

# The Effects of Performance Criteria on Learning and Retention of Spelling Words

Melissa J. Shirley and H. S. Pennypacker

Two eighth grade boys participated in a study on the effects of performance criteria on learning and retention of spelling words. Baseline conditions consisted of weekly tests with no criteria, whereas treatment conditions consisted of daily tests with: (a) no criteria, (b) an accuracy criterion, (c) a fluency criterion. Follow-up tests were given approximately 10 days after completion of each list. Correct frequencies and accuracy ratios were calculated for both learning trials and follow-up probes. Retention accuracy ratios were also calculated. The data indicate that teaching to a fluency criterion is most effective for learning spelling words.

Frequency measures have long been used by precision teachers to differentiate between performance that is proficient and that which is merely accurate. Certain frequencies of performance, termed fluent frequencies, must be reached in order for a skill to be useful, remembered, and applied (Haughton, 1980; Johnson & Layng, 1992; Maloney, Desjardins, & Broad, 1990).

Ivarie (1986) investigated the effects of different proficiency criteria on the retention of an Arabic/Roman numeral task. Subjects were required to perform the task to one of two frequency criteria. After meeting the criteria, subjects were given immediate, 1-, 2-, and 3-month posttests. The experimenter found that average and below-average subjects required to perform at a higher proficiency rate (70 rpm) achieved higher percent correct scores on a posttest than subjects required to perform at a lower proficiency rate (35 rpm) on the same task.

Berquam (1981) used a paired associates task with two classes of third graders. One class received fluency training during the initial learning phase, while the other received only unpaced practice. Fluency training resulted in better performance (measured in correct words per minute) and retention than unpaced practice. When the control group children later received fluency training, their retention improved as well.

One limitation of using frequencies as performance standards in the classroom is the lack of research-supported proficiency, or

fluency, aims (Evans & Evans, 1985). Fluency aims are the combination of frequencies and accuracy levels which have been shown to produce fluent performance. Several methods have been suggested for determining such aims. Most common is to measure the performance of normal or competent performers (White, 1984). Such standards can be obtained from either peers (Epstein & Cullinan, 1979; Jones, 1981) or competent adults (Wood, Burke, & Kunzelmann, 1978).

White (1984) points out that in order for a new behavior to compete effectively with an existing behavior, the new behavior must be performed at a higher frequency than the existing behavior. He suggests a frequency twice as high as the frequency of the old behavior. He proposes that if this frequency is not sufficient, an adult frequency may be necessary. Alternatively, he suggests that frequencies of new skills be half to two-thirds as high as tool skill frequencies. For example, Maloney et al. (1990) found that writing 160 digits per minute is required for writing math facts at 80 per minute. Eaton (1978) suggests a more complex method of setting aims which involves both tool skill frequency and comparison to competent adults.

Experts in the field have provided suggested fluency aims for many academic areas including math and reading. There is, however, a lack of research-supported fluency aims for other areas, including spelling. Likewise research has yet to show that fluency is related to retention in spelling skills. One reason for the lack of

fluency criteria in spelling may be that spelling, like many other academic skills, is usually taught by the calendar (Rowell, 1972). That is, tests are given each Friday, and new word lists are distributed each Monday, regardless of test performance. This calendar-based procedure provides little opportunity for the students to attain fluent performance.

The present experiment was designed to answer the following questions with respect to spelling: (1) Do daily performance sessions result in greater learning, follow-up performance, and retention than the weekly sessions typical of most classrooms; (2) Does a 100% accuracy criterion result in greater learning, follow-up performance, and retention than daily sessions with no criterion; and (3) Does fluency training produce greater learning and retention than accuracy training with no frequency criterion?

## Method

Two 8th-grade boys diagnosed with learning disabilities in reading and spelling were referred by their teacher for tutoring in spelling. Terry was a 15 year old attending special education classes on a full-time basis. Daniel was a 14 year old in special education reading only, though he was having difficulty in his other classes as well.

### *Pretest*

Prior to the study, the boys were tested on words from the Florida List for the Assessment of Spelling, the months of the year, and the numbers 1 through 20. Testing continued until a list of 200 unknown words was compiled for each boy. To get a measure of handwriting frequency, the students were asked to write their names as many times as possible in a one-minute period.

**Baseline I.** Baseline one was designed to replicate a typical spelling classroom. During this phase, which lasted four weeks, the experimenter met with the students three times per week. Each Monday the students were given 10 words. The students were asked to read each word aloud, and write it on paper. On Wednesdays the students wrote each word five times and in a sentence. On Fridays the students

participated in timed performance sessions. Immediate feedback was provided following the completion of the performance sessions. The students received new words on Monday, regardless of performance on Friday sessions.

**Baseline II.** Baseline two was identical to baseline one except that the students were given two lists of 10 words each week. The two lists were treated separately as two lists of 10 words each, not one list of 20 words. The duration of this phase was approximately four weeks.

**Phase III.** Phase three was designed to test the effects of a 100% accuracy criterion on learning. Sessions were conducted on each school day. On the first day each student was presented two lists of 10 words each. The students copied the words, wrote each five times and in a sentence. The next day the students participated in performance sessions, as in baseline. If a student did not spell all 10 words from one list correctly, he was required to perform on the same two lists again on the next day. Performance sessions continued until the student spelled all 10 words from one list correctly. When the student reached 100% accuracy on one list, he was presented with two new lists on the following day, and the procedure was repeated. The purpose of the second "yoked" list, for which there was no criterion, was to control for the effects of daily sessions. The duration of this phase was approximately three weeks.

**Phase IV.** Phase four was designed to test the effects of a fluency criterion on retention, controlling for practice (repetitions). The procedure for this phase was identical to that of phase three, but the criteria differed. In this phase, the student was required to perform with 100% accuracy on both lists. In addition to this criterion, the student was also required to perform one list at a "fluent" rate. Fluency was defined for each student by the frequency at which he wrote the letters in his name during the one-minute timing described previously. Fluency criteria were thus set at 91 and 105 letters per minute for Daniel and Terry, respectively. The two lists were yoked, such that both lists would be practiced the same number of times. Performance sessions continued until each student achieved 100% accuracy on one list and 100% accuracy with fluency on the other. When these criteria were met on both lists, the

student received two new lists, and the procedure was repeated. This phase lasted approximately four weeks.

### *Follow-Up Sessions*

Follow-up sessions were conducted approximately 10 days after the final performance session for each list in order to obtain a measure of retention. Procedures for follow-up sessions were identical to those for initial performance sessions.

### *Dependent Measures*

Spelling performance was measured in terms of correct and incorrect letters on each list. Correct frequencies and accuracy ratios were calculated for each performance session. Follow-up accuracy ratios were calculated by dividing the total number of correct letters by the total number of incorrect letters during the follow-up session. Retention accuracy ratios were calculated as for follow-up accuracy ratios using only the letters that were scored correctly during the final acquisition session.

## **Results**

### *Pretest*

Students were presented words until each misspelled 100. This required the presentation of 237 words to Terry and 234 words to Daniel. Accuracy ratios for Terry and Daniel were  $x1.9$  and  $x1.4$ , respectively.

