Journal of Research Methodology, Volume 18, Number 3 (September - December 2005)

วารสารวิธีวิทยาการวิจัย ปีที่ 18 ฉบับที่ 3 (กันยายน - ธันวาคม 2548)

Learning and Development Curves: Their Implications on Measuring Gains

Teara Archwamety
Kamonwan Tangdhanakanond
Somwung Pitiyanuwat

ABSTRACT

The present paper presented the various shapes of common learning and development curves. The mathematical models of these learning curves were also presented. Next, four implications of the learning curves were discussed. They were implications on (a) the sizes of standard deviations at various points through the course of learning in group data, (b) the case against examining only two points in time in learning and developmental studies, (c) the assumption of homogeneity of variance in ANOVA significance testing, and (d) the issue of measuring gain. Two improved indexes of measuring gain-Relative Gain Score and Gain Size—over the traditional gain score were then compared. Finally, the possibility of a few next generation gain indexes was brought up.

Introduction

Human learning or development as a function of time is not linear (see Figure 1). Instead, it is curvilinear. The curve could be in the form of decrease in acceleration as shown in Figure 2, or first increase in acceleration and then decrease in acceleration (S-shape) as shown in Figures 3 and 4. Note that in Figure 3, the lower portion of the S-shape is smaller than the upper portion while the Figure 4, the two portions are quite symmetrical.

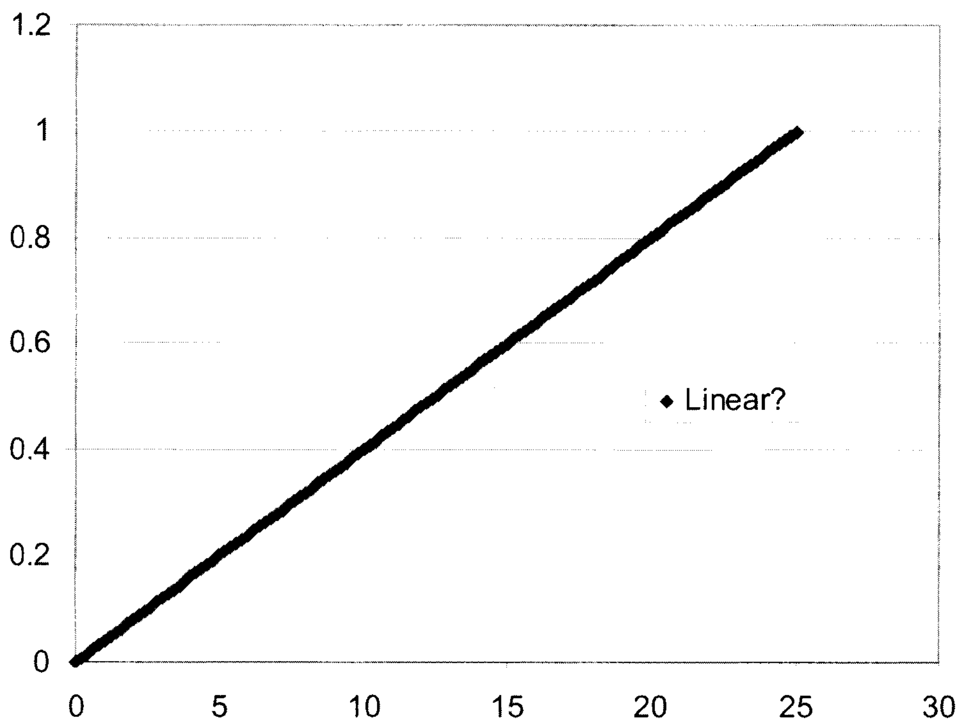

Figure 1 Is learning a linear function of time?

Examples of curves like Figure 2 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). Examples of curves like Figure 3 appear in Restle and Greeno (1970, p. 22), and in Zajonc and Markus (1975, p. 77). Examples of curves like Figure 4 appear in Brown and Saks (1985, p. 121), and in Hilgard and Bower (1966, p. 153). Mayer (1999, p. 42) also shows a multi-stage S-shape learning curve.

