DOCUMENT RESUME

ED 310 560 EC 220 764

AUTHOR Munger, Gail F.; And Others
TITLE How Often Do You Need To Collect Student Performance Data? A Study of the Effects of Frequency of Probe Data Collection and Graph Characteristics on Teachers' Visual Inference.
PUB DATE 88
NOTE 18p.; In: Baine, David, Ed.; And Others. Alternative Futures for the Education of Students with Severe Disabilities (Edmonton, Canada, May 6-8, 1987). For proceedings, see EC 220 763; for other selected individual papers, see EC 220 765-782.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Decision Making; Diagnostic Teaching; Evaluation Methods; *Graphs; Recordkeeping; Severe Disabilities; *Special Education Teachers; *Student Evaluation; *Teacher Attitudes; *Trend Analysis
IDENTIFIERS *Data Based Program Development; *Frequency of Measurement

ABSTRACT

This study addressed the question of whether teachers' judgments differ according to the frequency of data collection on student performance and the type of trend. The study also investigated whether teachers' judgments, based on different types of graphs (ascending, descending, flat, or variable), vary with the frequency of data collection. A set of 16 graphs of actual student performance data was analyzed by 59 teachers of students with moderate to profound handicaps. Results indicated that, when asked to evaluate student performance, teachers' judgments tended to be consistent and accurate when the graphed data represented continuous and systematic improvement in performance. However, when the data represented a decrease in performance, no change, or highly variable performance, teachers' judgments tended to differ according to the frequency of data collection. When asked to make program recommendations, teachers' judgments tended to differ according to the frequency of data collection for all types of trends. (Author/DB)

# How Often Do You Need to Collect Student Performance Data? A Study of the Effects of Frequency of Probe Data Collection and Graph Characteristics on Teachers' Visual Inference



Gail F. Munger
Martha E. Snell
Brenda H. Loyd
*University of Virginia*

*Abstract*

*Teachers often rely on visual analysis of graphed student performance data to evaluate progress and to make program decisions. However, because collecting data can be time consuming and interfere with instruction, teachers would like to know how much data is necessary to make reliable judgments. To investigate the effect of frequency of data collection on teachers' judgments and decisions, this study addressed the question of whether teachers' judgments differ according to frequency of data collection and type of trend. The study also investigated whether teachers' judgments, based on different types of graphs (ascending, descending, flat or variable) vary with the frequency of data collection. A set of 16 graphs of actual student performance data was analyzed by 59 teachers of students with moderate to profound handicaps. The resulting data were analyzed by a two-factor repeated measures design. The results indicated that when asked to evaluate student performance, teachers' judgments tended to be consistent and accurate when the graphed data represented continuous and systematic improvement in performance. However, when the data represented a decrease in performance, no change, or highly variable performance, teacher judgments tended to differ according to the frequency of data collection. When asked to make program recommendations, teachers' judgments tended to differ according to the frequency of data collection for all types of trends.*

Teachers often rely on their visual analysis skills to read and interpret graphs of student performance data and to monitor the effects of instructional programs. Based on such analysis, teachers may decide whether to change an intervention program or determine what changes are most likely to improve the students' performance of a target behavior.

There are, however, many factors that may distinguish teachers' visual analyses of data from the analyses performed by researchers and other professionals trained to read and interpret graphs. For example, teachers usually examine AB (baseline-intervention) data, whereas conventional visual analysis is customarily taught and practiced using single-subject designs (Parsonson & Baer, 1986). Even when teachers have received training in the visual analysis of data, the training has generally not included the interpretation of single-subject designs. Furthermore, teachers must often collect and analyze data under significant time pressures amid the general confusion common to many classrooms.

On the other hand, a teacher's ongoing participation in instruction and involvement with students is likely to produce additional clues regarding a particular subject's learning trend (Utley, Zigmond, & Strain, 1987) and may lead the teacher to discount or ignore data regarded as inaccurate (Grigg, 1986). Any of these factors may strongly differentiate a teacher's visual analysis of data from that of a researcher.

The literature on visual analysis has repeatedly demonstrated the problem of "interpretive inconsistency," regardless of who is examining graphed data (college students, teachers, researchers, behavioral journal reviewers, etc). There is substantial disagreement in the judgments made about the trend of the data viewed and about the functional relationship existing between the intervention and the target behavior. White (1971) found that teachers trained in visual inspection interpreted identical data differently to the extreme of disagreeing about whether graphs were ascending or descending. Jones, Weinrott and Vaught's (1978) findings showed that there was essentially no consensus regarding treatment effects among 11 skilled behavior analysts viewing published data from a respectable journal; their mean interjudge reliability coefficient was 0.39 with a range of 0.04 to 0.79.