### *Acquisition*

Performance during acquisition sessions was measured in terms of both correct frequencies (Charts 1 & 2) and accuracy ratios (Charts 3 & 4). For both students, implementation of the accuracy criterion in phase three resulted in correct frequencies that were slightly higher than baselines one and two, as well as daily sessions without an accuracy criterion. The change to daily sessions had a positive effect on the accuracy of Terry's performance, both with and without an accuracy criterion. Daniel, who initially showed the same effect, began performing better in the presence of the accuracy criterion.

Implementation of the fluency criterion in phase four led to an approximate doubling of correct frequencies for both students. Students' performance on lists on which fluency was required was consistently higher than that on which only accuracy was required, both across and within phases. Accuracy ratios remained high, as in all conditions in which the accuracy criterion was in effect.

### *Follow-Up Sessions*

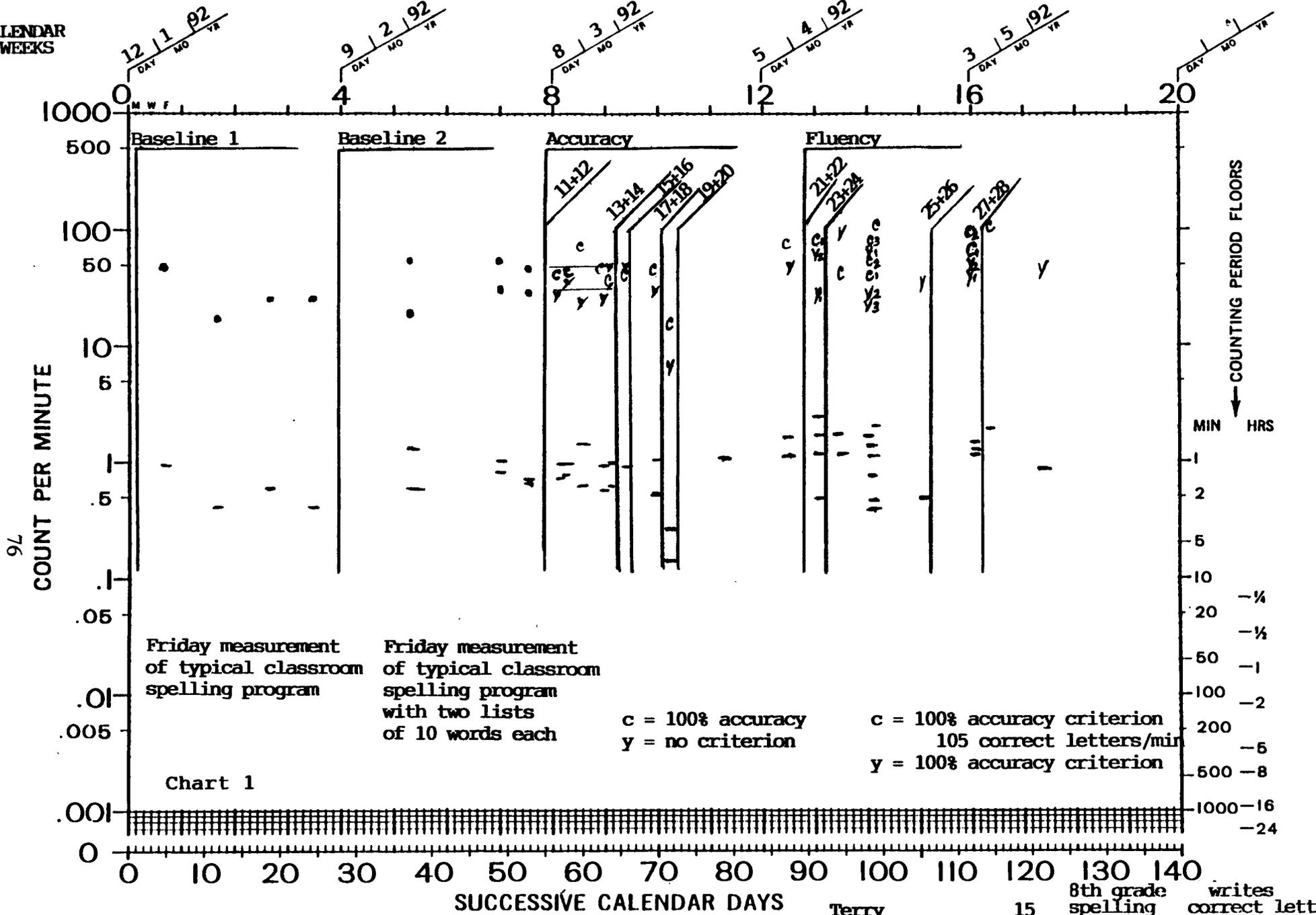
Correct frequencies (Charts 5 & 6) and accuracy ratios (Charts 7 & 8) were again used as measures of performance on follow-up sessions. Examination of between phase data reveals fairly stable performance on follow-up sessions throughout the study. Inspection of within phase data, however, exposes differences in follow-up performance related to experimental condition. In phase three, for example, both boys had higher correct frequencies on accuracy lists than on yoked lists. Whereas Terry's accuracy was also greater on accuracy lists, Daniel's data indicated no consistent effect of the accuracy criterion. Thus, within phase data, support the use of an accuracy criterion.

Examination of the data in phase four also reveals differential performance related to performance criteria. Implementation of the fluency criterion resulted in accuracy ratios that were higher than those obtained from the yoked condition. Correct frequencies were also slightly higher than with the accuracy criterion alone. Again, though not supported by across phase data, these data suggest positive effects of a fluency criterion.

### *Retention*

Retention data were calculated to determine what portion of correct performance was maintained after approximately ten days without practice. Retention accuracy ratios are presented in Charts 9 and 10. Across phase retention, data reveal differential trends across subjects. While Daniel's accuracy remained relatively stable throughout the study, Terry's accuracy ratio progressively declined over time. The within-phase data also differ from subject to subject. For Terry, the most stringent criterion in each phase resulted in the greatest amount of retention in that phase. Daniel's data, however, are not as clear. His performance in the third phase indicates that daily sessions, with no accuracy

CALENDAR WEEKS



Friday measurement of typical classroom spelling program

Friday measurement of typical classroom spelling program with two lists of 10 words each

c = 100% accuracy  
y = no criterion

c = 100% accuracy criterion  
105 correct letters/min  
y = 100% accuracy criterion