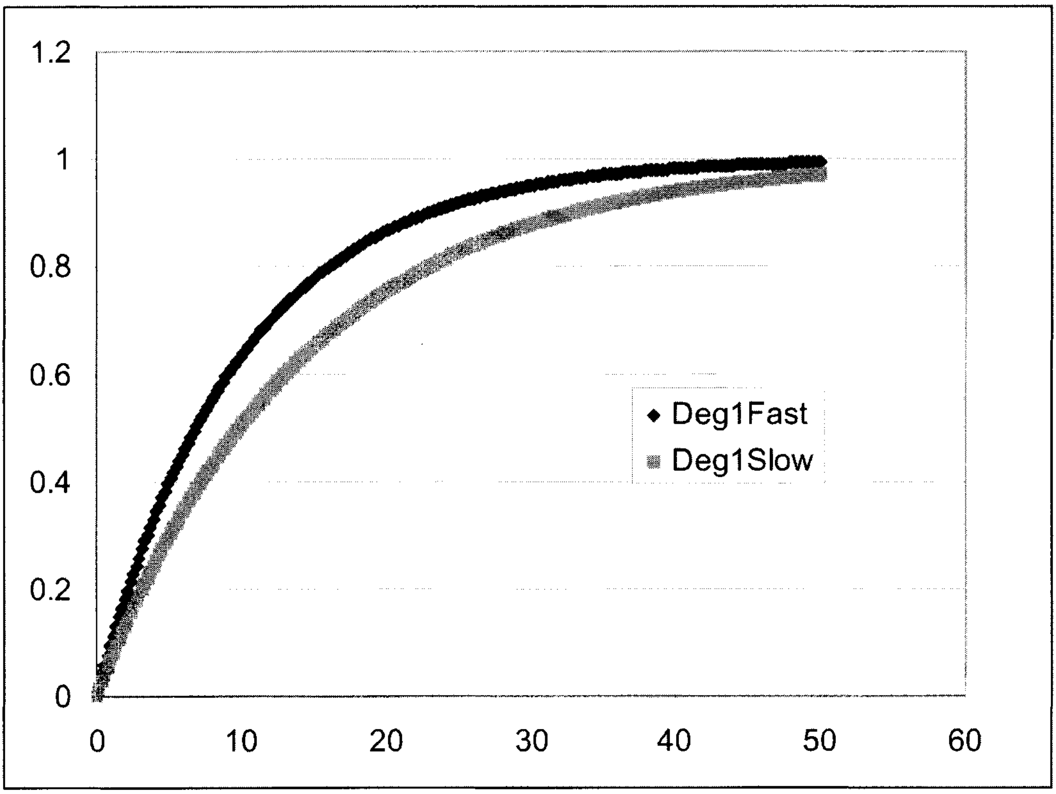

Figure 2 Learning curves modeled by $Y = 1-e^{-aX}$. In the upper curve a = 0.1 and in the lower curve a = 0.07.

Curves like Figure 2 could be modeled by a first-degree sigmoid equation:

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

where
- Y is a measure of learning or development
- X is a measure of time
- a indicates the speed the curve approaches maximum