Finally, DeProspero and Cohen (1979) obtained a modest, mean, interrater agreement correlation of 0.61 when reviewers of behavioral journals inspected graphs illustrating four influential factors: both pattern and degree of mean shift, within-phase variation, and trend. The four graphic factors studied appeared to influence judges interactively, not singly, emphasizing the complexity of visual analysis even under highly controlled conditions with trained judges.

For teachers to evaluate performance and to make appropriate program recommendations, it is necessary that they be able to analyze accurately whether a student's performance is improving, deteriorating, or remaining the same. In a study examining the effects of the form of data documentation and the type of trend on the ability of teachers to analyze the trend in frequency data, Utley et al. (1987) considered three trends (upward, level, and downward) and four levels of documentation, ranging from observation only to a combination of observation, raw data, graphs, and a six-day line of progress. Although the main effect for level of documentation and interaction between level of documentation and type of trend were found to be significant, the main effect for type of trend was not. All subjects were able to analyze upward trends accurately, but those in the "observation only" group were unable to analyze level and downward trends accurately. When any form of data was provided, the difference in accuracy across groups tended to be relatively small.

The findings of Utley et al. (1987) further confirm the necessity of collecting and analyzing data to evaluate student performance. However, these authors also found that as the amount of documentation increased, the accuracy of trend analysis did not increase concomitantly. This observation may suggest that further research is needed to determine whether sophisticated data analysis strategies do in fact improve the accuracy of trend analysis, and what effect frequency of data collection has on the accuracy and reliability of visual inference.

Teachers using visual analysis are likely to make more frequent errors if they have inadequate data on which to base a decision (Parsonson & Baer, 1986). Yet, it is far from clear how much data, probe or training, is necessary to make reliable judgments. The demands of teaching limit the amount of time available to all teachers for data collection, and when their students have moderate to profound disabilities, teachers have additional considerations. For example, the collection of training data, essential for making accurate day-to-day instructional decisions, may interfere with the use of "hands-on" systematic prompting procedures; the collection of probe data (under criterion conditions of no reinforcement or assistance) means a reduction of instructional time; the extension of baseline conditions to eliminate variability in a student's performance, or reversing to baseline conditions to demonstrate control, can result in a delay of treatment or a threat to improvement; and the collection of probe data in the community, where a majority of school-age instruction must take place to promote generalization, increases the number of potentially dangerous and stigmatizing situations the student experiences. These factors, which must be considered when teaching students with extensive disabilities, act to reduce the data available for analysis.

In a review of community-based research concerning students with severe disabilities, Snell and Browder (1986) found that when training was conducted daily, probe data were collected approximately once a week. So while these researchers examined both types of data to judge experimental effects, they generally had only one fifth the amount of probe data as training data.

Several studies have examined whether a reduced frequency of data collection yields adequate data for teachers to make consistent judgments about student progress or decisions about program changes. Bijou, Peterson, Harris, Allen, and Johnston (1969) studied the effects of varying the frequency of observations or data collection and found that sampling every other day beginning with the first session, sampling every other day beginning with the second session, and sampling every third day beginning with the first session yielded results that only slightly deviated from those attained when data were collected daily.

The effect of frequency of data collection and graph characteristics on visual inference was investigated by Munger and Loyd (1987). They reported that teachers tended to agree in their judgments regarding student progress and their decisions about program changes when performance data represented systematic improvement, but when graphed data represented a decrease in performance, no change, or highly variable performance, judgments tended to differ according to the frequency of data collection.

To investigate further the effect of frequency of data collection on decisions or judgments, this study replicated that of Munger and Loyd (1987) in addressing the questions of whether teachers would make similar decisions when student data was obtained each day, three times a week, twice a week, or once a week, whether different trends on graphs (ascending, descending, flat or variable) produced different judgments, and whether judgments based on different frequencies of data collection vary with the characteristics of data such as variation and trend.

## Method

### Graphs

To answer the research questions, four graphs of actual student acquisition data were selected from intervention programs of students with moderate to profound mental retardation. The graphs represented student performance of functional, multiple-step skills. The horizontal axis of each graph represented 60 days of data collection with baseline and intervention phases indicated. The vertical axis of each graph represented the percent of steps correctly performed by students during probe (test) observations of the target skills.

The four graphs were selected to illustrate four different trends: one graph showed an ascending trend (improvement in performance); one showed a descending trend (decline in performance); one graph represented neither an ascending nor descending trend but tended to be flat; and one represented neither an ascending nor descending trend but was variable, showing both advances and declines across the 60 days of data collection (see Figure 1). The trend of each graph was determined by statistical inference (testing for significant slope) and professional judgment.

Because teachers tend to change programs in which student performance is clearly decreasing, no graphs were located which represented a descending trend across 60 days of data collection. Therefore, a graph with 40 descending data points was selected to illustrate a descending trend; several nondescending data points were eliminated and additional descending points included in order to create a descending graph representing data collected across 60 days.