Chart 1

SUCCESSIVE CALENDAR DAYS

Terry 15 8th grade spelling writes correct lett

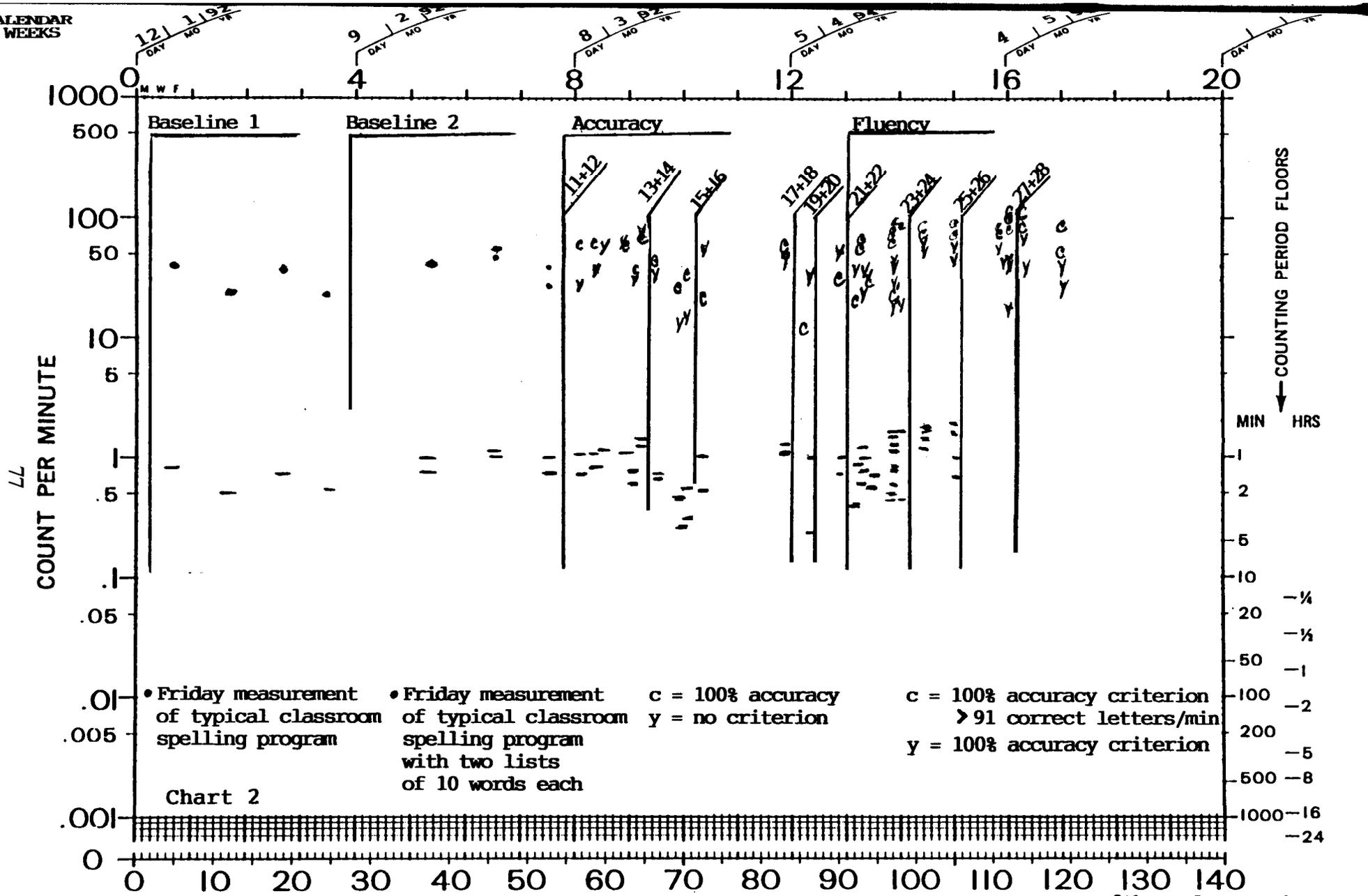
SUPERVISOR      ADVISER      MANAGER

BEHAVIOR      AGE      LABEL      COUNTED

DEPOSITOR      AGENCY      TIMER      COUNTER

CHARTER

CALENDAR WEEKS



• Friday measurement of typical classroom spelling program

• Friday measurement of typical classroom spelling program with two lists of 10 words each

c = 100% accuracy  
y = no criterion

c = 100% accuracy criterion > 91 correct letters/min  
y = 100% accuracy criterion

Chart 2

SUCCESSIVE CALENDAR DAYS

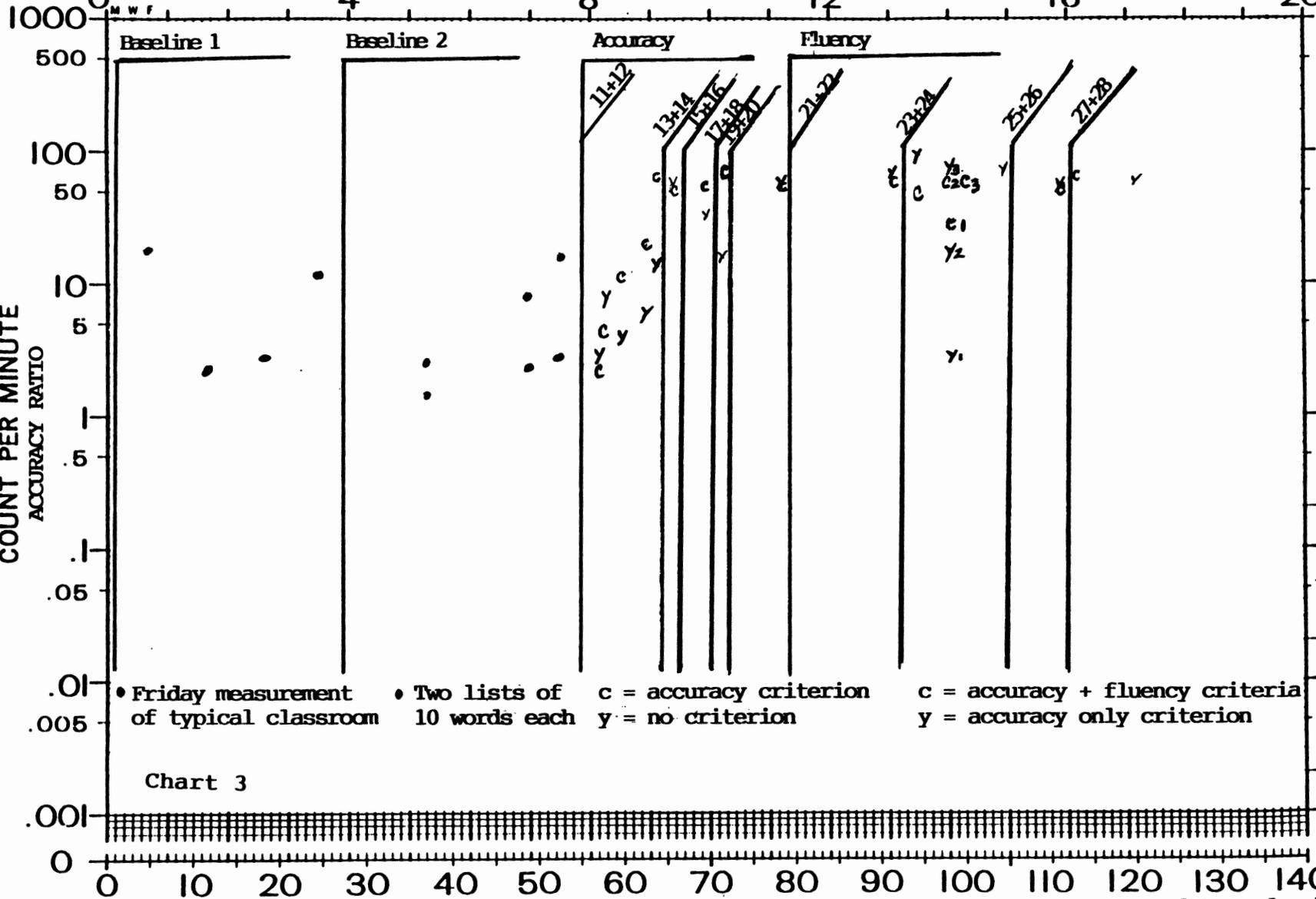
SUPERVISOR \_\_\_\_\_ ADVISER \_\_\_\_\_ MANAGER \_\_\_\_\_  
DEPOSITOR \_\_\_\_\_ AGENCY \_\_\_\_\_