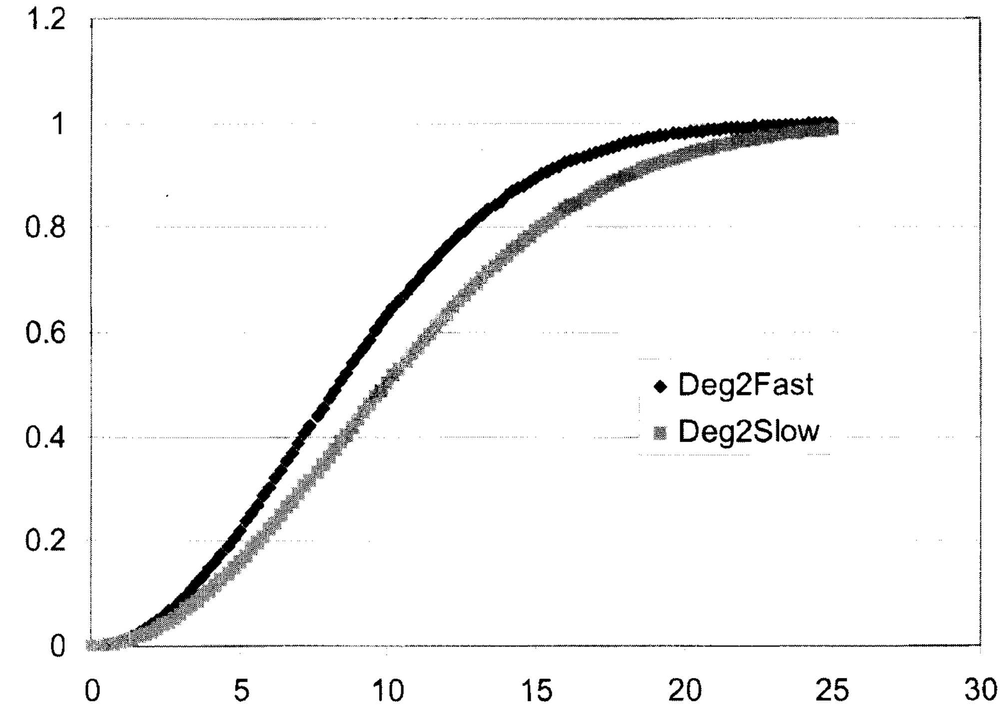

Figure 3 Learning curves modeled by $Y = 1-e^{-(aX)^2}$. In the upper curve a = 0.1 and in the lower curve a = 0.07.

Curves like Figure 3 could be modeled by a second-degree sigmoid equation:

$$Y = 1-e^{-(aX)^2}$$

where Y is a measure of learning or development

X is a measure of time

a indicates the speed the curve approaches maximum

According to Zajonc and Markus (1975), Equation 2 could be used to model intellectual development. For spatial intellectual development, the value of coefficient a = 0.1 whereas for verbal intellectual development, a = 0.07 which indicates a slower development.

