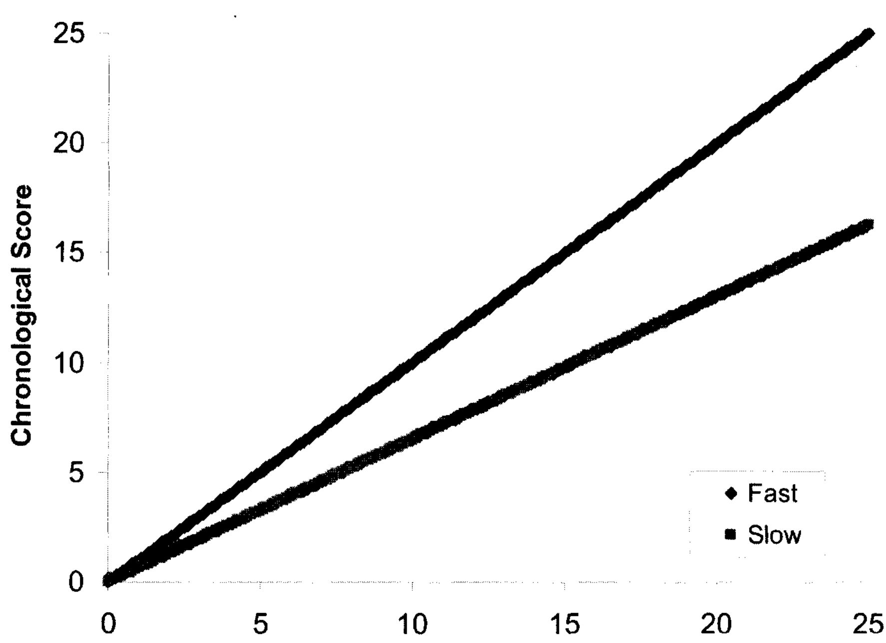

Figure 5 Chronological Equivalence Scores of a fast learner vs. slow learner as a function of Time.

Discussion

The finding in the previous section that a t-test for related measures, when Grade Equivalent Scores are used compared with raw scores, is less sensitive in detecting a significant difference, speaks against using grade equivalent scores when one is in need to show a significant difference. In such a situation raw scores or some other derived scores would be more appropriate. Many researchers also have found various cases against the use of Grade Equivalent Scores. For example, Reynolds and Willson (1984) criticized GES for having little to offer teachers or diagnostic personnel. Berk (1981) discussed the danger of using GES to identify students with learning disabilities. Weiner and Zibrin (1979) compared five reading achievement standardized tests and found dissimilarities in Grade Equivalent Scores generated which posed a threat to criterion-related validity. Finally, Taylor (1978) preferred standard scores to grade equivalents in measuring change.

On the positive side, Carver (1989) converted silent reading rates into grade equivalent units and found GES provide reasonable validity in evaluating the status and progress of individual students or groups. The authors of the present paper also think that the use of Grade Equivalent Scores may be appropriate in case studies where the concept of standard deviation of a group is irrelevant. In a case study, being able to say that John Doe Jr. has gained one grade level may sound better than he has gained only one point. Another possible appropriate use of grade equivalent scores is the area of "Hierarchical Linear Model" or HLM-an increasingly popular new approach to data analysis.

Application of Grade Equivalent Scores in a Hierarchical Linear Model

The concept of the Chronological Equivalence Score or Grade Equivalent Score discussed in this paper could also be applied to the recently popular statistical technique: Hierarchical Linear Model (HLM). A typical HLM starts with a person unit such as a student and then moves on to larger groupings such as class unit then school unit and so on. For example, in Pong and Pallas (2001) study, level-1 units were individual students with eighth-grade math achievement as outcome variable. Level-2 units were "classes" with class size as a variable. In Raudenbush, Rowan, and Cheong (1991) study, level-1 units were classes, level-2 units were teachers, and level-3 units were schools.

HLM actually could start with smaller than a person unit such as a time point when a measurement of a series of measurements of a person is taken. Jitendra, DuPaul, Volpe, Tresco, Junod, Lutz et al. (2007) studied 167 elementary school children with attention deficit hyperactivity disorder (ADHD) who were randomly divided into two consultation groups. Academic outcomes were assessed four times over a period of 15 months. One of the academic outcomes was R-CBM. Two-level Hierarchical Linear Modeling (HLM) was used in data analyses as follows (also see Raudenbush & Bryk, 2002).

