DOCUMENT RESUME

ED 411 299		TM 027 327
AUTHOR	Moxley, Roy A.	
TITLE	Treatment-Only Single-Case Designs with Student Self-Recording for the Teacher as Researcher.	
PUB DATE	1997-03-00	
NOTE	44p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).	
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC02 Plus Postage.	
DESCRIPTORS	Academic Achievement; *Case Studies; *Educational Assessment; Educational Research; Elementary Secondary Education; Higher Education; Hypothesis Testing; *Measurement Techniques; *Research Design; *Researchers; Self Evaluation (Individuals); Teacher Role	
IDENTIFIERS	*Self Report Measures	

ABSTRACT

One valuable ingredient that is missing from common teacher practice is the measurement of ongoing student performance in the classroom in a way that supports improvement in student performance and in teacher instruction that can be shared with parents, teachers, and other professional colleagues. Treatment-only designs (TODs) with self-recording by students can provide that measurement for educational practices in public school classrooms, and they may be a useful way for educators and educational researchers to share useful information. Such designs may be considered as an elementary or case study level of single-case research. Examples are given of inferences from TODs and how they may lead to more sophisticated designs. The development of this approach allows for a two-way interaction between technology and science that promises some of the benefits that occur when technology and science are in close collaboration. Examples are given in the following areas in which TODs support: (1) self-recording; (2) self-recording with graphs rather than tables; (3) improving writing by increasing the rate of writing; (4) introducing more advanced single-case designs; (5) questioning a hypothesis; and (6) establishing the usefulness of types of instruction. (Contains 5 figures and 74 references.) (SLD)

Running Head: TREATMENT-ONLY DESIGNS

Treatment-Only Single-Case Designs with Student Self-Recording for the Teacher as Researcher

Roy A. Moxley

Department of Curriculum & Instruction

West Virginia University

Morgantown, WV 26506

Correspondence should be addressed to Roy A. Moxley, 604 Allen Hall, P.O. Box 6122, Department of Curriculum & Instruction, West Virginia University, Morgantown, WV 26506

Abstract

One valuable ingredient that is missing from common teacher practice is the measurement of ongoing student performance in the classroom. This measurement should be such that it supports improvements in student performances and in teacher instruction that can be shared with parents, teachers, and other professional colleagues. Demonstrable measures that can be cited as an indicator of improved practice are a hallmark of virtually all the technologies that developed into sciences. Treatment-only designs with self-recording provide that measurement for educational practices in public school classrooms, and they may also provide an avenue for sharing useful information with other professional educators, including educational researchers. Such designs may be considered as an elementary or case study level of single-case research whose primary outcome is teacher satisfaction, but these designs can also provide reasonable inferences about instruction. Examples are given of inferences from such designs and how they may lead to more sophisticated designs. The development of this approach allows for a two-way interaction between technology and science which promises some of the benefits that occur in areas where technology and science enjoy a close collaboration.

Treatment-Only Single-Case Designs with Student Self-Recording for the Teacher as Researcher

Although researchers commonly point out the implications of research for practice, Kaestle (1993) finds that much educational research appears to have little impact on mainstream practitioners. Moreover, Kaestle (1993) does not mention any responsiveness by educational researchers to data from practicing teachers. Indeed, one of the recommendations made by Kaestle was that researchers should involve "practitioners in designing and conducting research" (p. 30). Exception may be taken to these claims, but the fact that such views are arguable indicates the relationship between educational practice and educational research has need of improvement.

Historically, the disparity between science and technology has been resolved through a two-way interaction in which the effective actions of practitioners lead to scientific explanations and scientific explanations lead to effective practitioner applications. In the development of this interactive process over time, technology adopted some of the practices of science, and science adopted some of the practices of technology. Remarkable achievements such as space flight are products of this continuing interaction (cf. Bennett, 1986; Moxley, 1989; Vincenti, 1990).

What is blocking the development in education of an effective collaboration between practice and theory? What is lacking in education that is present in other areas of successful collaboration between technology and science? One obvious missing ingredient is a more effective method of measurement by teachers. Virtually all technological practices that contribute to science use measurement of some kind, and indeed it may be argued that the major contribution of technology to science has been measurement (for a brief overview of the history of graphical measurements introduced by technology but with a long and continuing interaction between science and technology, see Moxley, 1989, pp. 51-54).

It is not difficult to see why measurement is so beneficial to the advancement of technology and science. Measurement make variations conspicuous that would otherwise be undetected. This permits a greater number of selections, or adjustments, to be made. Improvements in a desired direction can be more readily detected, and the selection of conditions followed by improvements can be repeated until a satisfactory result is achieved. With repeated measurement, the carpenter can more readily make a level desk with smoothly sliding drawers. With repeated measurement, the scientist can make more accurate predictions and more readily refute or modify theory. With repeated measurement, the teacher can more readily keep instruction that works (instruction that results in improved student performances) and revise or abandon what doesn't work (instruction that results in poorer student performances). There are more opportunities for teachers to detect variations in student performances more opportunities for teachers to modify instruction and their interpretations of effective instruction.

What practical, repeated measurement of student performance is there in the public school classroom? The use of teacher-designed measures is limited by the amount of time a teacher is able to invest in correcting tests, and traditional standardized tests are not designed for frequent use during the school year. In addition, such tests commonly focus on ranking students. However, information on how students in a class rank in comparison to one another at different times during the school year provides no useful information for improving instruction. If the allotment of A's (e.g, 10%), B's (e.g., 20%) C's (e.g., 40%), D's (e.g., 20%), and E's (e.g., 10%) is predetermined, the same distribution will appear no matter how effective or ineffective the instruction. No matter how differently students perform under different instructional conditions, the students can always be ranked; and such rankings will fail to show useful differences between different instructional conditions. If the information is confined to rankings, the only change that can appear is in the order in which individual students are ranked. In addition, an emphasis on the rankings between students fosters competitive interactions in the classroom rather than the cooperative interactions many teachers prefer. Even when the distribution of grades is permitted to vary, a test designed primarily to furnish norm-referenced information on how students rank will be relatively insensitive to the effects of instruction as compared to tests designed to furnish criterion-referenced information on what students have achieved. This different between norm-referenced and criterion-referenced information is brought out further in the distinction between formative and summative evaluation as well as in the different ways information is referenced for different purposes (Moxley, 1974; 1979).

An alternative is to use frequently repeated measures of individual student performances to assess within-individual progress on what students are learning. To be feasible, measurement with a minimum of costs to teachers is a virtually prerequisite for initiating and continuing any widespread use of frequent measurement by teachers. It is not realistic to expect the typical teacher of a class of 20 to 30 students will readily record each student's individual progress in all the different subject areas on a daily. Even a weekly collection of data on student performances across all subjects may be problematic. Fortunately, there are ways to reduce the costs of measurement to the teacher. This can be done by (1) having the children do their own self-recording with monitoring by the teacher and (2) integrating the practices of self-recording into the existing curriculum so that time spent on self-recording is not seen to come at the expense of other curriculum areas. This means using the very activity of self-recording as a means of introducing skills in math (e.g., counting, addition), science (e.g., classification, prediction), language arts (e.g., reading, writing, and interpreting), and social science (e.g., cooperative interaction with group contingencies). If it takes more time to introduce young children to self-recording, this means these children will learn more academic skills through the very activity of self-recording. Teaching math skills and accompanying scientific skills by means of graphing has already become a widely accepted practice in the early grades (e.g., Choate & Okey, 1989; Caratello & Caratello, 1990; Lee & Miller, 1993; Preddy, 1993; Hynes, 1995), and it is a short step from that to teaching students how to self-record their progress.

Once routine measurement is established in the classroom—e.g., when students self-record repeated measures of their performances—the data obtained can be linked to interpretative practices that may eventually contribute to scientific knowledge. With such data, teachers can make repeated instructional revisions on the basis of the data as it is collected during the course of the instructional treatment. The teacher can also make repeated educated guesses on conditions that may improve student performances and try new conditions out in a repeated, evolving way. Instructional conditions that seem linked to improved performances may be kept, those that do not may be eliminated. A focus on such ongoingly-modified instructional treatment can be considered as a treatment-only design (TOD), a case-study variation or precursor of the more sophisticated single-case designs that are commonly found in behavior analysis. TODs are simpler than typical single-case designs in that they do not plan for more than one phase, the treatment phase. They do not plan, for example, for a baseline although baselines and multiple phases may develop after the treatment begins.

The following makes the case for using TODs with student self-recording in mainstream public school classrooms on a routine basis. The first section explains what TODs are. The second section explains the advantages of using student self-recording. The third section identifies the benefits of coupling TODs with self-recording. And the fourth section provides examples of interpretations from using TODs with student self-recording.

Treatment-Only Designs

A recent publication from the International Reading Association has advocated combining the methods of case study designs with the methods of single-case experimental designs (Bisesi & Raphael, 1995, pp. 104); and some textbooks on single-case designs (e.g., Barlow, Hayes, & Nelson, 1984; Kazdin, 1982) have shown how to do so in their consideration of early, elementary-level, single-case designs. Hawkins & Hursh (1992) refer to such single-case studies as treatment-only designs (TODs) because TODs share one data collection phase with more sophisticated single-case designs: the treatment phase. This means that TODs do not require an accompanying baseline or reversal procedure in their original plan to serve as a control for what would otherwise happen.

