Exceptional Children, Vol. 55, No. 1, pp. 38–44.

Measuring Progress on IEPs: A Comparison of Graphing Approaches

DOUGLAS MARSTON

ABSTRACT: The accuracy of predictions of future student performance on the basis of graphing data on semi-logarithmic charts and equal interval graphs was examined. Predictions made for 83 students on the basis of reading and written expression data collected over 7 weeks were compared to actual data collected for weeks 8, 9, and 10. Analyses of deviations between predictions and actual scores indicated that predictions were more accurate when the data had been graphed on equal interval graphs. Implications for training are discussed.

☐ Implementation of Public Law 94-142, the Education for All Handicapped Children Act, requires that an Individualized Education Plan (IEP) be written for all handicapped students receiving special education services. In addition, the same legislation mandates the use of fair and nondiscriminatory assessment practices in monitoring student progress toward IEP objectives. The use of traditional achievement and intelligence tests for such purposes, however, may be a tenuous exercise. Salvia and Ysseldyke (1988) warn that many tests lack evidence of validity or reliability. Some researchers point out that achievement tests differentially sample student curricula and therefore provide educators with questionable data about actual student performance (Good & Salvia, 1988; Jenkins & Pany, 1978). And finally, norm-referenced tests are not adequately designed to measure pupil progress (Carver, 1974; Hively & Reynolds, 1975).

An alternative assessment strategy for measuring pupil progress is the use of repeated measurement and time series analysis of the student's academic skills (Fuchs, Deno, & Mirkin, 1984; Marston, 1988). This methodology has been outlined in several approaches to delivering special education services: Precision Teaching (Lindsley, 1971), Exceptional Teaching (White & Haring, 1980), and Data-Based Program Modification (Deno & Mirkin, 1977). Common to all three models is the frequent measurement of student skills on various academic tasks related to IEP goals. The collected data typically are plotted on graph paper and the results subjected to a time series analysis. In this way, student progress toward IEP goals and the effectiveness of instructional strategies may be evaluated. Fuchs and Fuchs (1986) determined that systematic monitoring over time produced an average gain of .7 standard deviation for monitored students compared to non-monitored students.

While there are some similarities among these repeated measurement models, there is an important difference involving the type of graph to be used in charting student data. Proponents of the Precision Teaching and Exceptional Teaching models advocate the use of the Standard Behavior Chart (Pennypacker, Koenig, & Lindsley, 1972), a graph on which variables measured along the ordinate (vertical axis) are recorded on a logarithmic scale. Those favoring this semi-logarithmic chart claim improvement in academic performance is proportional, not arithmetical, and is measured and predicted best on the logarithmic scale (Howell, Kaplan, & O'Connell, 1979). The implication for the exceptional student who may initially acquire academic skills at a slow

DOUGLAS MARSTON *is School Psychologist and Project Director of the Formative Evaluation Research Project, Department of Special Education, Minneapolis Public Schools, Minnesota.*

rate is illustrated by White and Haring (1980). These authors noted that graphing typical student performance data on an equal interval chart may be misleading since initially progress is slow. Using the equal interval chart, a successful instructional plan might be abandoned because of lack of improvement, whereas the same set of student data on the Standard Behavior Chart will appear more orderly in its display of growth.

However, many proponents of time series analysis indicate that equal interval graphs serve educational needs just as well. Most graphing procedures used in major texts emphasizing time series analysis in education employ equal interval graphs (Glass, Willson, & Gottman, 1975; Kratochwill, 1978; O'Leary & O'Leary, 1972; Sulzer-Azaroff & Mayer, 1977). The dilemma facing practitioners is choosing the more technically adequate graphing procedure.

Brandstetter and Merz (1978) addressed the issue of the efficacy of both graphing procedures. They found higher rates of student achievement when data were charted on the two types of graphs compared to not charting data. However, their research did not directly compare the semi-logarithmic chart with the equal interval graph. The question of which graph to use is not a trivial one. Test standards developed by the American Psychological Association, American Educational Research Association, and the National Council for Educational Measurement (APA, 1985) require that scaling methods used in assessment procedures be technically adequate. Users of the Standard Behavior Chart and equal interval graph must attend to these recommendations, for they are integral to assessment procedures currently recommended for exceptional students.

PURPOSE

The research presented here compares the two approaches. Since proponents of the Standard Behavior Chart maintain that a significant characteristic of the semi-logarithmic chart is the ability to better predict student performance, we have focused on an analysis of the accuracy of the different predictions generated by each type of chart. Each chart produces different predictions. As may be seen in Figure 1, the projections of student performance from the same set of data on the two types of charts are very different. For example, using identical data sets of 10 weekly points, the prediction of performance by the 20th week is 48 words correct for the linear model and 60 words correct for the logarithmic model. The research presented here examines efficacy of the linear model (Equal Interval Graph) and the logarithmic model (Standard Behavior Chart).