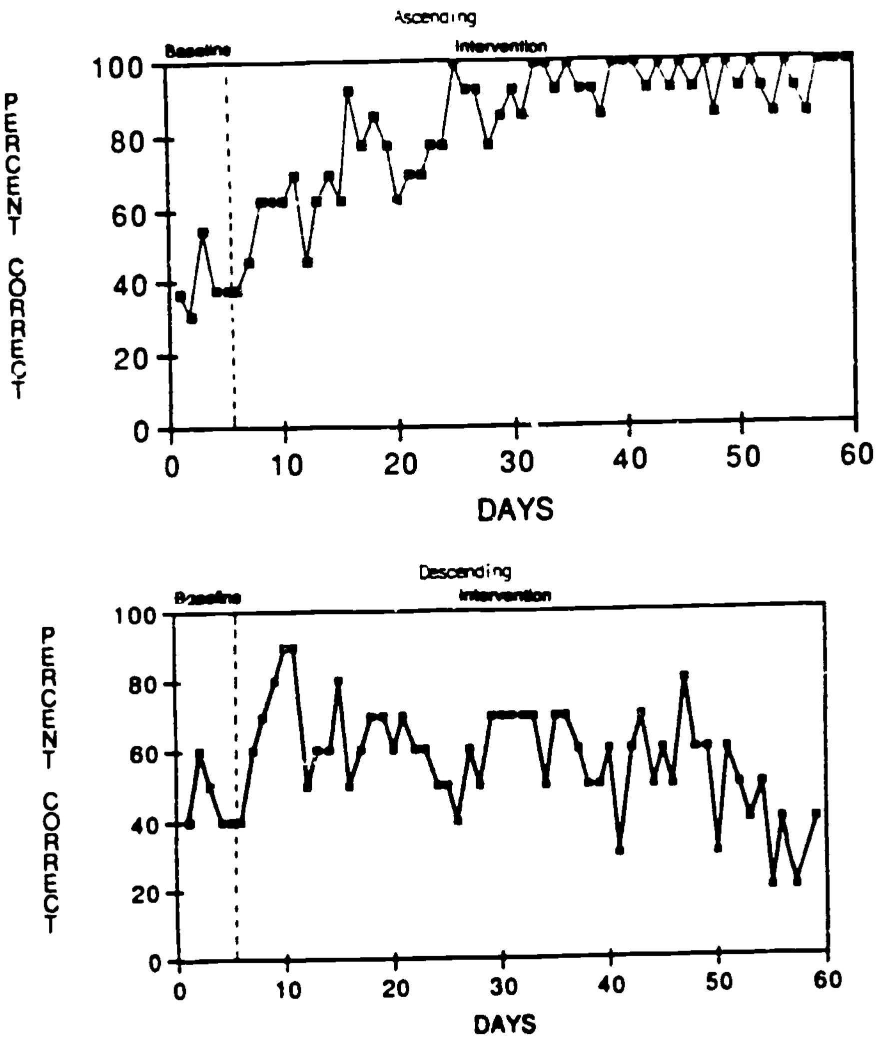


Figure 1: Graphs used to represent skill acquisition data collected five times per week and illustrating four trends: ascending, descending, flat, and variable