As an alternative terminology, treatment-only studies might be termed <u>baseline-only</u>

studies. However, the teacher using such a design is not seeking to achieve the stability sought for in baselines but an improvement in student performances. Other descriptors may also be considered. Kazdin (1982) characterized designs such as TODs as pre-experimental single-case designs (p. 87). If Kazdin's terminology is preferred, TODs for a class of students may be described as pre-experimental, multiple single-case designs. In relation to the A (baseline) B (treatment) terminology of single-case designs, such a design may be called a "B" or a "B-only" design (cf. Barlow & Hersen, 1984, p. 142). Yin (1994) indicates that studies with multiple subjects (as in a classroom) could be designated as multiple case study designs. In arguing to retain the designation treatment-only design, the case can be made that the descriptor treatment-only is fairly brief, suggests a distinction from traditional case studies, and suggests a relation to the more commonly used single-case designs in behavioral research.

Whatever designation is used, such designs share the advantages of traditional case-study designs:

> Among their advantages, the case study method can be used to (1) foster clinical innovation, (2) cast doubt on theoretic assumptions, (3) permit study of rare phenomena..., (4) develop new technical skills, (5) buttress theoretical views, (6) result in refinement of techniques, and (7) provide clinical data to be used as a departure point for subsequent controlled investigations. (Barlow & Hersen, 1984, p. 141)

In addition, the repeated measures that distinguish TODs from traditional case studies make arguments stronger. They are supported by more data points and allow for a comparison of the effects of different interventions on the same performances. Although these comparisons may only suggest rather than demonstrate a causal relation, they may eventually lead to such a demonstration.

An argument in favor of TODs is that they are an entry-level introduction to more sophisticated research and may provide a foundation for later work:

> Traditional research methodologists have tended to put case studies into a separate (and lower) class. Their functions are seen as quite limited. Case studies are said to be of use

> primarily because they suggest hypotheses, for example. While this is surely so, it can amount to damnation by faint praise because virtually anything can suggest hypotheses. Such a view discourages practitioners from analyzing their cases. Without this beginning, where will more elaborate analyses come from?...Even those analyses that end up as a lower level (e.g., a B-only [treatment-only], an A/B) will form a foundation for additional work.
>
> By encouraging case studies, we increase the likelihood of even more meaningful output. If practitioners are discouraged from doing case studies, why should they use time-series methodology at all? In this sense, an openness to case analysis and to the case studies that may often result is a cornerstone of applied time-series methodology. (Barlow et al., 1984, p. 281)

On this view, TODs have justification as a foundation or preparation for further research.

More specifically, Hawkins & Hursh (1992) identify three levels in a hierarchy of single-case research designs. Level 1 (Accountable Service Delivery) relies on TODs. The primary measurement task is to collect data during the treatment. At this level, the treatment-only level, satisfactory improvement is sought, and the treatment may continue until satisfaction is reached if not otherwise terminated by an established time framework such as commonly occurs in educational settings. Improvement simply means any change in performance in the desired direction with no implication of causal attribution. Indeed, many practitioners are primarily interested in seeing that performances change in a desirable direction and are only secondarily interested in attributing causal sources for the changes. An initial demonstration of causal explanation can be made at Level 2 (Semi-scientific evaluation of procedures or programs). Level 2 requires a credible baseline for an "AB design" or "some comparison condition or group, though it falls short of the ideal that an audience of scientists usually demands" (p. 66). The baseline may be introduced prior to the treatment, or the treatment itself may be regarded as a baseline for a subsequent treatment. The comparison condition may be one that occurs within the client's history or one that occurs between the client and other clients. The clients may be exposed to similar or to different treatments for comparison purposes. Level 3

(Scientific-quality Research) provides a more convincing demonstration of causal relations. This level requires "some form of replication, such as using a multiple-baseline design, an ABAB withdrawal design, or an alternating treatments design" (p. 68). Such designs can substantially isolate the relation between the intervention and the effects and are able to withstand the criticism that alternative cause and effect relations, such as maturation and its effects, may account for the changes observed.

These stages suggest how a two-way interaction between technology and science may work. As the above indicates, Level 1 may become level 2 when the initial TOD becomes a baseline for a subsequent treatment; and level 2 may become level 3 when replication occurs. In this way, TODs may be regarded as entry level designs that may lead to more sophisticated designs that support stronger cause and effect statements. Such a sequence would establish a direct connection from technology to science. A sequence from science to technology may also occur when the outcomes of level 3 research provide explanatory principles that can be introduced back into level 1 for guiding practitioner activities. However, just as all instances of technological satisfaction do not evolve into scientific explanation, all instances of scientific explanation do not evolve into technological satisfaction. The implementation of scientific knowledge may encounter a variety of obstacles which cannot all be overcome. In addition, there is no evidence for an effective scientific implementation at the technological level unless there is data collected at the technological level. The need for technological measurement in order to develop a two-way partnership between technology and science should be fairly evident.

Whether or not TODs lead to more sophisticated research, the primary theoretical contribution of TODs is the development of tentative, plausible hypotheses. The logic used to make these tentative hypotheses is the logic of making probabilistic inference from the consequences—which follow or accompany a TOD—that are consistent with contexts. These contexts include supporting literature as well as the current classroom setting. This inferential procedure is quite different from formal logic, such as occurs in syllogisms, where arguing from the consequent to the antecedent is considered invalid—a case of affirming the consequent—because the argument cannot be made with certainty. A lack of certainty, however, is not necessarily fatal in the empirical world and may be regarded as one of its abiding features. We work successfully with probabilities. Inferring an hypothesis from various facts has been called abduction and is one of three modes of argumentation employed in research along with induction and deduction (see Gallie, 1966, pp. 93-108; Hanson, 1958, pp. 85-92; Overholt & Stallings; 1976 Peirce, 1931-1963, 5.151-5.317). Abduction is also called retroduction, or hypothesis by Peirce as well as inference to the best explanation by others. Although abduction is sometimes regarded as a form of induction, Peirce distinguished a tentative abductive inference that unifies diverse facts from the increasing predictive probability of an inductive inference derived from an accumulation of similar instances. Examples of such abductions are found in criminal trials in which a prosecutor may argue from the diverse facts of motive, opportunity, and means to the guilt of the accused. Such an abduction stands in contrast to the induction an expert witness may provide on the probability, from many previous DNA blood matches, that the blood found at the scene of the crime matches that of the accused.

Illustrating the procedures of abductive research, Overholt and Stallings (1976) observed that an anthropological study does not begin with an hypothesis held constant in the manner of hypothetico-deductive experimental research, "Rather, the anthropologist manipulates hypotheses in order to arrive at statements that account for as many of the observed facts as possible with the greatest degree of economy, simplicity, and elegance possible" (p. 14). Accordingly, "The anthropologist is much more in the hypothesis tailoring than the hypothesis testing business" (p. 14). A promising hypothesis/explanation in itself may be considered a worthwhile outcome of a study. Darwin's theory of natural selection, which did not rely on demonstrative logic or repeatable experiments, is a singularly outstanding example of an inference to the best explanation. It is a theory that accounts for a diversity of facts; e.g., the fossil record, the variety of species found today, the different species found under different conditions, embryological development, available time, vestigial organs, the lack of intermediary species, heredity, and so on. But inferences do not need to be that spectacular to be useful, and teachers may rapidly generate and act on a variety of such hypothetical explanations during the course of a TOD before writing a final summation.

It is not just that some interpretations may become more plausible from a case study, but that other interpretations may become less plausible. One example may raise questions about, if not disconfirm, a previous account. In regard to evolution, this was dramatically illustrated in an observation by Nietzsche (1881/1982):

> The impartial investigator who pursues the history of the eye and the forms it has assumed among the lowest creatures, who demonstrates the whole step-by-step evolution of the eye, must arrive at the great conclusion that vision was <u>not</u> the intention behind the creation of the eye, but that vision appeared, rather, after chance had put the apparatus together. A single instance of this kind—and 'purposes' fall away like scales from the eyes! (p. 125)

A single instance such as this is enough to raise questions about accounts in terms of special creation or in terms of previously existing designs. This same instance may also lead to hypothesizing or supporting a theory of natural selection.

As Kazdin (1982) points out, "Even uncontrolled case studies may permit one to rule out rival interpretations" (p. 101); and Kazdin offered the following guidelines for strengthening the inferences from pre-experimental studies:

> The conclusions that can be reached from case studies and other pre-experimental designs are greatly enhanced when objective measures are used, when performance is assessed on several occasions over time, when information is available regarding the stability of performance over time, and when marked changes in behavior are associated with the intervention. Pre-experimental designs that include these features can closely approximate single-case designs in terms of the inferences that can be drawn. (p. 103)

Some if not all of these features may be included in the TODs that teachers use in the classroom.

Student Self-Recording

In using TODs in the public school classroom, the data on "things done" (Lee, 1994) can be collected and graphed largely by means of student self-recording. Research in a wide variety of areas indicates that self-recording in itself tends to have a favorable effect on improving the performances that are recorded (cf. McLaughlin, 1976, pp. 649-653; Morgan, 1984, p. 22; Moxley, Kenny & Hunt, 1990; O'Leary & Dubey, 1979; Rosenbaum & Drabman, 1979; Studwell & Moxley, 1984). Barlow et al. (1984) refer to this effect as the reactivity of self-monitoring; and it is fairly robust, even in the face of inaccurate recording:

> When clients self-record, the behavior that is being self-recording changes in frequency. This phenomenon is well established and has been termed the reactivity of self-monitoring (Kazdin, 1974; Nelson, 1977). Generally, this reactivity is therapeutic. The direction of the reactive behavior change is in a desirable direction. In other words, through self-monitoring, desirable behaviors increase in frequency and undesirable behaviors decrease in frequency (Cavior & Marbotto, 1976; Kazdin, 1974; Sieck & McFall, 1976).
>
> This reactivity is also independent of the accuracy of self-recording. This notion was suggested by Nelson and McReynolds (1971) and has been experimentally confirmed (Nelson, 1977). In other words, even if the self-recorded data are not accurate, the phenomenon of reactivity occurs. (p. 112)

Inaccurate data, however, opens up problems for interpretation; and Barlow et al. (1984) found in their survey of studies that the accuracy of the data may be increased 1) if the self-recorder "knows his or her accuracy can be checked on a random basis" and 2) if "self-recorded data is reinforced for accuracy" (p. 109). These recommendations apply to children as well as to adults (Barlow, et al., 1984. p. 110). To assure reasonable accuracy, the classroom teacher may have students check the accuracy of each other's data, and the teacher may also spot check that data.