METHOD

Design

Student performance on direct, repeated measures of reading and written expression were collected weekly over a 2½-month period for 83 low-achieving elementary students. Using a computer program to simulate charting on both equal interval and semi-logarithmic graphs, each student's data were entered into the computer at the end of the 7th week. Regression equations for each set of student data for both charts were calculated. The slope of each student's performance on both the semi-logarithmic and equal interval chart then was used to predict student performance at weeks 8, 9, and 10 of the data collection period. The estimates of student performance at these times were contrasted with the actual data collected at weeks 8, 9, and 10 by determining the absolute deviation between the scores.

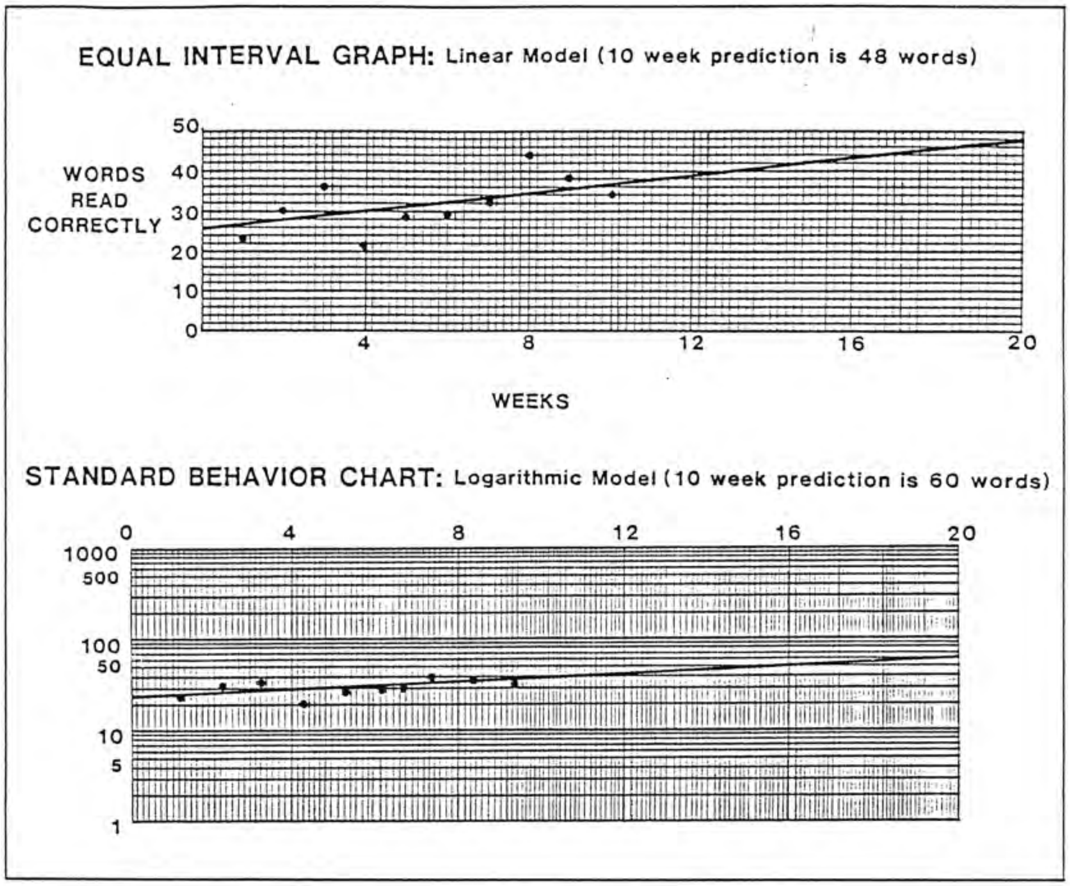
The size of the deviation scores for the semi-logarithmic chart (logarithmic model) was then compared to the magnitude of the deviations on the equal interval graph (equal interval model) for each student on each measure with a paired t test analysis. On those comparisons where significant differences were found, the graphing approach with the smaller average deviation score was considered to be the one making better predictions of student performance.

Subjects

Selection of this low-achieving population resulted from the screening of all 785 elementary students from grades 3 to 6 enrolled in three elementary schools. The schools were located within communities in rural settings, yet each was within 50 miles of the metropolitan Twin City area. According to 1970 Census figures, the three communities ranged in size from 2,281 to 8,876.

Screening procedures involved a short duration measure of written expression that significantly discriminated LD and non-LD students (Deno, Marston, & Mirkin, 1982). Each student was administered a story starter and asked to write a composition. For each student, the number of words written in the composition (Total Words Written) was computed. Students who had no history of special education services and scored at or below the 15th percentile were asked to participate in the repeated, direct measurement phase of the study.