TIMER \_\_\_\_\_ COUNTER \_\_\_\_\_ CHARTER \_\_\_\_\_

Daniel 14 8th grade writes spelling correct let  
BEHAVIOR AGE LABEL COUNTED

CALENDAR WEEKS

87  
COUNT PER MINUTE  
ACCURACY RATIO



COUNTING PERIOD FLOORS

MIN HRS

1  
2  
5  
10  
20  
50  
100  
200  
500  
1000  
24

SUPERVISOR ADVISER MANAGER

DEPOSITOR AGENCY

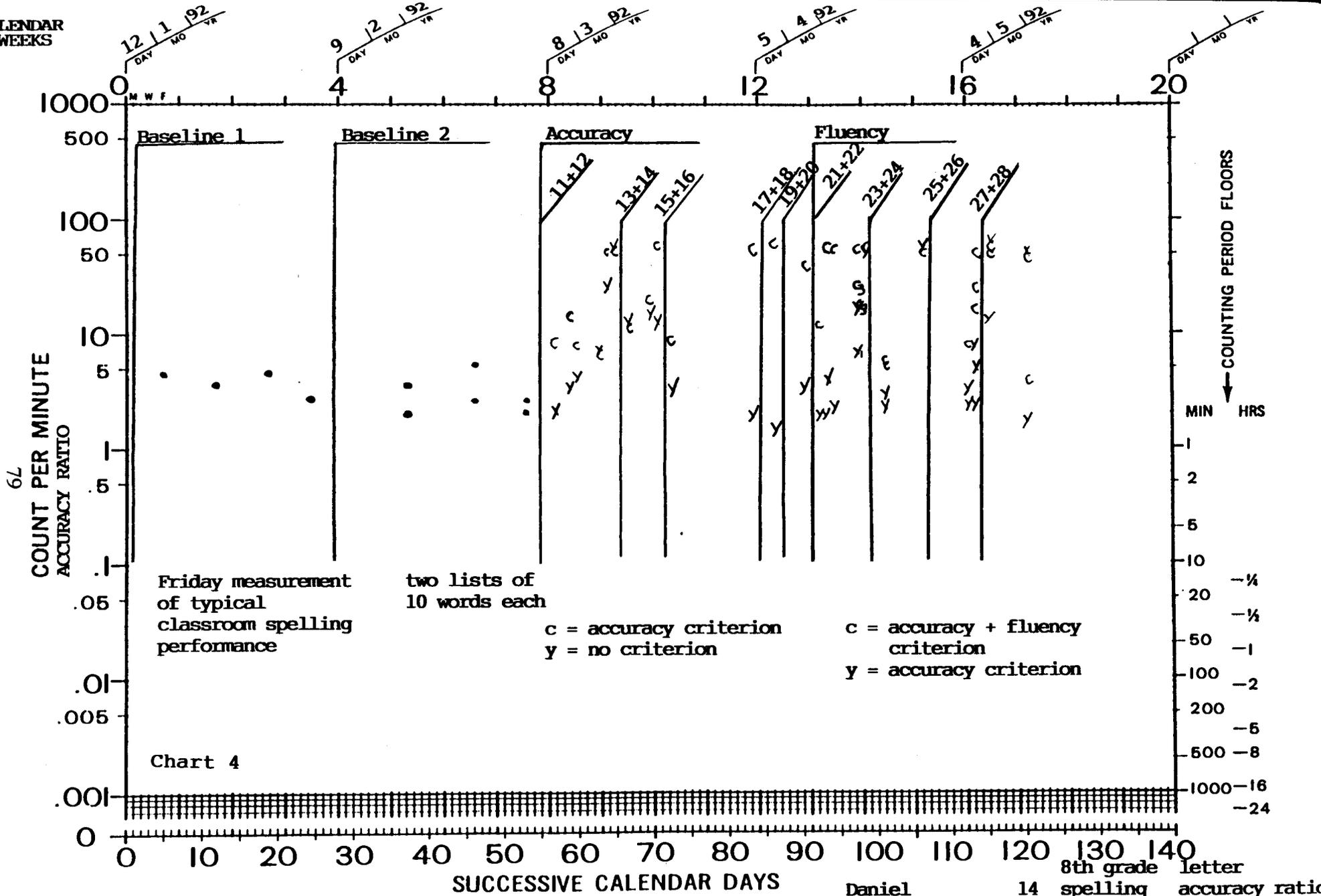
TIMER COUNTER

Terry 15  
BEHAVIOR AGE LABEL COUNTED

CHARTED

8th grade letter spelling accuracy ratio

CALENDAR WEEKS



SUPERVISOR      ADVISER      MANAGER

DEPOSITOR      AGENCY

TIMER      COUNTER