With the records before them, teachers as well as students may readily be confronted with any need to reexamine their goals, scales, indicators, and displays. What are the valued curricular objectives? How large should the steps be to accommodate all the students in the class? What are useful indicators of progress toward these objectives? What are appropriate displays of this progress? The displays of data provide opportunities for feedback and discussions of possible cause and effect relations that may suggest future revisions to the teacher. Such discussion may be among teachers, students, and teachers and students together. Discussions among teachers are further enhanced when more than one TOD for a similar performance can be compared. Such comparisons can show different performances in relation to different treatments and this can suggest alternatives for improving performances. Collations of multiple designs using similar performances allow practitioners to see what other teachers have done and to borrow ideas they find promising.

For many teachers, one important use of student self-recording is to facilitate cooperative interactions. The emphasis in such records need not be on between-student comparisons, but on within-student progress. Group summaries of progress derived from individual data can be prominently displayed although the individual data may or may not be publicly displayed. When there is an acknowledgment or celebration of class improvements or achievements, class records on the performance of the class as a group may encourage cooperative interactions. Unlike competitive relations where there is commonly some disadvantage for one student to help another student because a gain in rank for one student requires a loss in rank for another, it is an advantage for students to help other students if a gain for one student means a gain for all students. In addition, substantial research exists to support the idea that, in many classrooms, cooperative classroom structures advance overall academic achievements more than strictly competitive or individual classroom structures (cf. reviews by Leming & Hollifield, 1985; Qin, Johnson, & Johnson, 1995). This research is also supported by behavior analytic studies of group contingencies (see Slavin, 1991).

Many recommendations for classroom graphing are more specific to circumstances. For example, if there is any concern about problematic comparisons between children, any public display of graphs (e.g., to show class progress toward a group consequence) should not show the names of individual children. Information on individual progress may be left to privately kept records. Various other adjustments may be made. Checklist graphs may be preferred with young children when they check off each letter, shape, color they learn with a sticker, self-inking stamp, or symbolic mark while a frequency count of the number of children who have achieved such skills may be used on a class graph. With class charts, some teachers prefer a percentage scale that takes into consideration the different opportunities for achievement when some students are absent. If a convenient introduction to comprehensive recording for all subject areas is sought, students may graph the number of words they write in different subject areas (various levels of criteria may be selected for what counts as a word written, e.g. freewriting at one level, proofread writing at another level). Or students may record the number of questions they ask and answer in different subject areas, including the rate of generating and answering questions. Accuracy in answers may be an initial concern, but fluency or speed in accomplishing a task should always be considered and often aids accuracy. Sometimes fluency may be developed in a part to whole procedure as when fluency in addition, subtraction, and multiplication are developed as prerequisites to division. At other times, fluency in creative or complex skills may be developed in a whole-to part procedure. An indicator such as words written in creative story writing may be increased along with arrangements and monitoring to see that component skills such as handwriting, spelling, and punctuation skills are increased along with it. The result may be a combined whole-to-part-to-whole procedure.

The particular graph scales that are used also depend on circumstances. Checklist scales are often used with young children who do not yet know how to count. Percentage scales may be preferred for class graphs when the absences of children is seen to have a misleading effect on the appearance of trends in class achievement. When the variations among children in the frequency of their beginning achievements are quite wide, the scales for different children may begin at a number that is near where they start. If the differences among the frequency counts at which different children improve is great, a different scale can be used for different children: a scale with larger increments for children with smaller counts, a scale with smaller increments for children with larger counts. If a child still reaches a frequency count that goes off the graph, a number of adjustments can be made. 1. The child can be given a new graph to record on that has a higher frequency on its scale. 2. The scale can be extended upward by adding another page of graph paper and folding it over the first page if the graph is to be stored in a folder or book. "Going off the chart" is a conspicuous indicator that is appealing to many young children. 3. The recording of the frequency can be extended over into the next column. If the scale goes from 1-10, it still remains fairly easy to see the total count at a glance although improvements are less conspicuous. 4. Standard semi-log scales can be used for all children when the teacher wants to quickly campare the rates of progress of different children. This scale emphasizes improvements in the rate of rate and is a distinctive feature of precision teaching (Pennypacker, Koenig, & Lindsley, 1972). Students who are unfamiliar with semi-log scales will need to become familiar with how to record on this scale and how to do semi-log scale interpolations.

With experience, teachers will find they can graph progress in any areas they can conceive of in empirical terms, including projects in problem solving and creativity. What is needed is to determine steps along the way to completing the project along with satisfactory indicators of the project's progress and success. These steps (e.g., brainstorming, discussion, etc.) and indicators (e. g., audience approval, expressions of preference, etc.) may all be assembled on a checklist. Progress and improvements may then be entered as they occur (cf. Moxley & Su, 1994). A guide to graphing in classrooms as well as a guide to writing up treatment-only studies with self-recording may be helpful here (e.g. Moxley, 1994)

Precedents for classroom recording exist in previous educational practices. Many of the practices found today in precision teaching—which is distinguished by the use of graphs with a semi-log scale for measuring fluency (cf. Binder, 1996; Lindsley, 1991, 1996; West & Young, 1992)—have educational precursors in August Dvorak's scientific management approach to typing (cf. Joyce & Moxley, 1988). These precursors include reinforcement of correct responses, timed probes, warmup, premium on frequency, modeling, group contingencies, contracts, charts, individual data collection, celeration, acceleration, and self-recording. In addition, Carlton Washburne (1922), an associate of John Dewey, recommended that students measure their own performance and keep "definite records of their improvement from day to day" (p. 203). In the area of reading, which has a long-standing tradition of using graphs, O'Brien (1926) found, "The direction of slant of the line tells the whole story....The pupil becomes determined to 'make that line go up'" and concluded, "The individual graph made one of the strongest appeals to the pupils and proved one of the most effective instruments in stimulating their speed in reading" (pp. 74-75).

The early educational history of graphing data on student achievements was influenced by industrial practices, which gave rise to the scientific management movement in education. In applying scientific management, Franklin Bobbitt (1913) recommended graphical displays for recording individual student progress:

> This putting of the educational product in the forefront of education means the establishment of a continuous record of progress in the case of each of the products....Such a continuous record must be kept, naturally, in the case of each of the many score educational products so as to show how each pupil at any time measures up against the standard. Simpler than parallel columns of figures would be graphical representation, the only objection being the necessity of increased space and labor. (p. 23)

Bobbitt makes no mention of self-recording here. His emphasis on educational goals in terms of products and standards allows free reign for competition between students and invites a thin schedule of positive reinforcement if only the achievement of the standard is to be reinforced. In fact, if aversive methods were used when standards were not met, then reaching a standard would be more of an escape from threatened punishments than gaining a reward. Such a circumstance might easily undermine the benefits from self-recording in as much as many students are resourceful in escaping from threatened punishments in directions that do not advance academic performance.

There are other significant ways in which methods of modern business management and its suggested applications to education differ from earlier concepts of scientific management. For example, TQM (Total Quality Management) emphasizes continuous improvements rather than reaching a fixed goal and it emphasizes cooperative rather than competitive relationships (Bonstingl, 1992, pp. 6-19, 77-85). The positive reinforcement schedule may be much denser here. Making note of improvements provides more opportunities for reinforcement than fixed goals do; and cooperation does not entail the punishments that are inherent in competition. Such differences may partially account for the failure of early student self-recording in education to become a more routine practice in public school classrooms.

Coupling Treatment-only Designs to Student Self-Recording

The idea of TODs for the classroom being advanced here is not merely to have students self-record but to have teachers (and perhaps students) engage in an interpretive enterprise—to make changes in instruction and to provide interpretations for the relation of these changes to changes in student performances regardless of how tentative these interpretations may be. Even if teachers do no more than infer an hypothesis based on the data they obtain, they are engaging in a data-based interpretive process. In doing so, TODs open up a more direct avenue to more sophisticated experimental single-case designs than would otherwise be the case. Teachers who use TODs with interpretations may receive positive reinforcement from sharing or publishing their work on internet or in educational journals. These advantages may also accrue to principals and educators in teacher education who collaborate with classroom teachers. In particular, the publication incentive may motivate teacher education faculty to introduce methods of self-recording and interpretations of such data to their undergraduate and graduate students.

The adoption of TODs with student self-recording may further the development of scientific knowledge in another way by providing a larger pool of teachers prepared to engage in more sophisticated research. The teacher who has used TODs with self-recording is better prepared for using more sophisticated designs; and the researcher who seeks to use more sophisticated research designs with teachers experienced with TODs faces a much easier task than the researcher who seeks to introduce such designs with teachers who are not so experienced. In other words, a lower initial hurdle into classroom recording with TODs argues for more sophisticated long-term research on theoretical explanations than would otherwise be the case. This may be regarded as a successive approximation approach to research design in which the teachers may be satisfied with the first approximation and may continue on into more sophisticated research designs as appropriate occasions arise.

In addition, when teachers discuss data, they are more likely to talk in more specific terms than they would in the absence of any data to refer to. Explaining or interpreting a child's performances on a graph-in-view engenders a more objective way of talking than the often far vaguer references that are resorted to in the absence of data. Such ways of speaking should also be picked up by student doing the self-recording, and some of it may be passed on to parents.