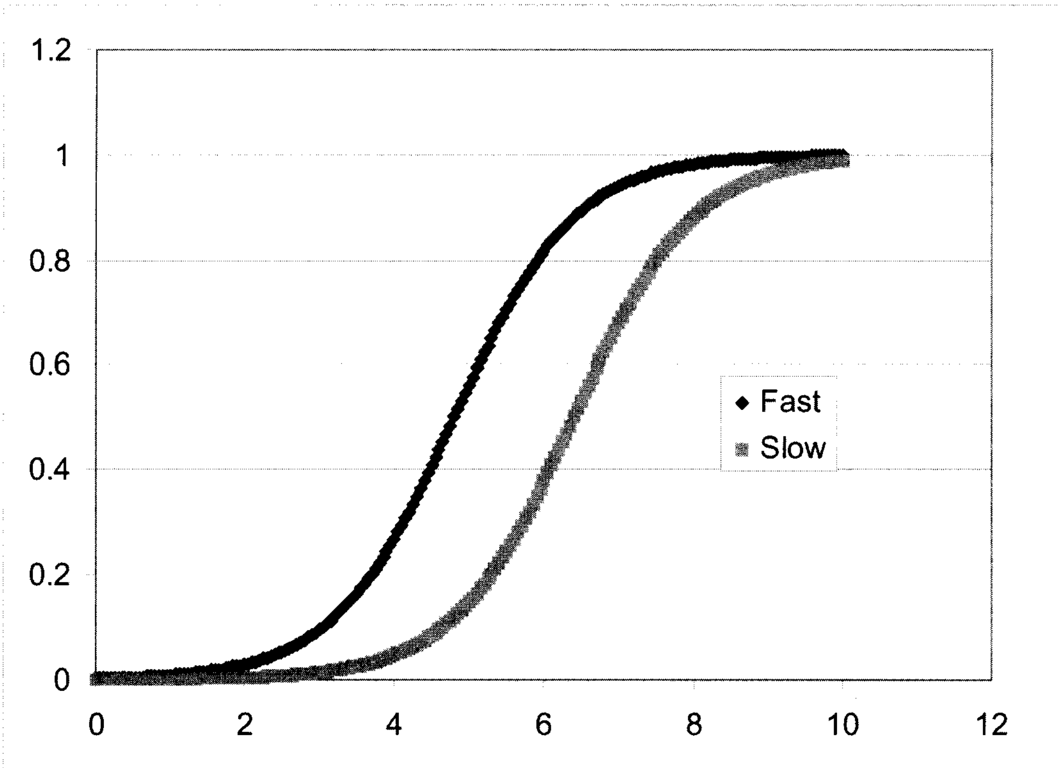

Figure 4 Learning curves modeled by $Y = \frac{1}{1+e^{-aX+b}}$ In the upper curve a = 1.25 and b = 6. In the lower curve a = 1.25 and b = 8.

Curves like Figure 4 could be represented by IRT models that provide an S-shaped ogive (see Embretson & Reise, 2000, p. 298). One such model is a logistic function shown in Equation 3:

$$Y = \frac{1}{1+e^{-aX+b}} \tag{3}$$

where Y is a measure of learning or development

X is a measure of time

a indicates the speed of the curve as it rises

b indicates the position of the S-shape along the X axis

Implications of the Learning Curve Models

The first implication of the learning curve models shown above is on the pattern of standard deviations from learning or developmental measures on a group of learners. All models of learning curves predict that the standard deviation of learning or skill development in a group growth study should be small at first. Then it will become larger and finally it will be smaller again. Deno, Marston and Mirkin (1982) studied children from 3rd grade to 6th grade using several writing skill measures such as total words written (in a period of fixed time), mature words, words spelled correctly, and letter sequence correct. The standard deviations for these measures were smaller in the 3rd grade, larger in the 4th and 5th grades, and in the great majority of cases smaller again in the 6th grade. The tendency towards this pattern could also be seen in Malecki and Jewell (2003, p.384) on writing skills measures such as total words written, words spelled correctly, and correct writing sequence. Finally, this small-large-small pattern of standard deviations was also found in Kirby, Parrila and Pfeiffer (2003, p.456) in their study of "work attack" and "word identification" skill development among kindergartens through 5th-graders.

The second implication of the learning curve models is that the practice of comparing student performance at two points in time—such as the popular pretest and posttest will not be able to detect if the performance is increasing in acceleration (for example, in the first half of the S-shape curve) or decreasing in acceleration (for example, in the second half of the S-shape curve). To detect a curve, at the very least, measuring the performance at three different points in is required. This point was succinctly emphasized in Rogosa (1982, p. 741).

The third implication follows directly from the first and second implications. In testing the significant difference over time of three or more performance averages as in the One-Way Analysis of Variance, homogeneity of variances has to be assumed (see Freund, 1971, p. 396; Hays , 1973, p. 482). However, this assumption will be violated as evident in the first implication discussed above. It has been suggested that the violation of this assumption would be less serious if the sample sizes are kept more or less equal (see, for example, Hays, 1973, p. 482). Therefore, care needs to be particularly exercised on equal-size samples when one performs ANOVA of this sort.

The fourth and probably most important implication of learning curves is on measuring gains at various intervals in time. If learning or development Y were a linear function of time X, the use of $Y_2 - Y_1$ as the indicator of gain at every time interval would have been OK. However, learning or development Y is NOT a linear function of time X. The function is curvilinear (most likely S-shaped) as shown in the introduction of this paper. If we take gain $Y_2 - Y_1$ at every equal interval from the beginning to the end of the curve, we will notice that the gain becomes smaller and smaller towards the end of the curve. This means accomplished learner will have difficulty making gain! Bereiter (1963 as cited in Embretson & Reise, 2000) emphasized this point as one of the many problems of the traditional gain score $Y_2 - Y_1$. Should we not then give higher and higher weights to gains made towards the end of the learning curve? In many Olympic competitions higher level of difficulty receives more weight in scoring.