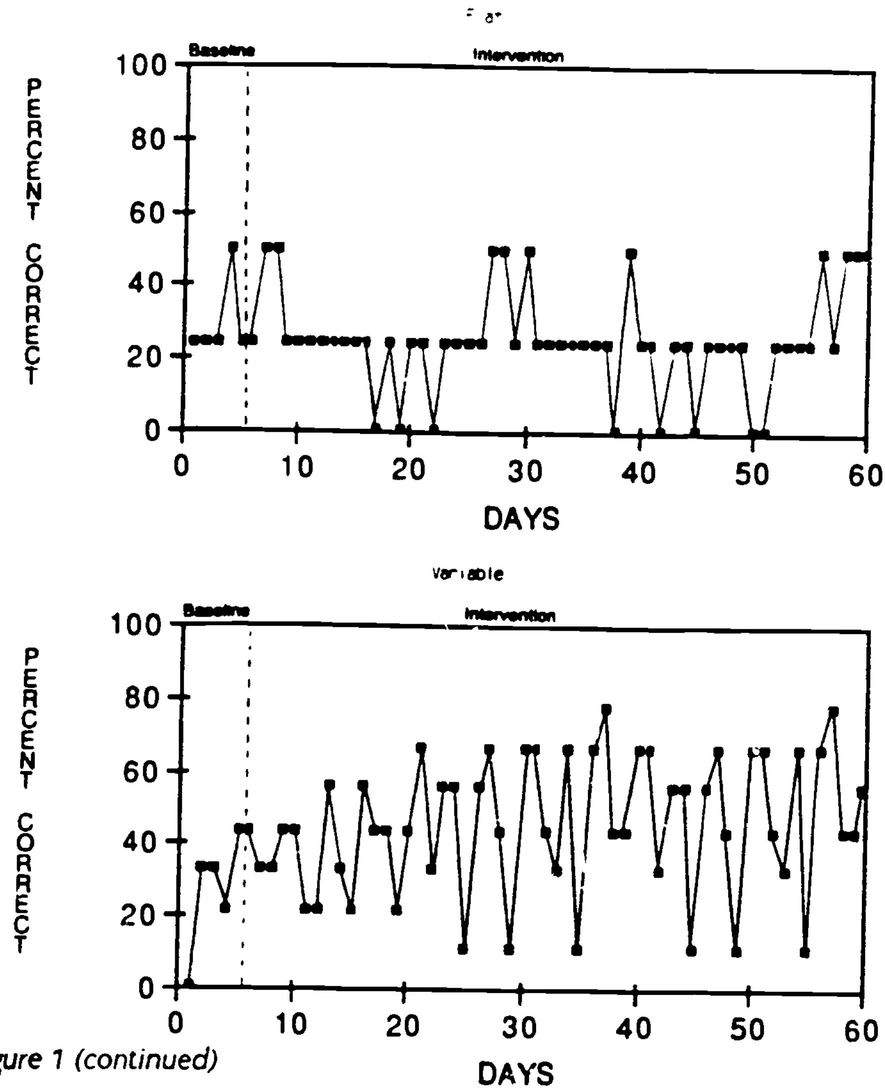


*Figure 1 (continued)*

From each of the original four graphs which represented data collected five times a week, three additional graphs were created to represent the sets of data as they would appear had the student performance data been collected three times a week, twice a week, and once a week. To create the 12 additional graphs, data points were selected as follows: to create the graphs representing data collected three times a week, only data collected on Mondays, Wednesdays, and Fridays across the 60 days were graphed; to create the graphs representing data collected twice a week, only data collected on Tuesdays and Thursdays across the 60 days were graphed; and to create the graphs representing data collected once a week, only data collected on Wednesdays were graphed. All graphs retained four days of baseline data.

The set of 16 graphs was arranged in a random sequence and analyzed by 59 randomly selected teachers of students with moderate to profound handicaps. The only information provided to the teachers was contained in the instructions which read as follows:

> These graphs represent actual performance data obtained from students with severe handicaps. Each graph summarizes three months of probe (not prompted or reinforced) sessions in skill acquisition programs. The vertical axis shows the percentage of steps performed correctly on a task-analyzed skill and the horizontal axis shows days on which the probe sessions were implemented.
>
> For each of the graphs that follows there are two questions. For the first question, check the statement that best describes the student performance represented by the graph. For the second question, check the statement that would most accurately reflect the program decision that you would make. Please make the best decision you can based on the information in the graph

## Teachers

The 59 teachers, employed by public school programs in eight states, were selected by seven university faculty members operating training programs for teachers of students with severe handicaps, and two directors of programs for students with severe handicaps. The Bachelor's degree was the highest degree of education attained by 66% of the teachers; the Master's degree was the highest degree of education attained by 31%; and 3% had completed the Education Specialist degree. Ninety-five percent of the teachers had received training in systematic instruction and data collection. Experience in teaching students with moderate to profound handicaps ranged from one to 19 years. The mean was 6.1 years. Eighty-eight percent of the teachers indicated that they collected training data daily; only 10% collected probe data daily.

For each graph, teachers were asked to evaluate the progress of the student by selecting one of five statements to describe the student's performance as represented by the graph:

1. definitely making progress;
2. probably making progress;
3. staying about the same;
4. probably decreasing in performance;
5. definitely decreasing in performance.

Teachers' judgments regarding student progress were assigned values from one to five respectively.

For each graph, teachers were also asked to make a program recommendation based on the student's performance as represented by the graph:

2. probably continue the program;
3. probably change the program;
4. definitely change the program.

Teachers' program decisions were assigned values from one to four respectively.

A two-factor, repeated measures design was used in the analysis of the data. The first factor was type of graph, with four levels: ascending, descending, flat, and variable. The second factor was frequency of data collection, with four levels: five, three, two, and one times per week. The dependent measures were: (1) student progress, as measured on a fi 'e-point scale from definitely making progress to definitely not making progress; and (2) program recommendation, measured on a four-point scale from definitely continue program to definitely change program.

The hypotheses of interest were whether different frequencies of data collection produced different teacher judgments and decisions, whether different trends on graphs produced different teacher judgments and decisions, and whether there was an interaction between type of graph and frequency of data collection.