Finally, the nature of the self-recorded data that is collected with TODs may encourage more variations in instructional treatment than would otherwise occur. With an emphasis on interpretation, more data may be collected on the basis of "Let's try it and see what happens." From the teacher's perspective, this may mean, "Let's try it this way and see if similar performances occur more efficiently" or "Let's try it that way and see if greater effects in performances result." These "trys" or probes may or may not occur with a return to the previous way of doing things. If the change results in a sharp departure from the original performance trend, the teacher may return to the previous way, as in an ABA design, in order to see if a causal interpretation for the change appears more justified. If convinced a better way has been found, however, some teachers may not wish to deprive students of its perceived advantages for even a few days.

Although there are advantages to coupling self-recording to TODs, there are also complications. Barlow et al. (1984) point out that self-monitoring and other treatment ingredients interact:

> For the research goals of scientist-practitioners, the reactivity of self-monitoring must be considered in making appropriate interpretations of results. If a scientist-practitioner is

conducting single-subject research in which one measure is self-monitored data, any changes from A (baseline) conditions to B (treatment) conditions must be interpreted as due to an interaction between self-monitoring and any other treatment ingredient. (pp. 112-113)

This interaction must be considered in determining whether it is worthwhile to plan for a more traditional baseline with treatment (AB) design. For example, if a teacher were to provide a baseline by doing all the recording of data on student performance before introducing a particular instructional treatment with student self-recording, the effects of the instructional treatment would interact with the self-recording. The performance during treatment may well trend upward from the baseline, but this may be largely attributable to the effect of the self-recording—an effect that we already expect to be there. Under these conditions using the baseline data as a reference is not particularly informative and the cost in collecting the baseline data may not be justified. In addition, if the original treatment-with-self-recording is used as a baseline when a subsequent unplanned-for treatment is introduced, then the baseline can be expected to be a rising one; and departures and returns to baseline need to be interpreted against a rising baseline.

Although these complications impose some restrictions on interpretations, they do not eliminate them. The following provides some examples of the inferences that may be generated from using TODs with self-recording.

Examples

The following illustrates ways in which TODs with self-recording can provide inferences that support or suggest a hypothesis as well as inferences that question or disconfirm a hypothesis. In addition, examples are given of TODs leading to more sophisticated designs that are initially unplanned.

<u>Support for self-recording</u>. In Studwell and Moxley (1984), the progress of a kindergarten class had been recorded by the teacher during the first semester. This was recorded as the number of students who had accomplished certain preassigned goals. In the second semester, this was changed to have the children self-record their own individual progress on checklist graphs in addition to the teacher's class graph of the number of children reaching an achievement. In

February, the children recorded their progress in reciting the ABCs, naming the numerals from 1 to 10, saying the month and day of their birthday, and saying where they lived and how to get there. The results show a dramatic upward trend. In March, a different set of tasks was recorded: naming the uppercase letters, naming the lowercase letters, saying a word that rhymed with a word presented to them, and identifying the left and right sides of their body. Another dramatic upward trend occurred. See Figure 1.

Insert Figure 1 about here

The results suggest the benefits of student self-recording. The self-recording was initiated as a TOD to which the teacher's original data was later attached. In effect, the teacher's previously collected data in the first semester served as a baseline for the self-recording treatment. Introducing the tasks in different sets in February and March also served to create multiple baselines, one for each set. The initial treatment in the first semester served as an unplanned baseline for the self-recording treatment in the second semester and strengthens a causal attribution for self-recording. Such an interpretation is also supported by the literature on self-control, which has reported gains in self-recorded performances for individuals and for groups of individuals in particular skill areas (e.g. McLaughlin, 1976; O'Leary & Dubey, 1979; Rosenbaum & Drabman, 1979).

Support for self-recording with graphs rather than tables. In Miller (1997), second graders recorded their progress in the number of words they wrote daily in connected text, usually stories, during an approximately 15 minute allotment of time. At first, the record was kept as a table in each child's writing journal, in which the date and the number of words written was entered. Later, this information was recorded in the form of a line graph. The graph made any change in performance much more conspicuous and allowed a class graph to be displayed in which changes in the overall class performance would be visible from almost anywhere in the classroom.

Insert Figure 2 about here

This graph shows a fairly stable baseline was attained when the record-keeping was done with the tables, at which point the number of words written by the children showed little evidence of either increasing nor decreasing. After the children switched to recording their performances on individual graphs, however, the number of words that the children wrote steadily increased.

Support for improving writing by increasing the rate of writing. Figure 3 shows the rate of self-recorded freewriting in four different classrooms. In as much as the dates of the weeks are not identical for all of the studies, which were done in different years, the difference in weeks is indicated by their approximate order in the respective studies. In each classroom, the observed quality of the writing in terms of expressiveness and concrete detail improved.

Insert Figure 3 about here

A comparison of the multiple TODs in Figure 3 suggests that some explanations for the increases in writing are more plausible than others. The relatively rapid increase in writing rates for the second grade class indicates a gradual maturation effect is an improbable explanation for the increase in writing because, no matter what baseline writing rate might have occurred, a sharp change in writing rate must have taken place during the treatment period. The third grade class shows a slower rate of writing performance than might be expected from a comparison with the first and second grades. This slower rate may be explained by 1) the optional opportunity for writing during an open activity period in the morning in which the children could select from various academically-related activities—unlike the specific writing periods set aside in the other grades; the 2) lack of response consequences that were provided in comparison to the other grades; and 3) the lack of tightly timed writing except for a brief period discussed below.

The comparisons between four different classrooms in this study support the use of various ways to increase children's rate of writing and thereby their experience in writing. Research with children already indicates that higher writing rates are an indication of a higher quality of writing (e.g., Deno, 1986; Deno, Mirkin, & Marston, 1980; Deno, Marston & Mirkin, 1982; Marston, 1989; Myklebust, 1965; Slotnik, 1972; Van Houten, Morrison, Jarvis, & McDonald, 1974/1988; Videen, Deno, & Marston, 1982), and this study supports and extends that finding to a wider variety of settings. More particularly, a comparison of the different studies shows that timed-writing for speed achieved higher rates per day than allowing children more time to write at a more leisurely pace. In addition, writing with weekly group incentives in the second grade produced a higher rate of writing than writing without such incentives in the first and third grades. This suggests that increases in the first and third grade may be due proportionately more to the effect of self-recording while the effects in the third and fourth grades may be due proportionately more to timed writing and more frequent incentives. In as much as the quality of writing generally improved in all the treatments, the limit at which increased writing rates would interfere with improved quality of writing appears to be well above the normal freewriting rates of these children.

The data from comparing these separate treatment-only designs leads to the hypothesis that, with student self-recording, almost any practical way of increasing children's freewriting rates beyond their normal freewriting rates offers a simple, convenient way to improve the quality of their freewriting. In addition, the timed-treatments tended to produce more freewriting overall than the treatments in which more time overall was allowed. Although the quality of the freewriting was not compared between the different treatments, it would be of educational value to see if timed freewriting produced a quality of writing that was equal to or superior to the untimed freewriting. It would also be of value to see if a higher writing rate resulted in improved quality of writing in comparison to writing the same or a similar number of words at a slower rate. In other words, is it merely the quantity of words written that matters or does a higher rate of writing contribute to improved quality.

Introduction To More Advanced Single-Case Designs

Once self-recording is established, a new treatment may be introduced after the original treatment even though this is not originally planned for. If this new treatment shows promise, i.e., an increasing trend of improving performance, it may be continued. In effect, this establishes an AB or baseline-treatment design (see Figure 1 and Figure 2). If the treatment after baseline is repeated with different components, this can be considered as a multiple baseline design (see Figure 1).

If the old treatment is returned to after the new treatment has been pursued for a time, this establishes a reversal, ABA or baseline-treatment-baseline, design. For example, the third grade data in Figure 4 provides an example of how a new treatment can be introduced within an existing treatment. The third grade classroom used a tight time constraint of 5 minutes for freewriting during ten days that covered three weeks in February. This period may be considered as the introduction of a different treatment in place of the original treatment. In effect, the original treatment functioned as a baseline for contrasting the effects of the new treatment. Although the students had been recording in their journals for several months by this time, it is apparent that these third graders were writing at a rate that was far below the writing rate they were capable of. The timed writing also showed more expressive detail. On return to the original treatment condition with loose time constraints, some students returned to their previous rate and manner of writing, but other students wrote more words with more expressive and extended detail than before the timed intervention.

Insert Figure 4 about here

In what has now become in effect an ABA design, rapid writing under a tight time constraint is shown as a likely influence for a substantial increase of freewriting. The original treatment (freewriting with a loose time constraint) now becomes a baseline for the new treatment (freewriting with a tight time constraint). Return to the original treatment becomes a return to baseline. Such treatments within treatments may be introduced as probes within an original TOD.

Questioning a hypothesis

Only one contrary instance is needed to contradict a hypothesis or to require its modification, and such instances may readily occur in TODs as well as in traditional case studies.

Phonological stage theory. For example, some studies of spelling development found a sequence of distinct stages in which the development of phonological strategies based on the sounds heard in words precedes the development of visual strategies based on what was seen in words (e.g., Gentry, 1981, 1982, 1987; Henderson, 1985). These stages were presented as applying to all children: "All children can learn to be competent spellers but they do so over time and in developmental stages" (Henderson, 1985, v), and these stages were presented as well defined and distinct: "Research in how children develop skill in spelling shows us that young people's writing moves through clearly defined stages" (Gentry, 1981, p. 378). For example, in the phonetic stage, " Letters are assigned strictly on the basis of sound, without regard for acceptable English letter sequence or other convention of English sound correspondences" (Gentry, 1982, p. 193). Examples such as EGL for eagle and TIP for type illustrate this stage, which is a mapping of letter-sound correspondences. It is not until the next stage—the transitional stage—that what Gentry terms as visual strategies occur: "Transitional spellers present the first evidence of a new visual strategy; the child moves from phonological to morphological and visual spelling" (Gentry, 1982, pp. 196-197). Examples of such visual spelling include letter reversals such as HUOSE for house.