Improving the Traditional Gain Score

Problems of traditional gain score ($Y_2 - Y_1$) were noted in many studies (see, e.g., Bereiter, 1963; Cronbach and Furby, 1970; Rogosa, Brandt & Zimowski, 1982; Embretson & Riese, 2000). Several improvements (for example, residual change measures) have been suggested but remained unsatisfactory (see Rogosa, Brandt & Zimowski, 1982; Embretson & Riese, 2000, p. 296). A more promising recent attempt to weight a traditional gain score throughout the range of the learning curve was proposed by Kanjanawasee (1989). The weighted gain score is known as the "Relative Gain Score" which is defined as:

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

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 a much improvement over the traditional gain score considering the nature of the learning curve. Relative gain score is applicable to both a single learner (corresponding to only one learning curve shown in Figure 2, 3 or 4), or a group of learners (corresponding to both learning curves shown in Figure 2, 3 or 4 as well as other curves imagined between the two curves). In the case of a group of learners we will be referring to gain score as $M_2 - M_1$ (Mean2 - Mean1, or *"group gain score"*) rather than $Y_2 - Y_1$. A "group learning curve" would be one imagined to be somewhere between the two curves shown in Figure 2, 3, or 4.

When considering a *group gain score*, in addition to weighting $M_2 - M_1$ by $F - M_1$ as in Kanjanawasee's Relative Gain Score, there is a competing index proposed by Glass, McGaw and Smith (1981) known as "Effect Size" (ES) that we could adapt for use in measuring gain. An effect size is defined as "the mean of the experimental group minus the mean of the control group" divided by "the standard deviation of the control group" (see Light & Pillemer, 1984). When adapting this for use in group gain score, we can define the "Gain Size" (GS) as "post-mean minus pre-mean" divided by the standard deviation of the pre-mean or $(M_2 - M_1)/SD_1$. Note that when Gain Size is taken towards the end of the learning curve (Figure 2, 3, or 4), SD_1 becomes smaller and smaller-thus, similar to Relative Gain Score, giving more weight to $M_2 - M_1$. Now, which is a better index to measure group growth, Relative Gain Score or Gain Size?

Figure 5 compares the graph of Relative Gain Score with the graph of Gain Size taken at small intervals throughout a learning curve. The Relative Gain Scores are plotted using the "medium" learning curve (see Figure 5). The Gain Sizes are plotted using the "medium" curve for $(M_2 - M_1)$s and the differences between the "fast" and "slow" learning curves (the range) in place of SD_1s (see Figure 5). The "range" is used in this plot instead of "standard deviation" because (a) real data are not used and therefore standard deviations are not available, and (b) the correlation between "range" and "standard deviation" is usually very high. Note from Figure 5 that (a) the Relative Gain Score "increases" slightly from interval to interval throughout the learning curve, but (b) the Gain Size, on the other hand, "decreases" from interval to interval with much faster rate at the beginning of the learning curve than towards the end of the curve. Towards the end of the learning curve, the Gain Size is essentially flat. The initial "steep" drop of the Gain Size graph lends inappropriate the use of Gain Size as an index to measure gain at the beginning of learning or development. This, together with the fact that Relative Gain Score can be used with both an "individual" learner and a "group" of learners but Gain Size can be used only with a group, makes Relative Gain Score a superior index between the two.