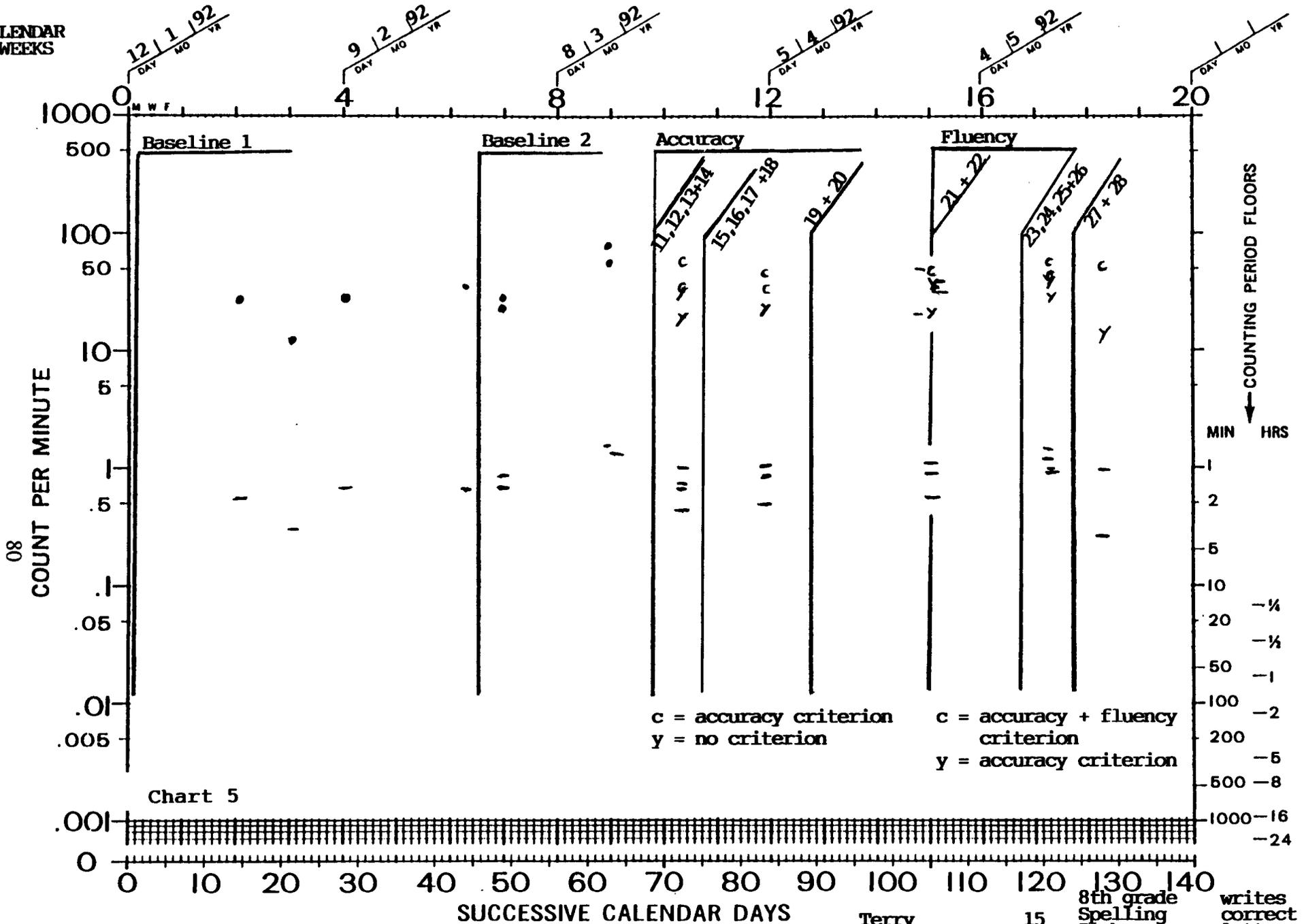
TIMER      COUNTER

Daniel      14      8th grade letter spelling accuracy ratio

BEHAVIOR      AGE      LABEL      COUNTED

CHARTER

CALENDAR WEEKS



SUPERVISOR      ADVISER      MANAGER

DEPOSITOR      AGENCY

TIMER

COUNTER

Terry  
BEHAVIOR

15  
AGE

8th grade  
Spelling  
Probe  
LABEL

writes correct  
letter  
COUNTED

CHARTER

CALENDAR WEEKS

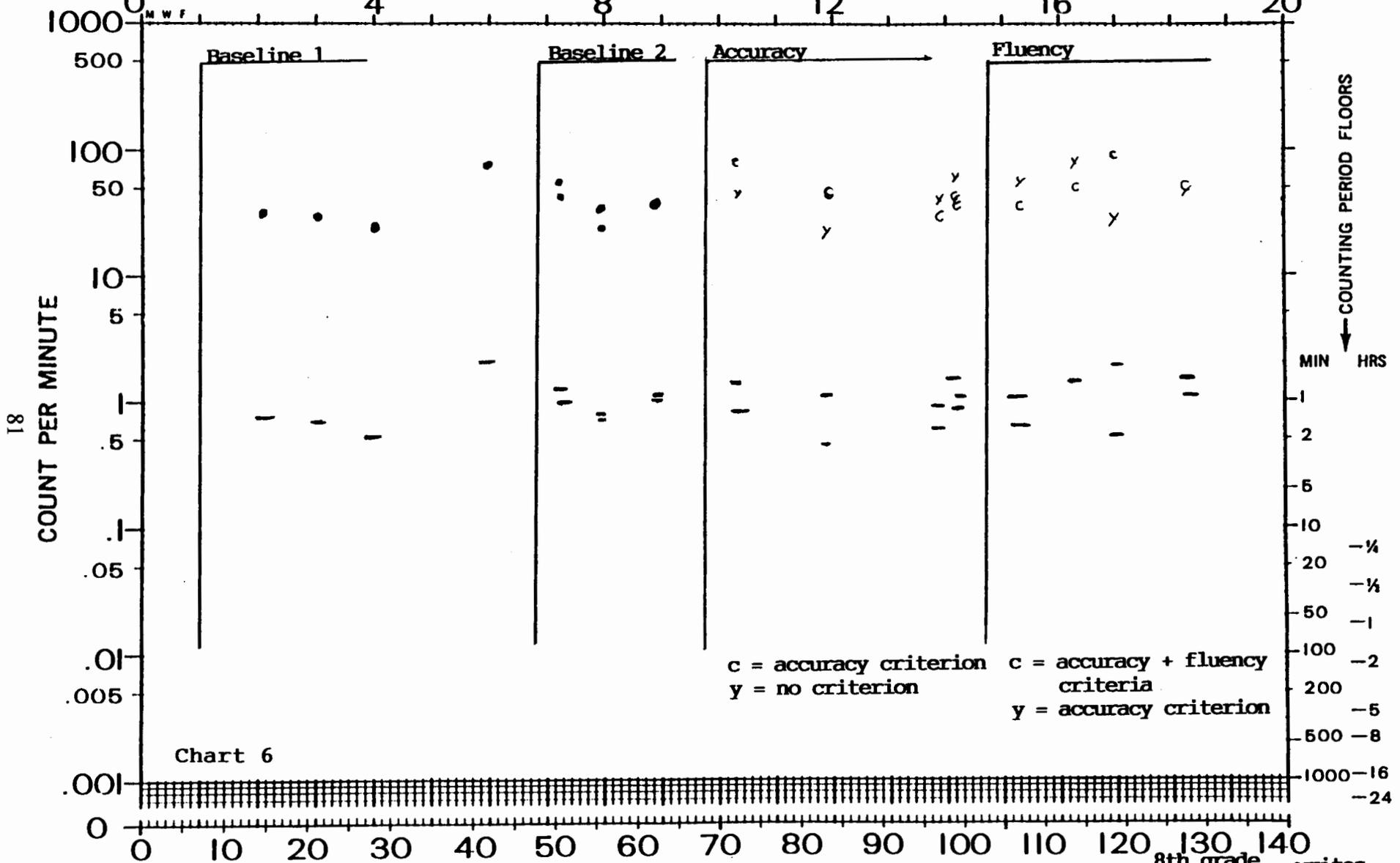


Chart 6

c = accuracy criterion  
 y = no criterion  
 c = accuracy + fluency criteria  
 y = accuracy criterion