In a TOD, Moxley and Joyce (1990) presented evidence from computer printouts on the spelling development of four children using computer word processing that would require the abandonment or revision of some of these statements:

> In various ways, the children showed a considerable degree of visual spelling development along with phonological spelling development. For example, the children produced a substantial amount of visual spellings along with semiphonetic spellings. Although some overlap would be understandable between adjacent stages like the phonetic and transitional stages, the more extensive the overlap between visual and early phonological strategies, the weaker the claim that phonological strategies precede visual strategies. The children also wrote standard visual spellings both before and after semiphonetic spellings for individual words. And some of their individual spellings developed without any evidence of phonological strategies. In fact, there was no period of time among the

> children in which "letters are assigned strictly on the basis of sound," much less did "transitional spellers present the first evidence of a new visual strategy." These results suggest that stage description of spelling development have provided more accurate detail on phonological spelling development than on visual spelling development. Any claim that phonological strategies precede visual strategies needs qualification. (pp. 11-12)

Although the spelling development of many individual words did follow the spelling stage sequence indicated in Gentry (1981, 1982, 1987) and Henderson (1985), many words did not. Some words spelled by the same child showed one stage while other words were more than one stage away. For example, on the same day, one student wrote semiphonetic FLR/flower and HRSS/horse, a phonetic HAVN/having that includes a representation of the short vowel sound, and transitional visual reversals in AEFFTR/after, SOWN/snow, SOWNING/snowing, and SOWNMAN/snowman. In addition to different stages appearing on the same day, some stages did not appear in the spelling development of some words over time. For example, although one child showed spellings of different stages at the same time, the sequence of spellings for it over a four month period was IT, IS, IT, IT, IT, THC, and then 12 standard spellings of IT. The influences for creative variations appear to come from other word forms (i.e., is and the) rather than phonological spellings.

Such findings are reasonable. Even a phonological letter-sound strategy requires some visual skill in selecting the letter for the sound, and the computer makes the visual features between words read and words written far more similar and salient than they can be in children's handwriting. Although this study involved the changes in spelling of many words by four children, even the data from one child in spelling one word over time could be sufficient to require some modification—or raise some doubt—regarding the unqualified claim that phonological spelling precedes visual spelling. Such evidence—which may lead to questioning, modifying, or abandoning a previous claim or hypothesis—is potentially able to occur in any TOD just as it is potentially able to occur in any case study.

Theory that literacy instruction in schools should be delayed. A later study with children using computer word processing showed how multiple TODs from separate designs can also provide information as to whether a treatment can contribute to accumulating effects from one year to the next; and more than one indicator(multiple measurements within the same design) may contribute to the explanatory account. Figure 5 compares children who wrote for one year on the computer from 1992 to 1993 (Moxley, Warash, Coffman, Geres, Roman, & Terhorst, 1994) with children who wrote for two years on the computer from 1993 to 1995 (Moxley, Warash, Coffman, Brinton, & Concannon, in press) using similar word processing programs. During the first semester of the three-year-old class, the children also used early literacy programs on the computer, but they were not asked to make (or invent) their own spellings. Accordingly, the record of spelling accuracy for the 1993-1995 study begins in February of the three-year-old class. The spelling proficiency (in terms of the percent of accurately spelled letters in words) of the 1993-95 children in the four-year old class with two years of experience on the computer surpassed the spelling proficiency of the 1992-93 children in the four-year-old class with one year of experience on the computer.

Insert Figure 5 about here

If instruction in literacy as a three-year old leads to improved literacy performances as a four-year-old, this raises a question as to whether literacy instruction for young children should be delayed. The view that formal instruction in literacy should be delayed until a child has a mental age of $6^1/2$ years or so has been supported by some developmental psychologists and early childhood curriculum professionals although little empirical evidence supports this view (see Hanson & Farrell, 1995, p. 913). Delay of early literacy instruction can be questioned from different directions. One direction is whether early instruction in literacy indeed has detrimental long-term effects on children's reading achievement, interests, and general attitude toward school—which Hanson and Farrell's (1995) evidence questions. Another direction is whether young children are indeed able to show improvements in early literacy skills after a feasible amount of instruction and whether there is any retention of these skills; and the above study with three and four year olds provides evidence of learning and retention. The evidence from these two directions supports questioning the assumption that early instruction in literacy in schools should be delayed.

Support for early instruction in literacy.

The results shown in Figure 5 also support Vygotsky's (1935/1978) argument to transfer the teaching of writing to the preschool years, which is more feasible with the use of computers which do not require handwriting skills. In other words, it supports the hypothesis that literacy may be acquired earlier by children if conditions support an earlier acquisition of literacy, and this gain need not disappear in later years (cf. Hanson & Farrell, 1995). The utility of this hypothesis is further enhanced by the fact that the Hanson and Farrell study also found higher grades, better attendance, and less remedial instruction in both elementary and secondary school by those who received formal reading instruction in Kindergarten. The views and data from Hanson & Farrell (1995), Moxley, Warash, Coffman, Brinton, & Concannon (in press), and Vygotsky (1935/1978) suggest that withholding literacy instruction from children 3-5 years of age—not introducing it—is developmentally inappropriate.

There is an apparent anomaly, however, in the spelling data for the children in the 1993-95 study when they were in the three-year-old class. The 1993-95 children start out with a greater spelling proficiency in the three-year-old class in February than the 1992-93 children do in the four-year-old class in September, but improvements in the spelling of the 1993-95 children from February through April in the three-year old class do not appear. How is this to be explained? One way this can be done is through the use of another indicator or multiple measures within a design.

The relation between different indicators can suggest explanations of treatment effects that would be less evident if only one indicator were used. Figure 5 shows two indicators in the 1993-95 study: The original indicator is for spelling accuracy from February in the three-year-old class through the following year to April in the four-year-old class. The additional indicator, which was recovered from the collected data, is for the percent of words written in stories (as defined by writing consisting of connected text in the form of one or more sentences)—in contrast to words written in titles or labels (in the form of a name or list of names)—as a percent of all writing from February through April in the three-year-old class. The indicator on story writing provides a plausible explanation as to why the spelling accuracy of the 1993-95 children in the three-year-old class did not improve from February through April: the children were spelling more difficult words, such as verbs and function words, that were not among the list of words with pictures that the children copied in displaying pictures on the screen.

The children first entered (typed) the names of pictures into the computer by selecting and copying them from a printed guide. The pictures for these names would then appear on the screen at locations selected by the child. The student monitor would then ask the children to tell them about their pictures. The monitors wrote down what the children said and dictated this back for the children to type in by themselves. When the children came to writing (or typing) about their pictures (without the printed guide), they had an advantage in spelling the names of the pictures because they had typed those names in earlier when they entered them into the computer in order to display pictures. Some of the children's writing might consist entirely of a list of names (or labels) for the pictures that they had typed in earlier. However, other parts of speech such as verbs and function words would not have been typed in earlier, and the children would be writing such words for the first time that day. Consequently, the more that children wrote "stories" with independent clauses, the more the children had to write parts of speech they had not seen and copied earlier in the session. Thus, the spelling of words in stories was likely to be more difficult than spelling a title or a list of labels. Encountering words that were more difficult to spell would be a plausible explanation for why improvements in spelling proficiency did not appear during the final three months of the three-year-old class. This suggestion is rendered more convincing by the increasing trend of percent of words written in stories.

Summary

TODs with self-recording were presented as a means of introducing measurement on a frequent and routine basis into the public school classroom. It was argued that practitioner measurement has been an important historical ingredient in linking technology and science in a collaborative manner to the mutual advancement of both. In education, coupling TODs with student self-recording contributes to furthering the connection between classroom practice and educational research. A single TOD may be sufficient to question an existing hypothesis or to suggest a new hypothesis, and variations of TODs may readily allow interpretations that can become increasingly convincing as TODs accumulate.

The routine collection of such data for TODs provides a direct avenue leading from the more tentative interpretations of TODs to the more decisive interpretations of higher level experimental research. Teachers may also collaborate with other professionals in sharing or publishing their data and findings. As teachers acquire a greater facility in rendering data-based accounts of their own, it is reasonable to expect they will also become more receptive to data-based research interpretations in general. This effect may extend to students and parents.

References

Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). The scientist practitioner: Research and accountability in clinical and educational settings. New York, NY: Pergamon.

Barlow, D. H., & Hersen, M. (1984). Single case experimental designs (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Bennett, J. A. (1986). The mechanics philosophy and the mechanical philosophy. History of Science, 24, 1-28.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. The Behavior Analyst, 19, 163-197.

Bisesi, T. L., & Raphael, T. E. (1995). Combining single-subject experimental designs with qualitative research. In S. B. Neuman & S. McCormick (Eds.), Single-subject experimental research: Applications for literacy (pp. 104-119). Newark, DE: International Reading Association.

Bobbitt, F. (1913). Some general principles of management applied to the problems of city-school systems. The twelfth yearbook of the national society for the study of education: Part I. The supervision of city schools (pp. 7-96). Chicago, IL: University of Chicago Press.

Bonstingl, J. J. (1992). Schools of quality: An introduction to total quality management in education. Alexandria, VA: Association for Supervision and Curriculum Development.

Breznitz, Z. (1987). Increasing first grades' reading accuracy and comprehension y accelerating their reading rates. Journal of Educational Psychology, 79, 236-242.