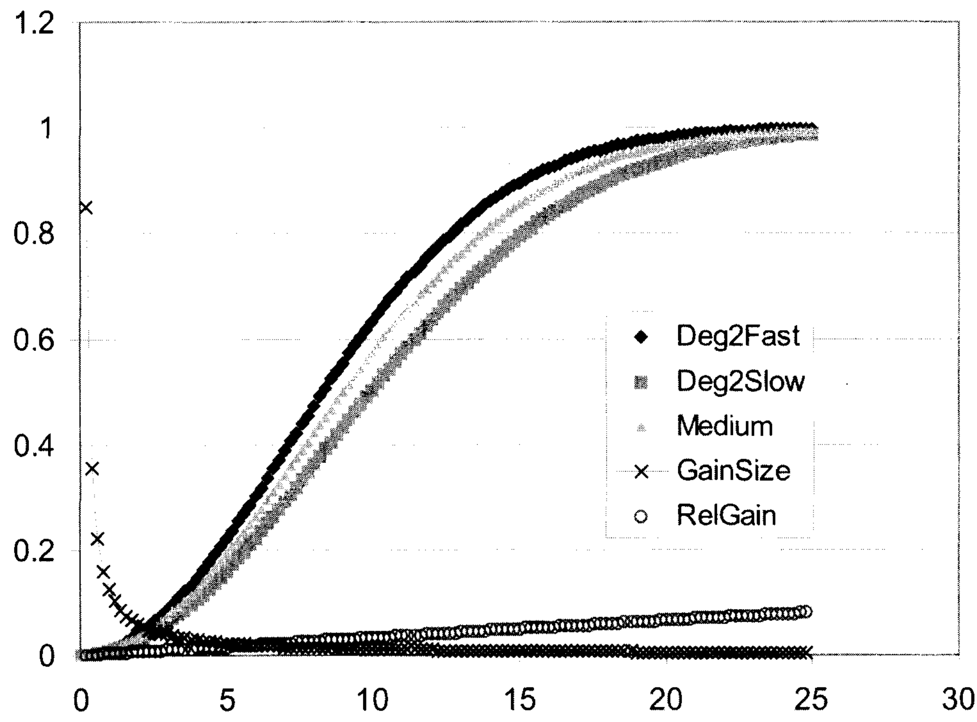

Figure 5 Comparing group Relative Gain Score and Gain Size at various small intervals throughout the medium learning curve.

Applying the Relative Gain Score and Gain Size to Real Data

The Relative Gain Score and Gain Size were applied to data from Darunsikkhalai School–the only full scale constuctionism school in Thailand that provides a totally project-based learning. A project period lasted nine weeks. Portfolios were used to assess the learning and development of the students. Four areas of outcomes were evaluated:

1. Mathematical skills. They included calculation and problem solving skill (CAL), as well as data presentation and analytical skill (DAT). Possible score = 3-12 each.

2. Thai language skills. They included listening skill (LIS), speaking skill (SPE), reading skill (REA), and writing skill (WRI). Possible score = 3-12 each.

3. Five disciplines. These were those proposed by Senge (1994, 2000) which included personal mastery (PM), mental model (MM), shared vision (SV), team learning (TL), and systems thinking (ST). Possible score = 1-4 each.

4. Four quotients. These were desirable personal qualities that the school conceptualized and believed consistent with modern Thai culture. They included emotional quotient (EQ), adversity quotient (AQ), technology quotient (TQ), and moral quotient (MQ). Possible score = 1-4 each.

The products in the students portfolios were evaluated at three points in time (once every three weeks) based on scoring rubrics. The results from a sample of 12 students ages 8-10 years (seven females and five males) were as shown in Table 1 and Table 2 with Relative Gain Scores and Gain Sizes included. Note that both the Gain Scores and Gain Sizes, in general, indicated a higher gain in the second period (Time 2-3) than the first period (Time 1-2). When the Relative Gain Scores were averaged by the academic achievement outcomes vs. desirable characteristics outcomes, it was found that gain in academic achievement was higher than gain in desirable characteristics at both the first period (24.27 vs. 14.29) and second period (42.02 vs. 33.27). Similarly, when Gain Sizes were averaged by the academic achievement outcomes vs. desirable characteristics outcomes, it was found that gain in academic achievement was also higher than gain in desirable characteristics at both the first period (0.77 vs. 0.49) and second