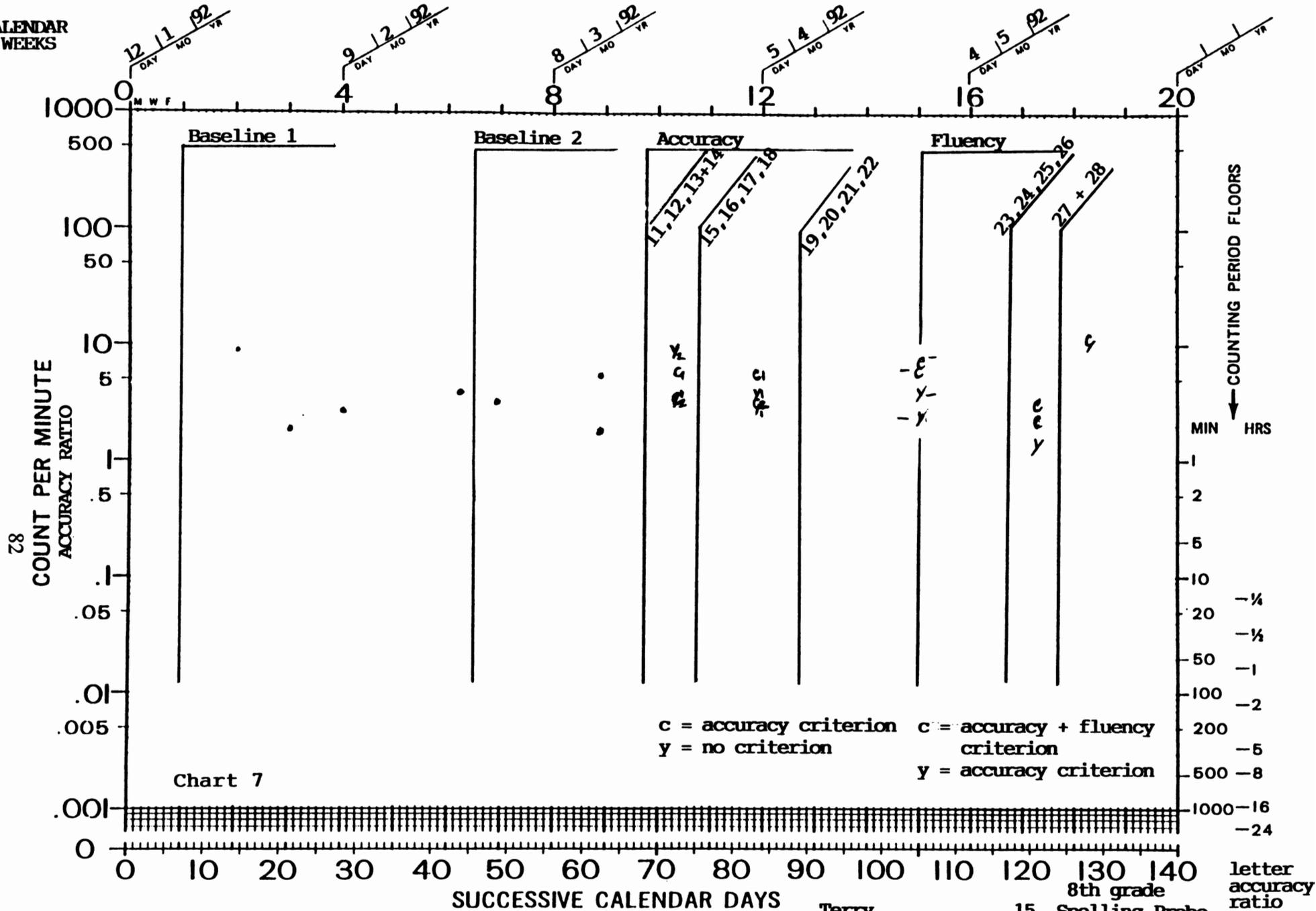
SUCCESSIVE CALENDAR DAYS

SUPERVISOR \_\_\_\_\_ ADVISER \_\_\_\_\_ MANAGER \_\_\_\_\_  
 DEPOSITOR \_\_\_\_\_ AGENCY \_\_\_\_\_

TIMER \_\_\_\_\_ COUNTER \_\_\_\_\_ CHARTER \_\_\_\_\_

Daniel 14 8th grade Spelling Probe writes correct letter  
 BEHAVIOR AGE LABEL COUNTED

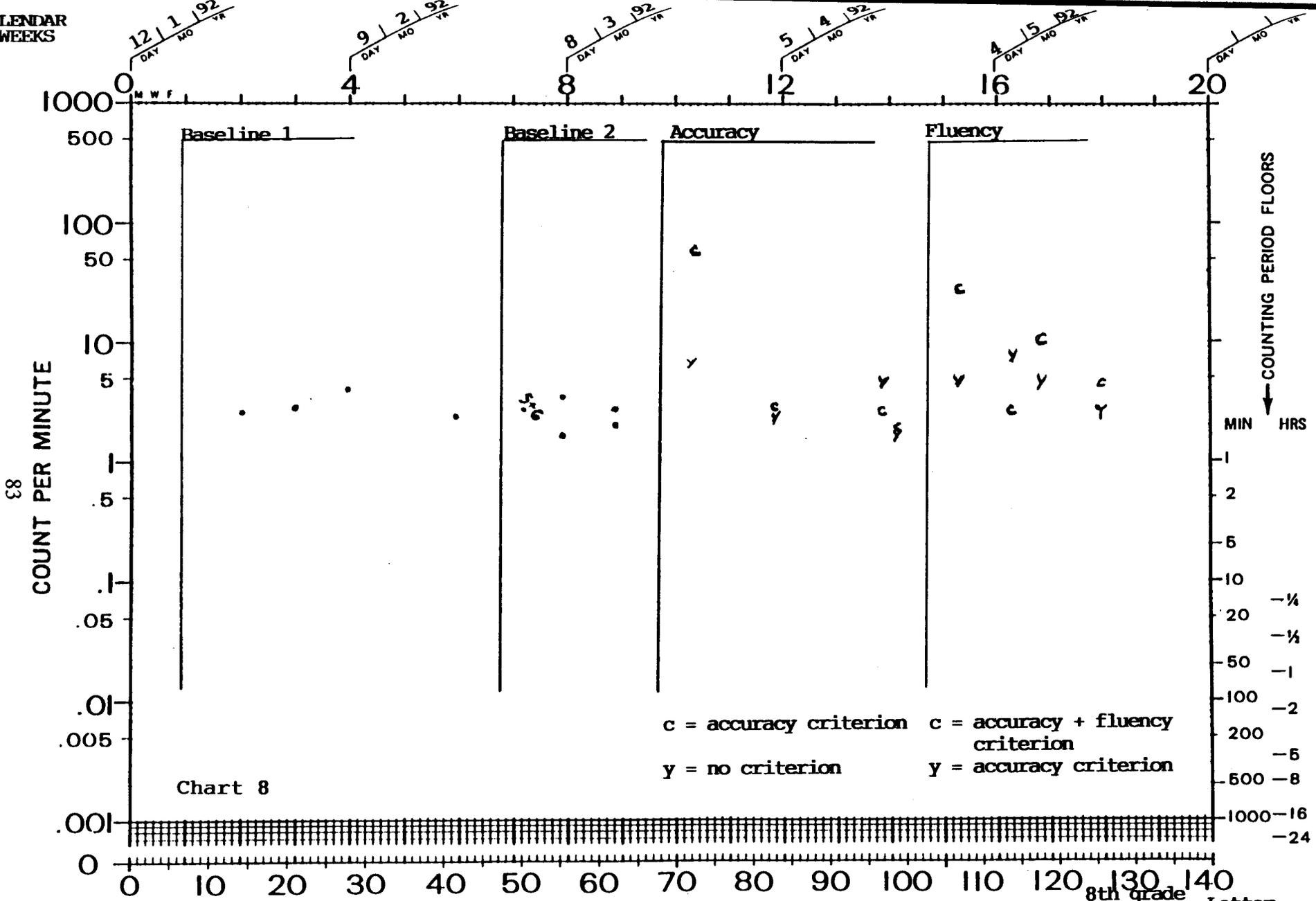
CALENDAR WEEKS



SUPERVISOR \_\_\_\_\_ ADVISER \_\_\_\_\_ MANAGER \_\_\_\_\_  
 DEPOSITOR \_\_\_\_\_ AGENCY \_\_\_\_\_ TIMER \_\_\_\_\_ COUNTER \_\_\_\_\_ CHARTER \_\_\_\_\_  
 Terry \_\_\_\_\_ 15 Spelling Probe \_\_\_\_\_  
 BEHAVIOR AGE LABEL COUNTED

letter accuracy ratio

CALENDAR WEEKS



SUPERVISOR			ADVISER			MANAGER			Daniel			14			8th grade Spelling Probe			Letter Accuracy Ratio								
DEPOSITOR			AGENCY			TIMER			COUNTER			CHARTER			BEHAVER			AGE			LABEL			COUNTED		