Caratello, J., & Caratello, P. (1990). Beginning charts, graphs & diagrams. Huntington Beech, CA: Teacher Created Materials.

Cavior, N., & Marbotto, C. M. (1976). Monitoring verbal behavior in a dyadic interaction: Valence of target behaviors, type, timing, and reactivity or monitoring. Journal of Consulting and Clinical Psychology, 44, 68-76.

Choate, L. D., & Okey, J. K. (1989). Graphing primer: Grades K-2. Palo Alto, CA: Dale Seymour Pub.

Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. School Psychology Review, 15, 358-374.

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. Exceptional children, 48, 368-371.

Deno, S. L., Mirkin, P. K., & Marston, D. (1980). Relationships among simple measures of spelling and performance on standardized achievement tests (Research Report No. 22). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Dukes, W. F. (1965). *N*=1. Psychological Bulletin, 64, 74-79.

Gallie, W. B. (1966). Pierce and pragmatism. New York, NY: Dover.

Gentry, J. R. (1981). Learning to spell developmentally. The Reading Teacher, 34, 378-381.

Gentry, J. R. (1982). An analysis of developmental spelling in GNYS AT WRK. The Reading Teacher, 36, 192-200.

Gentry, J. R. (1987). Spel ... is a four-letter word. Portsmouth, NH: Heinemann.

Hanson, N. R. (1958). Pattern of discovery. Cambridge, England: Cambridge University Press.

Hanson, R. A., & Farrell, D. (1995). The long-term effects on high school seniors of learning to read in kindergarten. Reading Research Quarterly, 30, 908-933.

Hawkins, R. P. (1989). Developing potent behavior-change technologies: An invitation to cognitive therapists. The Behavior Therapist, 12, 126-131.

Hawkins, R. P., & Hursh, D. E. (1992). Levels of research for clinical practice: It isn't as hard as you think. The West Virginia Journal of Psychological Research and Practice, 1, 61-71.

Henderson, E. H. (1985). Teaching spelling. Boston, MA: Houghton Mifflin.

Hynes, M. C. (1995). IDEAS: NCTM Standards-Based Instruction: Grades K-4. Reston, VA: National Council of Teachers of Mathematics.

Joyce, B., & Moxley, R. A. (1988). August Dvorak (1894-1975): Early expressions of applied behavior analysis and precision teaching. The Behavior Analyst, 11, 33-40.

Kaestle, C. (1993). The awful reputation of educational research. Educational Researcher, 22, 23-31.

Kazdin, A. E. (1974). Reactive self-monitoring. The effects of response desirability, goal setting, and feedback. Journal of Consulting and Clinical Psychology, 42, 704-716.

Kazdin, A. E. (1982). Single-case research designs: Methods for clinical and applied settings. Oxford, England: Oxford University Press.

Lee, M., & Miller, M. (1993). Great graphing: Grades 1-4. New York, NY: Scholastic.

Lee, V. L. (1994). Organisms, things done, and the fragmentation of psychology. Behavior and Philosophy, 22, 7-48.

Leming, J., & Hollifield, J. (1985). Cooperative learning: A research success story. Educational Researcher, 14 (3), 28-29.

Lindsley, O. R. (1991). Precision teaching's legacy from B. F. Skinner. Journal of Behavioral Education, 1, 253-266.

Lindsley, O. R. (1996). The four free-operant freedoms. The Behavior Analyst, 19, 199-210.

Lipinski, D. P., & Nelson, R. O. (1974). The reactivity and unreliability of self-recording. Journal of Consulting and Clinical Psychology, 42, 118-123.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), Curriculum-based measurement: Assessing special children (pp. 18-78). New York: The Guilford Press.

McLaughlin, T. F. (1976). Self-control in the classroom. Review of Educational Research, 46, 631-663.

Miller, K., Kasper, S., & Moxley, R. A. (1997). How teachers can introduce their students to self-recording. Manuscript in preparation.

Morgan, M. (1984). Reward-induced decrements and increments in intrinsic motivation. Review of Educational Research, 54, 5-30.

Moxley, R. A. (1974). Formative and Non-formative Evaluation. Instructional Science, 3, 243-283.

Moxley, R. A. (1979). Subjective, Individual, and Aggregate References in Educational Research. Instructional Science, 8, 169-205.

Moxley, R. (1989). Some Historical Relationships between Science and Technology with Implications for Behavior Analysis. The Behavior Analyst, 12, 45-57.

Moxley, R. A. (1994). Continuous Improvement of Instruction: Instructional Design with Frequent Measurement and Positive Evaluation (PreK-4). Morgantown, WV: NextPrint.

Moxley, R. (in press). Treatment-only designs and student self-recording as strategies for public school teachers. Education and Treatment of Children.

Moxley, R. A., & Joyce, B. G. (1990). Early visual spelling on the microcomputer. Journal of Computing in Childhood Education, 1, 3-14.

Moxley, R. A., Kenny, K. A., & Hunt, M. K. (1990). Improving the instruction of young children with self-recording and discussion. Early Childhood Research Quarterly, 5, 233-249.

Moxley, R. A., Lutz, P. A., Ahlborn, P., Boley, N., & Armstrong, L. (1995). Self-recorded word counts of freewriting in grades 1-4. Education and Treatment of Children, 18, 138-157.

Moxley, R. A., & Su, B. (1994). Continuous improvement in writing with a high-frequency, wide-ranging record of progress. The West Virginia Journal of Psychological Research and Practice, 3, 49-60.

Moxley, R. A. Warash, B., Coffman, G., Geres, L. M., Roman, T., & Terhorst, L. E. (1994). Computer writing development in a prekindergarten Class of 4 Year Olds. Journal of Computing in Childhood Education, 5 (2), 211-229.

Moxley. R. A., Warash, B., Coffman, G. Brinton, K., & Concannon, K. R. (in press). Writing development using computers in a class of three-year-olds. Journal of Computing in

Childhood Education.

Myklebust, H. R. (1965). Development and disorders of written language: Vol. 1: Picture story Language Test. New York: Grune & Stratton.

Nelson, C. M., & McReynolds, W. T. (1971). Self-recording and control of behavior: A reply to Simkins. Behavior Therapy, 2, 594-597.

Nelson, R. O. (1977). Methodological issues in assessment via self-monitoring. In J. D. Cone & R. P. Hawkins, (Eds.), Behavioral assessment: New directions in clinical psychology. New York, NY: Brunner/Mazel.

O'Brien, J. A. (1926). Silent reading. New York, NY: Macmillan.

O'Leary, S. G., & Dubey, D. R. (1979). Applications of self-control procedures by children: A review. Journal of Applied Behavior Analysis, 12, 467-485.

Overholt, G. E., & Stallings, W. M. (1976). Ethnographic and experimental hypotheses in educational research, 5 (8), 12-14.

Peirce, C. S. (1931-1963). Collected papers of Charles Sanders Peirce (C. Hartshorne & P. Weiss, Eds., 8 vols.). Cambridge, MA: Belknap.

Pennypacker, H. S., Koenig, C. H., & Lindsley, O. R. (1972). Handbook of the standard behavior chart. Kansas City, Kansas: Precision Media.

Preddy, M. (1993). Introduction to Graphing: Grades K-1. Greensboro, NC: Carson-Dellosa Pub. Co.

Qin, Z., Johnson, D. W., &, Johnson, R. T. (1995). Cooperative versus competitive efforts and problem solving. Review of Educational Research, 65, 129-143.

Rosenbaum, M. S., & Drabman, R. S. (1979). Self-control training in the classroom: A review and critique. Journal of Applied Behavior Analysis, 12, 467-485.

Sidman, M. (1960). Tactics of scientific research. New York, NY: Basic Books.

Sieck, W. A., & McFall, R. M. (1976). Some determinants of self-monitoring effects. Journal of Consulting and Clinical Psychology, 44, 958-965.

Slavin, R. E. (1991). Cooperative learning and group contingencies. Journal of Behavioral Education, 1, 105-115.

Slotnick, H. B. (1972). Toward a theory of computer essay grading. Journal of Educational Measurement, 9, 253-263.

Studwell, P., & Moxley, R. (1984). Self-recording in kindergarten: A study in naturalistic evaluation. Psychology in the Schools, 21, 450-456.

Van Houten, Morrison, E., Jarvis, R., & McDonald, M. (1988). The effects of explicit timing and feedback on compositional response rate in elementary school children. In B. Sulzer-Azaroff, R. M. Drabman, R. D. Greer, R. V. Hall, B. A. Iwata, & S. G. O'Leary (Eds.), Behavior analysis in education 1968-1987 (pp. 287-295). Lawrence, KS: Society for the Experimental Analysis of Behavior. (Original work published 1974)

Vargas, J., & Moxley, R. (1980). Teaching for thinking and creativity: The radical behaviorist's viewpoint. In A. E. Lawson (Ed.), The Psychology of Teaching for Thinking and Creativity: The Seventh Yearbook of the Association for the Education of Teachers in Science (pp. 268-295). Columbus, OH: Association for the Education of Teachers in Science.

Videen, J., Deno, S. L., & Marston, D. (1982). Correct word sequences: A valid indicator of proficiency in written expression (Research Report No. 84). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Vincenti, W. G. (1990). What engineers know and how they know it. Baltimore, MD: The Johns Hopkins University Press.

Vygotsky, L. S. (1978). The prehistory of written language. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), Mind in society (pp. 105-119). Cambridge, MA: Harvard U. Press. (Original work published 1935).

Washburne, C. (1922). Educational measurement as a key to individual instruction and promotions. Journal of Educational Research, 5, 195-206.

West, R. P., & Young, K. R. (1992). Precision teaching. In R. P. West & L. A.

Hamerlynck (Eds.), Designs for excellence in education: The legacy of B. F. Skinner (pp. 113-146). Longmont, CO: Sopris West.