FIGURE 1
Predictions of Student Performance Using Linear and Logarithmic Models



Parental permission was received for 83 students. Twenty-six of the students were third graders, 17 were fourth graders, 19 were fifth graders, and 21 were sixth graders. Thirty-two of the low-achieving population were females.

Procedures

All 83 students were administered short duration measures designed for direct, repeated measurement of reading (Deno, Mirkin, & Chiang, 1982) and written expression (Deno, Marston, & Mirkin, 1982) on a weekly basis for 10 weeks.

Reading. Lists of words selected randomly from the third-grade level of the Harris-Jacobson Word List (Harris & Jacobson, 1972) were used for the reading tasks; these were administered each week. For each list, the student was asked to read aloud for one minute. Test instructions read verbatim to the subject were:

> Here is a word list that I want you to read. When I tell you to start, you can read across the page. Please read as fast and accurately as you can. If you come to a word that you don't know, move on to the next one. I will tell you when to stop reading. Are there any questions? Ready? Begin.

The child then was timed for 60 seconds while the teacher followed along and recorded mistakes on a sheet identical to the one from which the student read. If a student did not respond after approximately 6 seconds, he or she was told to move on to the next word. At the end of the timing, recording sheets were

TABLE 1
Comparison of Linear and Logarithmic Models in Predicting Performance at Week 8

Measures	*Sample Size*	*Type of Graph*	*Mean Deviation*	*Standard Deviation*	*T-value*	*Probability*
		Reading Measures				
Words Read Correctly	75	EQ Interval	9.55	7.6	1.67	.100
3rd Grade Level		Logarithmic	9.84	8.3		
Words Read Incorrectly	75	EQ Interval	2.87	1.1	.83	.409
3rd Grade Level		Logarithmic	3.24	3.8		
Words Read Correctly	55	EQ Interval	5.99	4.7	.29	.772
Grade Level		Logarithmic	5.97	4.6		
Words Read Incorrectly	54	EQ Interval	2.77	2.6	1.62	.111
Grade Level		Logarithmic	3.12	3.1		
		Written Expression Measures				
Total Words Written	75	EQ Interval	6.90	5.2	1.35	.183
		Logarithmic	7.17	5.7		
Words Written Correctly	75	EQ Interval	6.49	5.3	1.20	.232
		Logarithmic	7.05	6.6		
Words Written Incorrectly	75	EQ Interval	1.81	1.5	2.45	.016
		Logarithmic	2.42	2.7		
Correct Letter Sequences	75	EQ Interval	28.38	21.3	1.60	.115
for Writing Task		Logarithmic	29.89	23.8		

collected and later scored by trained judges. For each list, the numbers of Words Read Correctly (WRC) and Words Read Incorrectly (WRI) were scored. Estimates of interrater agreement ranged from .94 to .98.

In addition to reading the third-grade lists, the fourth, fifth, and sixth graders were asked to read a list of words selected from their grade level from the Harris-Jacobson list. For example, each week the fifth graders read both a third-grade list and a fifth-grade list. For each of these lists, the number of Words Read Correctly from Grade Level (WRCG) and the number of Words Read Incorrectly from Grade Level (WRIG) were counted.

Written expression. Story starters were used to obtain weekly writing samples from the 83 students. Directions to the students were:

> I want you to write a story. I am going to read a sentence to you first, and then I want you to write a short story about what happens. You will have a minute to think about a story to write and then you will have three minutes to write it. When I say 'please start writing' you may begin.

Students' responses to each story starter were scored by a trained judge. The compositions were scored for Total Words Written, Words Written Correctly, Words Written Incorrectly, and Correct Letter Sequences Written (White & Haring, 1980). Interrater agreement was .87.

RESULTS

On all *t* test comparisons, the .05 level of probability was used as the criterion level for significance. In the first analysis, the linear (equal interval) and logarithmic models were compared on predictions of student performance at week 8 based upon the slope of the first seven weekly measurements. As may be seen in Table 1, only one of the contrasts was significant; the difference favored the linear model in measuring Words Written Incorrectly.