## Results

The means of the teachers' ratings of student progress and program decisions for the four types of trends and four frequencies of data collection are presented in Table 1 and Figures 2 and 3 The group means of teachers' ratings for the four graphs depicting an increase in performance or ascending trend were 1.089 for student progress (1=definitely making progress) and 1.208 for program decisions (1=definitely continue the program). The group means of teachers' ratings for the four graphs with a downward or descending trend were 3.890 for student progress (4=probably decreasing in performance) and 3.474 for program decisions (4=probably change the program). The group means of teachers' ratings of the four graphs that were generally flat, depicting no change in performance across the 60 days, were 2.809 for student progress (3=staying about the same) and 3.152 for program decisions (3=probably change the program). The group means of teachers' ratings of the four variable graphs were 2.534 for student progress (3=staying about the same) and 2.847 for program decisions (3=probably change the program).

The group means of teachers' ratings of student progress based on performance data collected five times a week, three times a week, twice a week, and once a week also varied only slightly, ranging from 2.623 to 2.737.

The results of the two-factor analysis of variance procedure using type of graph and frequency of data collection as the independent variables and student progress ratings as the dependent measure are presented in Table 2. Main effects for type of graph and frequency of data collection and interaction effects were all statistically significant at the .05 level.

The results of the two-factor analysis of variance procedures using type of graph and frequency of data collection as the independent vari-

Table 1
Means of Ratings of Student Progress and Program Decisions by Type of Graph and Frequency of Data Collection

| Type of Graph | Teacher Decision: Student Progress (1-5 Scale) | Teacher Decision: Program Decision (1-4 Scale) |
|---|---|---|
| **Ascending** | | |
| 5 times/week | 1.034 | 1.085 |
| 3 times/week | 1.136 | 1.373 |
| 2 times/week | 1.017 | 1.068 |
| 1 time/week | 1.169 | 1.305 |
| **Descending** | | |
| 5 times/week | 4.119 | 3.542 |
| 3 times/week | 3.932 | 3.508 |
| 2 times/week | 4.356 | 3.627 |
| 1 time/week | 3.153 | 3.220 |
| **Flat** | | |
| 5 times/week | 3.068 | 3.559 |
| 3 times/week | 2.424 | 2.627 |
| 2 times/week | 3.186 | 3.525 |
| 1 time/week | 2.559 | 2.898 |
| **Variable** | | |
| 5 times/week | 2.373 | 2.576 |
| 3 times/week | 2.508 | 2.983 |
| 2 times/week | 2.441 | 2.729 |
| 1 time/week | 2.814 | 3.102 |

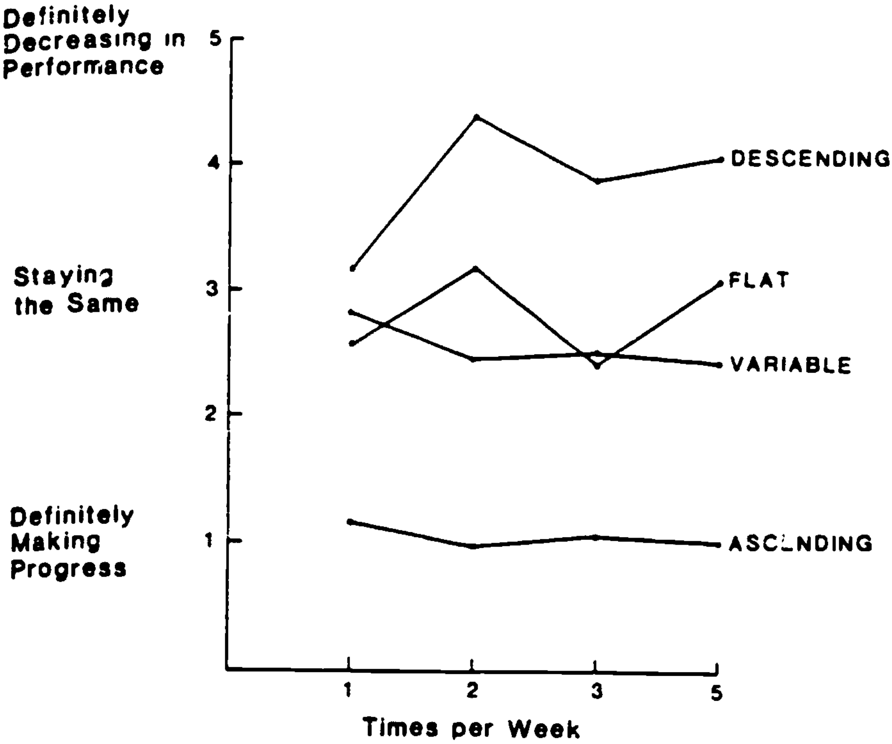


*Figure 2: Mean ratings of student progress by graph type and frequency of data collection*