CALENDAR WEEKS

12 | 1 | 92  
DAY MO YR

9 | 2 | 92  
DAY MO YR

8 | 3 | 92  
DAY MO YR

5 | 4 | 92  
DAY MO YR

4 | 5 | 92  
DAY MO YR

DAY MO YR

1000  
500  
100  
50  
10  
5  
1  
0.5  
0.1  
0.05  
0.01  
0.005  
0

Baseline 1

Baseline 2

Accuracy

Fluency

COUNT PER MINUTE

COUNTING PERIOD FLOORS  
MIN  
HRS

1  
2  
5  
10  
20  
50  
100  
200  
500  
1000  
-16  
-24

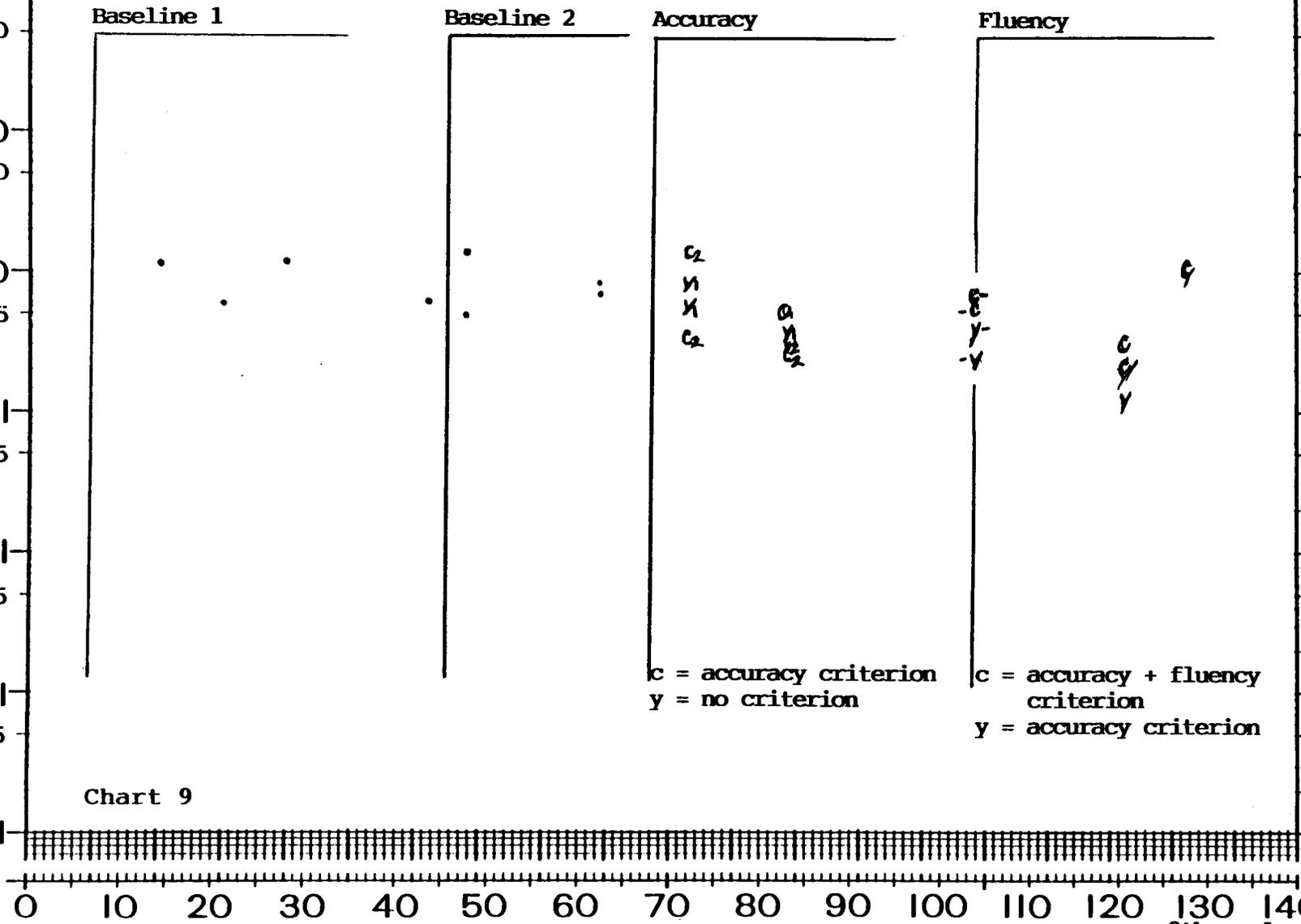


Chart 9

SUCCESSIVE CALENDAR DAYS

Terry  
BEHAVIOR  
15  
AGE  
8th grade letter  
Spelling retention  
accuracy ratio  
COUNTED