TABLE 2
Comparison of Linear and Logarithmic Models in Predicting Performance at Week 9

Measures	*Sample Size*	*Type of Graph*	*Mean Deviation*	*Standard Deviation*	*T-value*	*Probability*
		Reading Measures				
Words Read Correctly	76	EQ Interval	9.69	9.0	1.96	.054
3rd Grade Level		Logarithmic	10.15	10.2		
Words Read Incorrectly	76	EQ Interval	2.81	1.3	2.52	.014
3rd Grade Level		Logarithmic	5.36	8.9		
Words Read Correctly	54	EQ Interval	7.30	7.3	1.25	.217
Grade Level		Logarithmic	7.13	6.9		
Words Read Incorrectly	54	EQ Interval	3.14	2.9	1.86	.069
Grade Level		Logarithmic	3.90	4.7		
		Written Expression Measures				
Total Words Written	78	EQ Interval	8.62	7.0	2.58	.012
		Logarithmic	9.45	8.2		
Words Written Correctly	78	EQ Interval	8.23	6.9	1.58	.119
		Logarithmic	9.85	12.1		
Words Written Incorrectly	78	EQ Interval	1.94	1.9	2.04	.045
		Logarithmic	3.19	5.7		
Correct Letter Sequences	78	EQ Interval	36.12	30.1	2.95	.004
for Writing Task		Logarithmic	40.51	35.2		

Predicting week 9 performance from the slope of the first 7 weeks was the focus of the second analysis. The results from this analysis are presented in Table 2. Four of eight comparisons (Words Read Incorrectly, Total Words Written, Words Written Incorrectly, and Correct Letter Sequences) were significant; all differences favored the linear model which exhibited smaller deviations between predicted and actual scores. Two other contrasts, which also displayed lower deviations for equal interval graphs, approached significance (Words Read Correctly, $p = .054$; Words Read Incorrectly on Grade Level Material, $p = .069$).

Week 10 estimates based upon the 7-week slope were examined in the third analysis. As can be seen in Table 3, four of the eight contrasts were significant at the .05 level; again, these favored the linear model. The comparison for Words Read Incorrectly on Grade Level Material approached significance ($p = .056$), with smaller deviations demonstrated on equal interval graphs.

DISCUSSION

The research described here focused on only one aspect of graphing, the use of time series data to predict future performance. Yet, this in itself is quite significant when the writing of IEP goals is considered. White and Haring (1980) proposed that an analysis of the slope of student data is useful in producing goals and objectives. Thus, generating predictions from time series data is helpful in the delivery of special services to the exceptional student. The question asked here is, which graph should be used to chart student performance?

The research indicates that predictions of student performance for weeks 8, 9, and 10, based on 7 weeks of data in reading and written expression, are more accurate when data are graphed on the equal interval chart. In no cases did comparisons showing a significant difference favor the Standard Behavior Chart.

The most important implication of this research is for training. Often educators resist using the

TABLE 3
Comparison of Linear and Logarithmic Models in Predicting Performance at Week 10

Measures	*Sample Size*	*Type of Graph*	*Mean Deviation*	*Standard Deviation*	*T-value*	*Probability*
		Reading Measures				
Words Read Correctly	79	EQ Interval	13.56	9.8	1.64	.105
3rd Grade Level		Logarithmic	14.05	10.6		
Words Read Incorrectly	79	EQ Interval	3.43	1.3	2.12	.037
3rd Grade Level		Logarithmic	8.50	21.3		
Words Read Correctly	54	EQ Interval	10.11	9.5	1.67	.101
Grade Level		Logarithmic	9.82	9.2		
Words Read Incorrectly	54	EQ Interval	4.37	5.1	1.96	.056
Grade Level		Logarithmic	5.80	7.7		
		Written Expression Measures				
Total Words Written	79	EQ Interval	9.87	8.6	2.95	.004
		Logarithmic	11.29	9.2		
Words Written Correctly	79	EQ Interval	9.31	8.3	1.44	.153
		Logarithmic	12.53	21.4		
Words Written Incorrectly	79	EQ Interval	1.75	1.6	2.21	.030
		Logarithmic	3.26	6.2		
Correct Letter Sequences	79	EQ Interval	43.03	34.6	2.19	.032
for Writing Task		Logarithmic	48.27	40.9		

Standard Behavior Chart for reasons such as "it's overly complex" or "I can't explain it to parents." In many instances, those discouraged by the semi-logarithmic graph but interested in graphing are more willing to use equal interval graphs. If increasing the likelihood that special educators will use repeated measurement strategies in their educational planning, interventions, and assessments is a function of the type of graph they prefer (usually the equal interval chart), it would appear the research presented here provides an empirical basis for making that choice.

REFERENCES

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.

Brandstetter, G., & Merz, C. (1978). Charting scores in precision teaching for skill acquisition. *Exceptional Children, 45*(1), 42-48.

Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist, 29*, 512-518.

Deno, S. L., Marston, D., & Mirkin, P. K. (1982). Identifying valid measures of written expression for use in continuous evaluation of educational programs. *Exceptional Children, 48*, 368-371.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: The Council for Exceptional Children.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36–45.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449-460.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*, 199-208.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). *Design and analysis of time-series experiments*. Boulder, CO: University of Colorado Press.