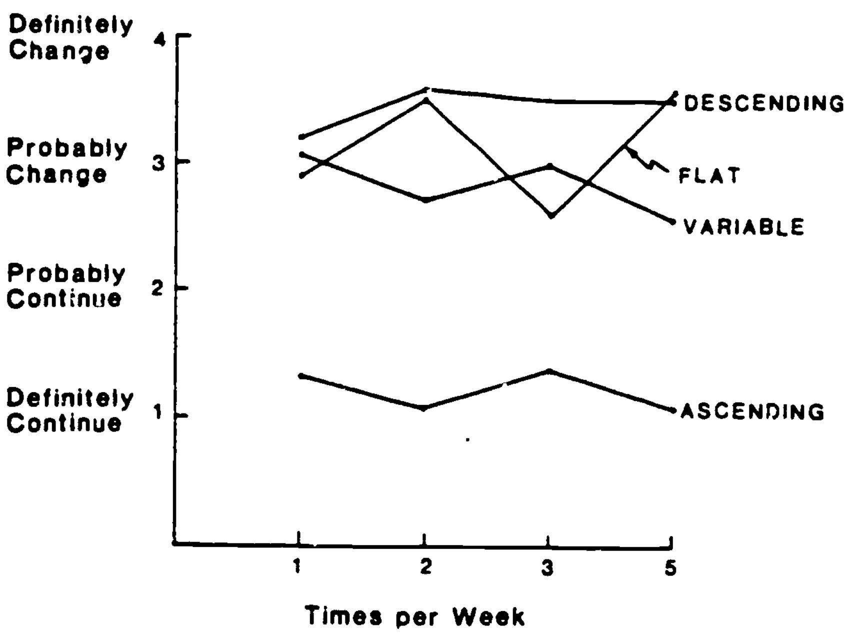


*Figure 3: Mean ratings of program decisions by graph type and frequency of data collection*

ables and teachers' program recommendations as the dependent measure are presented in Table 3. Main effects for type of graph and interaction effects of type of graph and frequency of data collection were again found to be statistically significant at the .05 level. Main effects for frequency of data collection were not significant.

Tukey's follow-up procedures were used to examine main effects of type of graph (ascending, descending, flat, and variable) on teachers' assessments of student progress. The analysis indicated that teachers generally were able to distinguish between the types of trends, as was

Table 2
Summary Table for Two-Factor Repeated Measures Design for Ratings of Student Progress by Type of Graph and Frequency of Data Collection

| Source | DF | MS | F |
|---|---|---|---|
| Type of Graph | 3 | 314.155 | 440.71* |
| Subject by Type | 174 | .713 | |
| Frequency | 3 | 5.065 | 14.26* |
| Subject by Frequency | 174 | .355 | |
| Type by Frequency | 9 | 7.264 | 25.17* |
| Subject by Type by Frequency | 522 | .289 | |

*$p<.05$

Table 3
Summary Table for Two-Factor Repeated Measures Design for Program Decisions by Type of Graph and Frequency of Data Collection

| Source | DF | MS | F |
|---|---|---|---|
| Type of Graph | 3 | 239.950 | 345.52* |
| Subject by Type | 174 | .694 | |
| Frequency | 3 | .682 | 1.78 |
| Subject by Frequency | 174 | .383 | |
| Type by Frequency | 9 | 6.204 | 18.27* |
| Subject by Type by Frequency | 522 | .340 | |

evidenced by significant differences in the mean ratings Only the difference in mean ratings between the flat and variable graphs was not found to be statistically significant. In contrast, post-hoc procedures examining main effects of frequency of data collection indicated that the only significant difference between level ans in the pairwise comparisons was between teachers' ratings of student progress as represented by data collected twice a week and ratings of student progress as represented by data collected once a week.

Post-hoc procedures examining interaction effects of type of graph and frequency of data collection indicated that when student performance data represented an ascending trend or systematic improvement, here were no significant differences between level means. That is, when the graphed data clearly represented an increase in student performance, teachers' assessments were similar when data were obtained each day, three times a week, twice a week, or once a week. When the graphed data represented a decrease in performance, no change in performance, or highly variable performance, several of the differences in means were statistically significant. When the trend of the student performance data was not ascending, teachers' ratings based on data obtained only once a week tended to be different than those based on data collected more frequently.

In examining main effects of type of graph on teachers' program decisions, follow-up procedures indicated that decisions based on grap representing systematic improvement in performance were significantly different than decisions based on the other three types of graphs. Decisions based on graphs representing a decrease in performance were also significantly different than those based on variable graphs.