Table 1

The Pair-wise Comparisons and the Means of Relative Gain Scores of Student Academic (Math and Thai) and Non-Academic (Five Disciplines and Four Quotients) Outcomes (n=12)

Students' learning	1st–2nd		2nd–3rd		1st–3rd
	Mean Difference	*Relative Gain Score (%)*	*Mean Difference*	*Relative Gain Score (%)*	*Mean Difference*
1. Academic achievements (max=12)					
1.1 Mathematics					
1.1.1 calculating and problem solving skill (CAL)	-1.570**	35.39	-1.120**	35.57	-2.69**
1.1.2 data presentation and analytical skill (DAT)	-1.400**	27.05	-1.080**	39.46	-2.48**
1.2 Thai					
1.2.1 listening skill (LIS)	-0.687**	18.88	-1.897**	63.04	-2.58**
1.2.2 speaking skill (SPE)	-0.540	13.47	-1.723**	54.13	-2.26**
1.2.3 reading skill (REA)	-0.977**	22.45	-0.713*	21.83	-1.69**
1.2.4 writing skill (WRI)	-1.200*	29.00	-0.887**	37.07	-2.09**
2. Desirable characteristics (max=4)					
2.1 5 disciplines					
2.1.1 personal mastery (PM)	-0.157	7.15	-0.473**	48.60	-0.63**
2.1.2 mental model (MM)	-0.213	15.73	-0.247	24.00	-0.46**
2.1.3 shared vision (SV)	-0.353*	19.46	-0.283**	24.69	-0.64**
2.1.4 team learning (TL)	-0.240	13.75	-0.330**	36.20	-0.57**
2.1.5 systems thinking (ST)	-0.373**	23.73	-0.313*	27.53	-0.69**
2.2 4 quotients					
2.2.1 emotional quotient (EQ)	-0.055	7.10	-0.235*	24.83	-0.29**
2.2.2 adversity quotient (AQ)	-0.228**	17.98	-0.183*	19.42	-0.14**
2.2.3 technology quotient (TQ)	-0.093	10.98	-0.280**	52.75	-0.37**
2.2.4 moral quotient (MQ)	-0.283*	12.75	-0.447**	41.75	-0.73**
Average		17.24		34.60	

* $p < .05$; ** $p < .01$

Table 2

Pair-wise Comparisons (Time 1-2 and Time 2-3) and Gain Sizes of Student Academic (Math and Thai) and Non-Academic (Five Disciplines and Four Quotients) Outcomes (n=12)

Students' learning	1st-2nd		2nd-3rd		1st-3rd
	Mean Difference	*Relative Gain Score (%)*	*Mean Difference*	*Relative Gain Score (%)*	*Mean Difference*
1. Academic achievements (max=12)					
1.1 Mathematics					
1.1.1 calculating and problem solving skill (CAL)	-1.570**	0.84	-1.120**	0.68	-2.69**
1.1.2 data presentation and analytical skill (DAT)	-1.400**	0.74	-1.080**	0.90	-2.48**
1.2 Thai					
1.2.1 listening skill (LIS)	-0.687**	1.13	-1.897**	2.02	-2.58**
1.2.2 speaking skill (SPE)	-0.540	0.55	-1.723**	1.78	-2.26**
1.2.3 reading skill (REA)	-0.977**	0.71	-0.713*	0.50	-1.69**
1.2.4 writing skill (WRI)	-1.200*	0.68	-0.887**	0.98	-2.09**
2. Desirable characteristics (max=4)					
2.1 5 disciplines					
2.1.1 personal mastery (PM)	-0.157	0.31	-0.473**	1.18	-0.63**
2.1.2 mental model (MM)	-0.213	0.53	-0.247	0.77	-0.46**
2.1.3 shared vision (SV)	-0.353*	0.61	-0.283**	0.79	-0.64**
2.1.4 team learning (TL)	-0.240	0.39	-0.330**	0.62	-0.57**
2.1.5 systems thinking (ST)	-0.373**	0.79	-0.313*	0.89	-0.69**
2.2 4 quotients					
2.2.1 emotional quotient (EQ)	-0.055	0.15	-0.235**	0.49	-0.29**
2.2.2 adversity quotient (AQ)	-0.228**	0.85	-0.183*	0.86	-0.14**
2.2.3 technology quotient (TQ)	-0.093	0.25	-0.280**	0.82	-0.37**
2.2.4 moral quotient (MQ)	-0.283*	0.51	-0.447**	1.50	-0.73**
Average		0.60		0.99	