Yin, R. K. (1994). Case study research: Design and methods (2nd ed.). Newbury Park, CA: Sage.

Captions for Figures

Figure 1. Self-recording in kindergarten.

Figure 2. Self-recording with graphs rather than tables.

Figure 3. Daily mean average number of words written by each student for grades 1-4.

Figure 4. Effect of timed writing on daily mean average of number of words written by each student in the third grade.

Figure 5. Progress in the story writing and spelling accuracy of the children in a two-year 1993-95 study compared to the spelling accuracy of the children in a one year 1992-93 study.

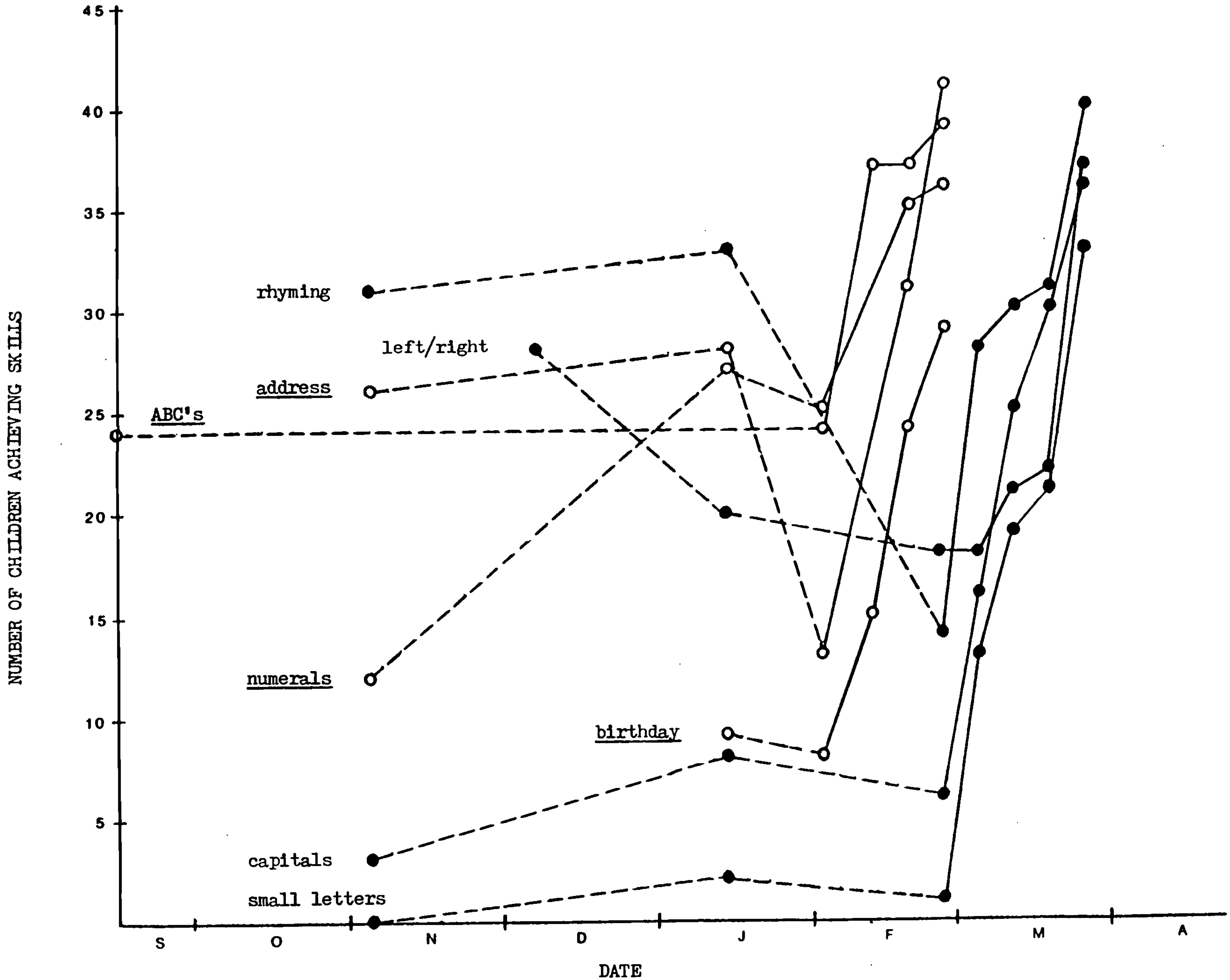


Figure 1. Self-recording in kindergarten.

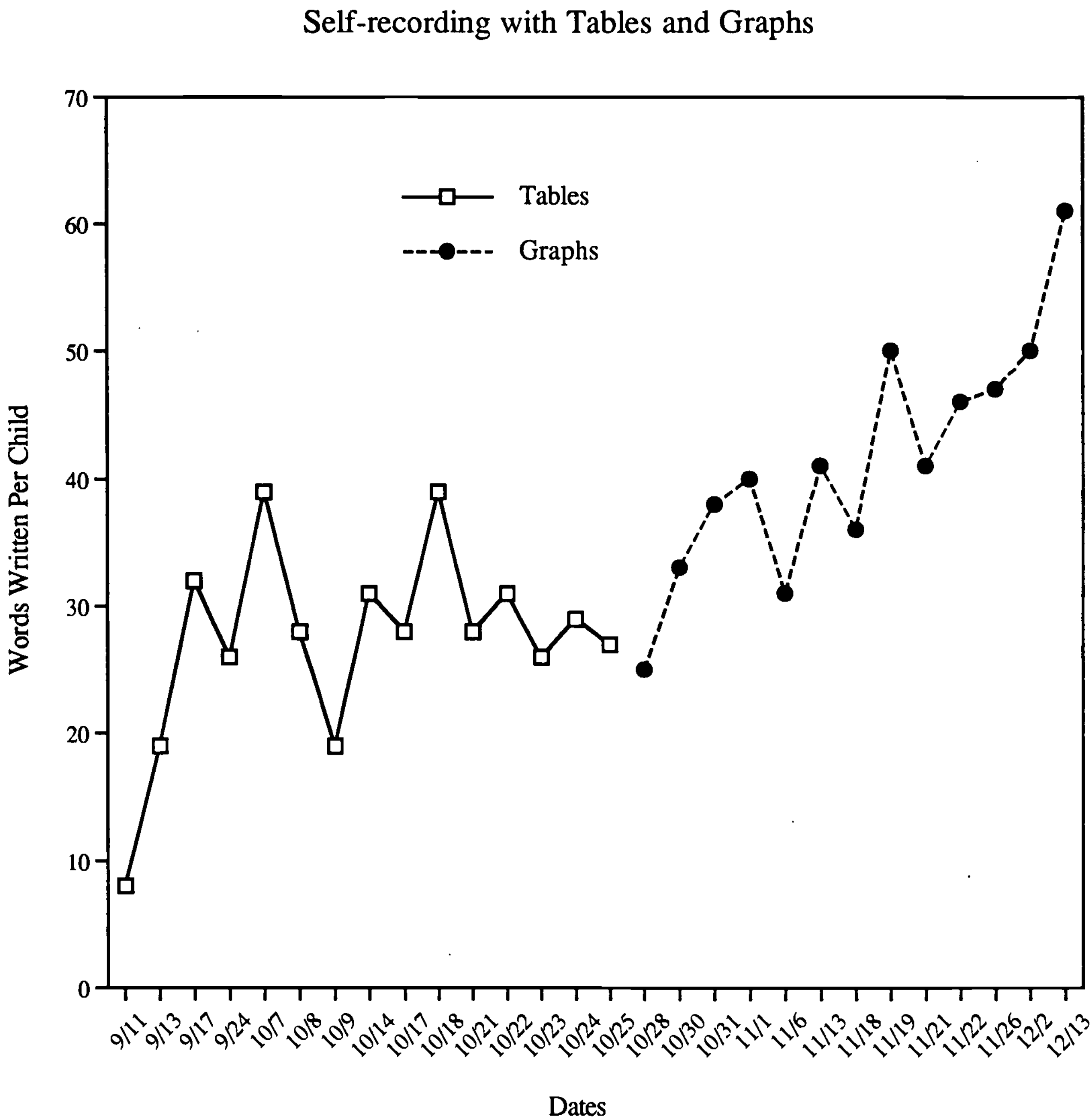


Figure 2. Self-recording with graphs rather than tables.