The results of follow-up tests examining interaction effects of type of graph and frequency of data collection on teachers' program decisions are somewhat unclear. When the graphed data represented an ascending trend, two of the six pairwise comparisons between level means were found to be statistically significant. However, these differences did not appear to be systematic. When the graphed data represented a decrease in performance, no change in performance, and highly variable performance, several of the differences in means were statistically significant. Although these differences also did not appear to be clearly systematic, program decisions based on data obtained only once a week tended to be different than those based on data collected more frequently.

## Discussion

The three questions addressed within the two-factor repeated measures design were:

1. Do teachers' judgments and decisions differ according to type of trend?
2. Do teachers' judgments and decisions differ according to frequency of data collection?
3. Do teachers' judgments and decisions based on different types of graphs vary with frequency of data collection?

The results of the two-factor analysis of variance and subsequent follow-up procedures suggest that teachers' judgments and decisions do tend to differ according to type of trend. When teachers were asked to assess student progress, the ascending and descending conditions were found to be significantly different from each other and from the flat and variable conditions. When teachers were asked to make program recommendations, the ascending condition again was found to be significantly different from the other three conditions, and the descending and variable conditions were significantly different from each other. These findings suggest that teachers are able to distinguish between most trends and can clearly distinguish ascending trends from other types. The ability of teachers to distinguish between trends of graphed data is an important skill, as the use of graphs to make instructional decisions is largely dependent upon this ability.

The absence of significant differences in the mean ratings of teachers' judgments and decisions when presented with graphs that did not represent a systematic improvement in performance may be, at least in part, a function of the nature of the rating scales. When presented with flat and variable graphs, in which student performance was neither systematically improving nor decreasing, teachers tended to evaluate performance by making a single choice: "staying about the same." When asked to make a program recommendation based on graphs that did not represent an ascending trend, but were descending, flat, or variable, teachers also tended to select one choice: "probably change the program." Although the differences between mean ratings of program decisions for the descending and variable graphs were significant, those between the descending and flat, and flat and variable graphs were not.

The results of testing the main effects of frequency of data collection were mixed. When teachers rated student progress, the main effect for frequency showed that the differences were statistically significant. However, when teachers made program recommendations, the main effect for frequency of data collection was not found to be significant.

When teachers were asked to evaluate student progress, and the performance data were ascending, the results were clear and consistent with the findings of Munger and Loyd (1987). That is, when student performance data represent systematic and continuous improvement, teachers' judgments were similar whether probe data were collected daily, three times a week, twice a week, or once a week. These findings suggest that when a student is clearly making progress, it may be necessary to obtain probe data only once a week to evaluate performance. These results support the findings of Utley et al. (1987) who reported that when students demonstrated an increase in performance, subjects were able to analyze ascending trends with approximately equal accuracy, regardless of the amount of documentation (e.g., observation only vs. raw data vs. data in graphic form).

When teachers were asked to make program decisions and the performance data were ascending, the results were less clear. Munger and Loyd (1987) reported that when graphed data represented a systematic improvement in student performance, teachers' decisions were similar when data were collected each day, three times a week, twice a week, or once a week. By contrast, this study found that teachers' decisions tended to differ by frequency of data collection for all types of trends.

When the graphed probe data represents a decrease in performance, no change in performance, or highly variable performance, teachers' judgments as well as program decisions tend to differ by frequency of data collection. When the trend of the student performance data is not ascending, ratings based on data obtained only once a week tend to be different than those based on data collected more frequently. These results are consistent with those of Munger and Loyd (1987) who also found that when the treatment was descending, flat, or variable the majority of the significant differences in means occurred between ratings based on data collected once a week and the other three frequencies.

The results of the current and previous (Munger & Loyd, 1987) studies suggest that when the graphed probe data clearly represent systematic and continuous improvement in student performance, it may not be necessary for teachers to collect data more than once a week to assess student progress. However, when the graphed probe data represent a decrease in performance, no change in performance, or highly variable performance, this study suggests that data be collected more often than once a week, as teachers' judgments and decisions are not the same when based on data collected once a week and data collected more frequently.

These findings should be welcomed by classroom teachers. When probe conditions are similar to those used in this study (no reinforcement, error correction, or prompting given), students tend to learn little about the target skill during the probe. Thus, obtaining the minimum amount of probe data necessary to make consistent judgments and decisions is desirable. Also, since time spent collecting probe data reduces the amount of time available for teaching students, a decrease in probe frequency might provide more time to teach.

However, there is one situation in which teachers may want to increase rather than decrease the frequency of probe data collection, even though the graphed data represent systematic improvement. This situation concerns the accomplishment of IEP objectives. When objectives are written with criteria specifying a certain degree of accuracy over a period of consecutive days of probe performance, teachers may want to increase the frequency of data collection as student performance approaches criterion. This will enable criterion performance to be documented more quickly than if infrequent probes, as recommended by the results of this study, are continued throughout the intervention phase.