* $p < .05$; ** $p < .01$

period (1.14 vs. 0.88). Thus the use of Relative Gain Score and Gain Size produced similar results in practice despite some differences in theory. Note another advantage of using either the Relative Gain Scores or Gain Sizes, i.e., they both allow direct comparison of measures that have different maximum scores (an academic achievement has a maximum score of 12 while a desirable characteristic has a maximum score of only 4). This is because both indexes are in terms of common units-Relative Gain Score is in percentage gain, and Gain Size is in standardized unit.

The results of data analysis at Darunsikkhalai School were presented in details elsewhere (see Tangdhanakanond, Pitiyanuwat & Archwamety, 2005)

Further Improvement of Relative Gain Score?

The previous section commented on the superiority of Relative Gain Score over Gain Size. Could the Relative Gain Score be improved? There are two observations.

1. As it stands now, a Relative Gain Score $\frac{Y_2 - Y_1}{F - Y_1} \times 100$ over a short interval and a Relative Gain Score $\frac{Y_3 - Y_1}{F - Y_1} \times 100$ over a longer interval with the same origin Y_1 will have the same denominator $F - Y_1$. However, Y3 is closer to the full or maximum score than Y2. Should the denominator be a little bit smaller? A more "balanced" relative gain score defined as:

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

could be an interesting concept for future studies.

2. As shown in Figure 5, as we move from the beginning towards the end of the learning/development curve, the Relative Gain Score gradually increases in size. Should the gain not be *constant*? An "*equalized*" gain score (the Equalizer) is another interesting concept to explore in future studies.

References

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3-20). Madison, WI: University of Wisconsin Press.

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.

Cronbach, L. J., & Furby, L. (1970). How should we measure change-Or should we? *Psychological Bulletin, 74*(1), 68-80.

Deno, S.L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children,* 48, 386-371.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates.

Freund, J. E. (1971). *Mathematical statistics.* Englewood Cliffs: NJ: Prentice-Hall.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis of social Research.* Beverly Hills, CA: Sage.

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

Hays, W. L. (1973). *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart and Winston.

Hilgard, E. R., & Bower, G. H. (1966). *Theories of learning.* New York: Appleton-Century-Crofts.

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

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

Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring Curriculum-Based Measurement writing probes. *Psychology in the Schools, 40*(4), 379-390.

Mayer, R. E. (1999). *The promise of educational psychology: Learning in the content area.* Upper Saddle River, NJ: Prentice-Hall.

Kanjanawasee, S. (1989). *Alternative strategies for policy analysis: An assessment of school effects on student's cognitive and affective mathematics outcomes in lower secondary school in Thailand.* Doctoral Dissertation. Education, Graduate School, University of California, Los Angeles.

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95*(3), 453-464.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research* Cambridge, MA: Harvard University Press.

Restle, F., & Greeno, J. G. (1970). *Introduction to mathematical psychology.* Reading, MA: Addison-Wesley.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726-748.

Senge, P. (1994). *The fifth discipline fieldbook.* New York: Doubleday Publishing Group.

Senge, P. (2000). A primer to five disciplines. In P. Senge (Ed), *Schools that learn* (pp. 59-100). New York: Doubleday Publishing Group.

Tangdhanakannond, K., Pitiyanuwat, S., & Archwamety, T. (2005). Constructionism: Student Learning and Development. *Academic Exchange Quarterly, 9*(3), 259-266.

Zajonc, R. B., & Markus, G. B. (1975). Birth order and intellectual development. *Psychological Review, 82*(1), 74-88.