SUPERVISOR  
ADVISER  
MANAGER

DEPOSITOR  
AGENCY  
TIMER  
COUNTER  
CHARTER

CALENDAR WEEKS

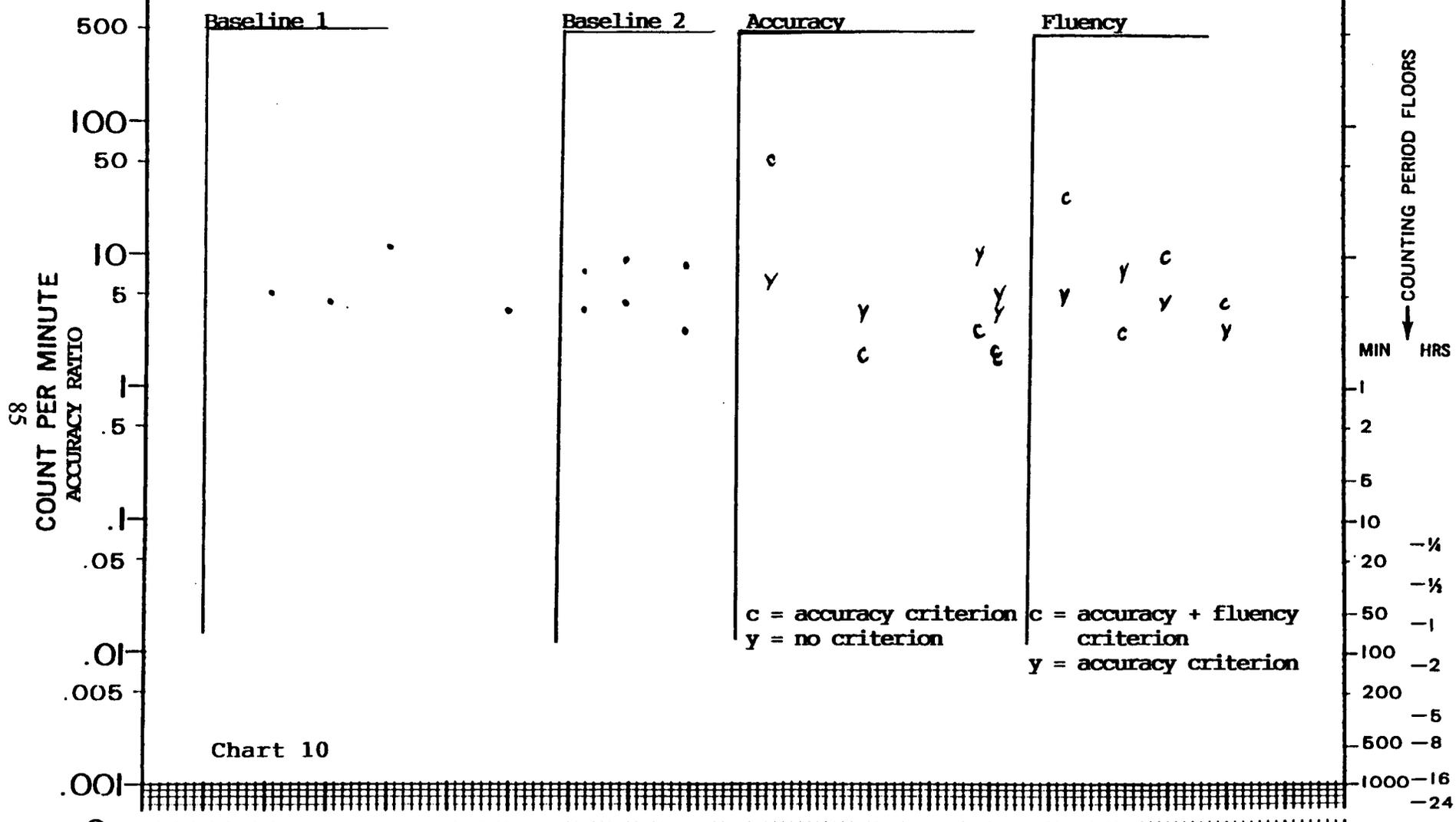
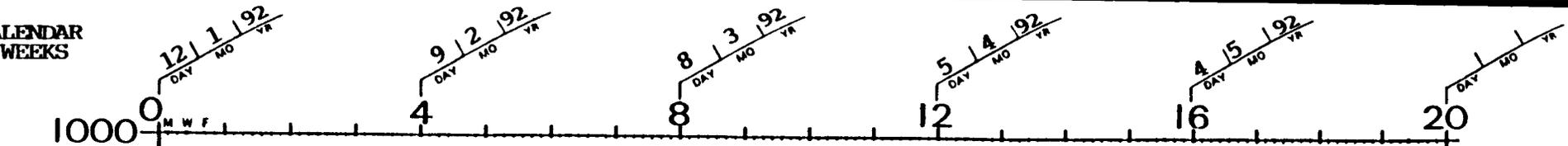


Chart 10

c = accuracy criterion  
y = no criterion

c = accuracy + fluency criterion  
y = accuracy criterion

SUCCESSIVE CALENDAR DAYS

SUPERVISOR \_\_\_\_\_ ADVISER \_\_\_\_\_ MANAGER \_\_\_\_\_

DEPOSITOR \_\_\_\_\_ AGENCY \_\_\_\_\_

TIMER \_\_\_\_\_

COUNTER \_\_\_\_\_

Daniel  
BEHAVIOR

14  
AGE

8th grade spelling retention accuracy ratio  
L A B E L C O U N T E D

CHARTER \_\_\_\_\_

criterion, resulted in greater retention than daily sessions with an accuracy criterion. The final phase suggests that the fluency criterion produces better retention than the accuracy-only criterion, though this conclusion is not supported by comparing phases three and four.

## Conclusion

It is possible to draw some conclusions from this study regarding the use of various criteria. First, daily sessions result in greater learning and follow-up performance than weekly sessions. Second, a 100% accuracy criterion results in greater learning, and for at least one student, better follow-up and retention performance, than daily sessions with no criteria. Third, a fluency criterion produces greater learning than an accuracy criterion alone.

It is possible that larger, more consistent, effects would have resulted from a higher fluency criterion or more repetitions at that level. Though these data suggest that benefits can be gained through the use of fluency criteria, various frequencies need to be investigated to determine which produce the greatest amount of learning and retention. In addition, the effects of requiring multiple performances at a fluent rate should be examined (Johnson & Layng, 1992). Perhaps requiring students to perform fluently on two or three consecutive occasions before proceeding to a new list would enhance retention. It is clear that altering criteria and the contingencies placed on performance, change the performance of the student. It is essential that researchers and teachers continue to work diligently to find the most effective ways to teach children.

## References

- Berquam, E.M. (1981). *The relation between frequency of response and retention, on a paired associates task*. Unpublished doctoral dissertation, University of Florida, Gainesville, FL.
- Eaton, M.D. (1978). Data decisions and evaluations. In N.G. Haring, T.C. Lovitt, M.D. Eaton, & C.L. Hansen, *The fourth R: Research in the classroom*. Columbus, OH: Charles E. Merrill.
- Epstein, M.H., & Cullinan, D. (1979). Social validation: Use of normative peer data to evaluate interventions. *Learning Disability Quarterly*, 2, 93-98.
- Evans, S.S., & Evans, W.H. (1985). Frequencies that ensure skill competency. *Journal of Precision Teaching*, 6(2), 25-30.
- Haughton, E.C. (1972). Aims, growing and sharing. In J. Jordan & L. Robbins, *Let's try doing something else kind of thing*. Reston, VA:CEC 20 - 39.
- Haughton, E.C. (1980). Practicing practices: Learning by activity. *Journal of Precision Teaching*, 1(1), 3-20.
- Ivarie, J.J. (1986). Effects of proficiency rates on later performance of a recall and writing behavior. *Rase*, 7(5), 25-30.
- Johnson, K.R., & Layng, T.V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist*, 47, 1475-1490.
- Jones, T. (1981). Peer-comparison percent and frequency scores for grades 8, 9, and 10. *Journal of Precision Teaching*, 1(4), 6-13.
- Maloney, M., Desjardins, A., & Broad, P. (1990). Teach your children well. *Journal of Precision Teaching*, 7, (2), 36-58.
- Rowell, C.G. (1972). A prototype for an individualized spelling program. *Elementary English*, 49, 335-40.
- White, O.R. (1984). Aims\*star wars. *Journal of Precision Teaching*, 5 (3), 55-64.
- Wood, S., Burke, L., Kunzelmann, H., & Koenig, C. (1978). Functional criteria in basic math skill proficiency. *Journal of Special Education Technology*, 2, 29-36.

---

**Are your dues to the Standard  
Celeration Society paid? They  
are due. Please send to Peggy  
Anderson, Treasurer, at Box  
247, Drummond, MT 59832.  
Half of your dues include your  
*Journal* subscription!**

---