This study suggests that it may be necessary to collect probe data more frequently than once a week to obtain consistent judgments when student performance data do not represent an ascending trend. Although probe conditions are clearly less conducive to learning than are training conditions, an increase in the frequency of probe data collection may enable teachers to have more confidence in their judgments and program decisions than if data were obtained only once a week.

These results leave many unanswered questions. First, when teachers collect, but do not graph, probe data, it remains uncertain whether these results apply, since the results are based on the visual analysis of graphed data. When teachers do not graph their probe data, judgments of trend are more difficult and the applications of those findings may yield more disagreement in teachers' judgments. Second, it is not clear whether these findings, based on graphed *probe* data, can be generalized to graphed *training* data. The teachers in this study indicated that, for the same programs, they collected training data more often than probe data. Because many teachers feel that data collection during teaching interferes with their effectiveness (Holvoet, O'Neil, Chazdon, Carr, & Warner, 1983), it would be useful to them to know whether they could collect training data less often during instructional sessions and still have confidence in their judgments and program decisions.

Finally, it remains unclear as to the amount of data a teacher must have or how long a teacher must wait to determine the trend of a graph and apply appropriate decision rules. Although this study used graphs which extended across a period of 60 days, those graphs representing data collected once a week had only 11 data points. Further research would be necessary to determine whether the findings of the present study would apply when teachers examined data collected across only 11 days. Although the practice of White and Haring (1980) and others

(Browder, 1987; Browder, Liberty, Heller, & D'Huyvetters, 1986; Liberty, 1972, 1985) is to examine five to 10 days of graphed data before making a decision based on trend, more research is needed to determine whether teachers' judgments would follow the same patterns revealed in this study if the data spanned a shorter period of time and if the graphs represented fewer data points.

*Notes*

1. This research was supported by Grant G008530150, Office of Special Education and Rehabilitation Services, U.S. Department of Education.

2. The authors gratefully acknowledge the assistance of Leslie Farlow, Mary Fisher, and Shara Walters, as well as the teachers who participated in this project.

3. Requests for reprints should be sent to Martha E. Snell, Curry School of Education, 405 Emmet Street, Charlottesville, Virginia 22903.

*References*

Bijou, S.W., Peterson, R.F., Harris, F.R., Allen, K.E., & Johnston, M.C. (1969). Methodology for experimental studies of young children in natural settings. *The Psychological Record, 19*, 177-210.

Browder, D.M. (1987). *Assessment of individuals with severe handicaps*. Baltimore: Paul H. Brookes.

Browder, D., Liberty, K., Heller, M., & D'Huyvetters, K. (1986). Self-management by teachers: Improving instructional decision-making. *Professional School Psychology, 1*, 165-175.

De Prospero, A., & Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. *Journal of Applied Behavior Analysis, 12*, 573-579.

Grigg, N.C. (1986). *The process of visual analysis of time series data in applied settings*. Unpublished doctoral dissertation, University of Virginia.

Holvoet, J., O'Neil, C., Chjazdon, L., Carr, D., & Warner, J. (1983). Hey, do we really have to take data? *Journal of the Association for the Severely Handicapped, 8*(3), 56-70.

Jones, R.R., Weinrott, M.R., & Vaught, R.S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. *Journal of Applied Behavior Analysis, 11*, 277-283.

Liberty, K.A. (1972). *Data decision rules*. Unpublished manuscript, University of Oregon, Regional Resource Center, Eugene.

Liberty, K.A. (1985). Enhancing instruction for maintenance, generalization, and adaptation. In K.C. Lakin & R.H. Bruiniks (Eds.), *Strategies for achieving community integration of developmentally disabled citizens* (pp. 29-71). Baltimore: Paul H. Brookes.

Munger, G.F., & Loyd, B.H. (1978, February). *Effects of frequency of data collection and graph characteristics on visual inference*. Paper presented at the Annual Meeting of the Eastern Education Research Association, Boston.

Parsonson, B.S., & Baer, D.M. (1986). The graphic analysis of data. In A. Poling & R.W. Fuqua (Eds.), *Research methods in applied behavior analysis* (pp. 157-184). NY: Plenum Press.

Snell, M.E., & Browder, D.M. (1986). Community-referenced instruction: Research and issues. *Journal of the Association for Persons with Severe Handicaps, 11*, 1-11.

Utley, B.L., Zigmond, N., & Strain, P.S. (1987). How various forms of data affect teacher analysis of student performance. *Exceptional Children, 53*, 411-422.

White, O.R. (1971). *Pragmatic approaches to progress in the single case*. Doctoral dissertation, University of Oregon, University Microfilms, 72-8618, Ann Arbor, MI.

White, O., & Haring, N. (1980). *Exceptional teaching*. Columbus, OH: Charles E. Merrill.