Level 1:

$$Y_{it} = \beta_{i0} + \beta_{i1}t + r_{it}$$

where Y_{it} = R-CBM measure of an individual student *"i"* at time *"t"*

β_{i0} = intercept representing initial performance level

$\beta_{i1}t$ = slope representing rate of growth

r_{it} = error term

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

where W_j = the group independent variable [j = 0, 1]

Other terms = dependent variables, intercepts, slopes and errors

In another similar study, Evans, Serpell, Schultz, and Pastor (2007) studied 79 middle school youth with ADHD randomly preassigned to treatment and control groups. Several outcomes including academic were assessed several times during a period of three years. The same HLM designed mentioned above was used in data analyses.

In both studies above the Level 1 dependent variables "academic outcomes" used were not Grade Equivalent Scores or Chronological Equivalence Scores and therefore most likely are not a linear function of "time"-the independent variable. For example CBM measures as a function of time are well-known to be non-linear. This violates the important "linearity" assumption of a linear regression analysis approach. Rescaling those academic outcome variables as Grade Equivalent Scores would have been more appropriate. The important characteristic of a Grade Equivalent Score is that it increases at a constant rate of one unit per year in the norm group (Schulz & Nicewander, 1997)

References

AIMSweb (2006). *Charting the path to literacy: Reading fluency.* Retrieved February 18, 2006, from http://www.aimsweb.com/norms/reading_fluency.htm.

Anderson, N., & Christensen, J. (2006). *A program evaluation study on accountability: Linking data to intervention.* Paper presented at the 2006 Annual Meeting of the Naitonal Association of School Psychologists (NASP), Anaheim, California, March 28 - April 1, 2006.

Archwamety, T., & Tangdhanakanond, K.(2007). Equalizing Gain Scores: Relative Gain Scores vs. Cumulative Relative Gain Score. *Journal of Research Methodology, 20*(1), 1- 21.

Archwamety, T., Tangdhanakanond, K., & Pitiyanuwat, S. (2005). Learning and development curves: Their implications on Measuring gains. *Journal of Research Methodology, 18*(3), 213-227.

Berk, R. A. (1981). What's wrong with using Grade Equivalent Scores to identify LD children? *Academic Therapy, 17*(2), 133-140.

Brown, B. W., & Saks, D. H. (1985). Economic analysis of time and school learning. In C. W. Fisher & B. C. Berliner (Eds.), *Perspectives on instructional time* (pp. 119-129). New York: Longman.

Carver, R. P. (1989). Silent reading rates in Grade Equivalents. *Journal of Reading Behavior, 21*(2), 155-166.

Evans, S. W., Serpell, Z. N., Schultz, B. K. & Pastor, D. A. (2007). Cumulative benefits of secondary school-based treatment of students with attention deficit hyperactivity disorder. *School Psychology Review, 36*(2), 256-273.

Haber, R. N., & Fried, A. H. (1975). *An introduction to psychology.* New York: Holt, Rinehart and Winston.

Hulse, S. H., Deese, J., & Egeth, H. (1975). *The psychology of learning* (4th ed.). Tokyo: McGraw-Hill Kogakusha.

Jitendra, A. K., DuPaul, G. J.,Volpe, R. J., Tresco, K. E., Vile Junod, R. E., Lutz, J. G., et al. (2007). Consultation-based academic intervention for children with attention deficit hyperactivity disorder: School functioning outcomes. *School Psychology Review, 36*(2), 217-236.

Kanjanawasee, S. (1989). Alternative strategies for policy analysis: An assessment of school effects on students' cognitive and affective mathematics outcomes in lower secondary school in Thailand (Doctoral Dissertation, UCLA, 1989). Dissertation A*bstracts International, 50*(2-A), 345.

Lefrancois, G. R. (1995). *Theories of human learning* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Pong, S., & Pallas, A. (2001). Class size and eighth-grade math achievement in the U.S. and abroad. *Educational Evaluation and Policy Analysis, 23,* 251-273.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Rowan, B., & Cheong, F. Y. (1991). *Teaching for higher-order thinking in secondary schools: Effect of curriculum, teacher preparation, and school organization.* East Lansing: Michigan State University, College of Education.

Renolds, C. R., & Willson, V. L. (1984). Standardized Grade Equivalents: Really! No. Well, sort of, but they are more confusing than helpful. *Journal of Learning Disabilities, 17*(6), 326-327.

Ruengtrakul, A. (2002). The Development and analysis of the quality of growth measurement methods based on classical test theory and item response theory. *Journal of Research Methodology, 15*(1), 115-125.

Schulz, E. M., & Nicewander, W. A. (1997). Grade equivalent and IRT representation of growth. *Journal of Educational Measurement, 34*(4), 315-331.

Taylor, H. (1978). The misuse of Grade Equivalent Scores. *School Guidance Worker, 33*(4),11-15.

Weiner, M., & Zibrin, M. (1979). Dissimilarities in Grade-Equivalent Scores on different standardized tests of achievement: A threat to criterion-related validity. *Educational and Psychological Measurement, 39*(4), 923-928.