Daily Average Number of Words Written by Each Student Grades 1-4

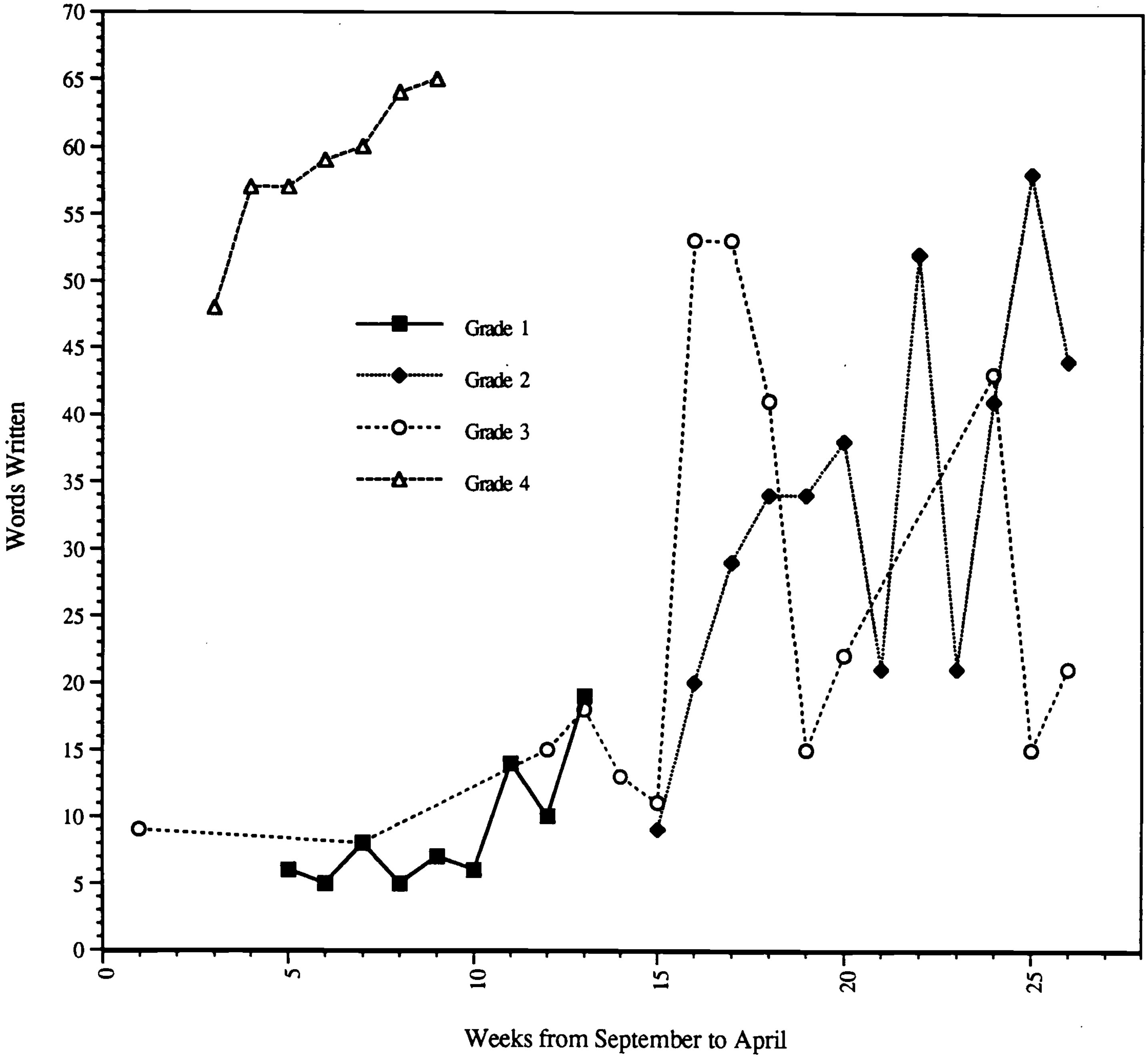
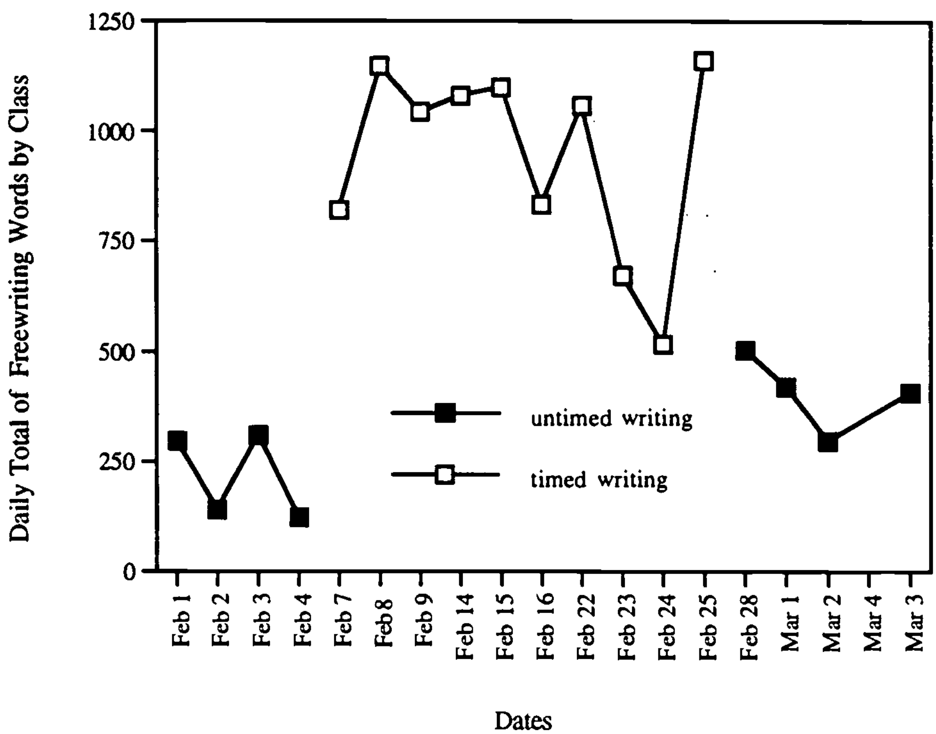


Figure 3. Daily mean average number of words written by each student for grades 1-4.

Effect of Timed Writing on Number of Words Written



<u>Figure 4.</u> Effect of timed writing on daily mean average of number of words written by each student in the third grade.

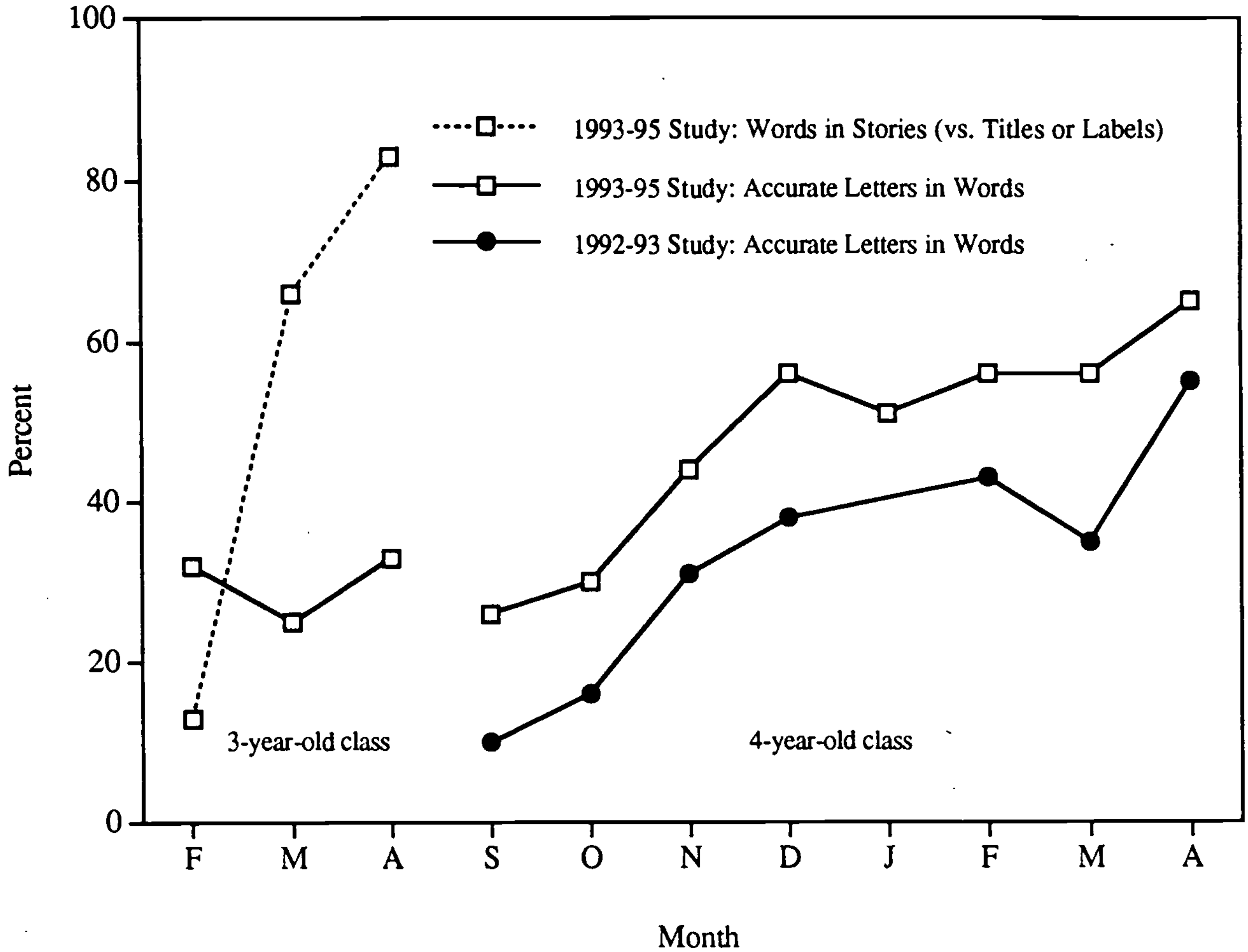


Figure 5. Progress in the story writing and spelling accuracy of the children in a two-year 1993-95 study compared to the spelling accuracy of the children in a one year 1992-93 study.

AREA 1997

TM027327

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Treatment-Only Single-Case Designs with Student Self-Recording for the Teacher as Researcher	
Author(s): Roy A. Moxley	
Corporate Source: West Virginia University	Publication Date: May 16, 1997

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

or here
Permitting reproduction in other than paper copy.

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Roy A. Moxley	Position: Professor of Education
Printed Name: Roy A. Moxley	Organization: West Virginia University
Address: 604 Allen Hall, West Virginia University, Morgantown, WV 26506	Telephone Number: (304) 293-3442
	Date:

OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 21, 1997

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae2.educ.cua.edu.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (523)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: AERA 1997/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

---

[1]If you are an AERA chair or discussant, please save this form for future use.