

1994

Looking at learning : alternative assessment in the kindergarten

Joan A. Damm
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Damm, Joan A., "Looking at learning : alternative assessment in the kindergarten" (1994). *Master's Theses*. 829.
DOI: <https://doi.org/10.31979/etd.hy5p-n4qb>
https://scholarworks.sjsu.edu/etd_theses/829

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700, 800/521-0600



Order Number 1359012

Looking at learning: Alternative assessment in the kindergarten

Damm, Joan A., M.A.

San Jose State University, 1994

Copyright ©1994 by Damm, Joan A. All rights reserved.

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106



LOOKING AT LEARNING:
ALTERNATIVE ASSESSMENT IN THE KINDERGARTEN

A Thesis

Presented to

The Faculty of the Division of Teacher Education
San Jose State University

In Partial Fulfillment
of the Requirements for the Degree

Master of Arts

by

Joan A. Damm

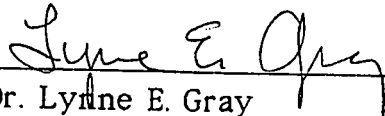
August, 1994

© 1994


Joan A. Damm

ALL RIGHTS RESERVED

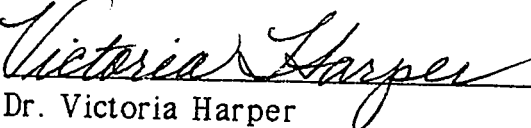
APPROVED FOR THE DIVISION OF TEACHER EDUCATION



Dr. Lynne E. Gray




Dr. Beverly Jensen



Dr. Victoria Harper

APPROVED FOR SAN JOSE STATE UNIVERSITY



ABSTRACT

LOOKING AT LEARNING: ALTERNATIVE ASSESSMENT IN THE KINDERGARTEN

by Joan A. Damm

Assessment practices can impact motivation, self-esteem, and achievement. Concern is growing that traditional assessment practices encourage the teaching of discrete, low-level skills and discourage critical thinking. Standardized testing is being criticized for driving instruction but neither taking recent research into account nor informing classroom practice.

This research is a descriptive case study of an attempt to implement alternative assessment techniques in a middle to high SES California kindergarten. The following techniques are explored: portfolio assessment of journal writing, assessment of verbal output patterns, performance assessment of individual and group problem solving skills, and performance assessment of word decoding skills.

Results confirm that alternative assessment techniques can be rich, informative, and motivating; they can both enhance and document the growth of learning over time. However, results also confirm that alternative assessment is time- and labor-intensive, complex, and expensive; it will demand new forms of support for teachers at all levels.

Acknowledgements

I would like to acknowledge the following people who aided me, some knowingly and some unknowingly, in the completion of my thesis:

My mother, May Snell Damm, who set the example;

Elizabeth Kurtich, my seventh grade teacher, who read to our class every day and taught me the love of reading; and Maude Willard, my high school English teacher, who taught me to write (teachers' teachers both; also both deceased--I hope they knew how much they influenced my education, both in the short run and in the long run);

My brother Jerry, who played with my data on his computer and who, with his wife, Suzy, engaged me in substantive conversation about assessment;

Karen Yamamoto, a new friend and neighbor, herself a graduate student, who gave me the encouragement to continue when I was ready to quit;

The people at my school, especially Jerry Krumbein, who was principal during my first thesis attempt and once took my yard duty so I could work on my project; my team partner Karen Hayes, who is unerringly supportive under all conditions; Kay Morshead and Mary Martin, who were Kindergarten aides during the first attempt; Bunny Dawson, who kept inquiring after my progress; Carol Day and Peg

Anderson, who gave me a reason to persevere; Libby Nickel, who helped me with video camera technicalities on my second thesis attempt; parents Sharon Kelley and Kim McCarthy, who shared the frustrations and the humor of video-taping; the parents of my class, who graciously gave their permission; Anne Creevy, who practices authentic assessment in her classroom; and Michele Sanders, who understands why it took me so long.

Also, the following people at San Jose State University:

Dr. Evelyn Neufeld, who gave me an excellent introduction to Piaget;

Dr. Beverly Jensen, the head of the Teacher Education Division, who offered to help and who served on my committee;

Dr. Victoria Harper, who also served on my committee;

Dr. Marty Krovetz, who offered valuable suggestions;

Dr. Clay Andrews, who, in retrospect, gave me the best advice I got ("Just write the damned paper!") and, finally, with special gratitude, to the two Dr. L. Grays who gave such a pleasing circularity to my academic life--Lillian, who gave me the confidence I needed to begin, and Lynne, who gave me the support I needed to finish.

Table of Contents

1) Chapter I. Statement of the Problem.....	Page	1
2) Chapter II. Review of the Literature.....	Page	10
3) Chapter III. Methodology.....	Page	43
4) Chapter IV. Results.....	Page	64
5) Chapter V. Discussion.....	Page	110

List of Figures

- 1) Figure 1 Very Important Person Word Count
Average Number of Words per Entry
by Trimester. Page 69
- 2) Figure 2 Very Important Person
Average Word Production by Gender. . . . Page 71
- 3) Figure 3 Year's Very Important Person
Word Production Graphs. Page 72
- 4) Figure 4 People Pieces (Pretest) Product Page 78
- 5) Figures 5.1 and 5.2 Teacher's People Pieces
Scaffolding Product. Pages 79 & 80
- 6) Figures 6.1 - 6.3 Group IV's Toad and Rabbit
(Posttest) Product and
Details from It. Pages 81 - 83
- 7) Figures 7.1 - 7.3 Purple and Green Squares
(Pretest) Products. Pages 92 - 94

List of Figures (Continued)

8) Figure 8 Teacher's Two-color Squares		
	Scaffolding Product.	Page 95
9) Figures 9.1 - 9.4 Blue and Orange Squares		
	(Posttest) Products.	Pages 96 - 99
10) Figures 10.1 - 10.8 Student-generated Nursery Rhyme Reconstructions.		Pages 102 - 109

List of Tables

- 1) Table 1 Comparison of Ways Individuals Sorted
on Group Sorting Products (People Pieces
[Pretest] and Toads and Rabbits [Posttest]). . . .Pages 85 & 86

- 2) Table 2 Children Compared to Themselves across
Individual Squares Pre- and Posttests. Pages 89 & 90

Chapter 1

INTRODUCTION TO THE PROBLEM

It has become increasingly apparent that many American children are graduating from high school unprepared to contribute to the democratic process (Shapiro, 1991), compete in the international marketplace (O'Neil, 1991), or lead lives satisfying to themselves (Dentzer, 1991). In most recent elections in the United States, voter turnout has been disturbingly low; U.S. businesses complain that entry-level workers lack necessary basic skills (Butler, 1985); American workers are losing their competitive edge to workers of Japan and Germany; and, for many U.S. citizens, the standard of living is declining (Dentzer, 1991).

As Americans attempt to solve these problems and focus more attention on the education of their children, there are increasing calls for school accountability and school restructuring (i.e., for schools to change the way they go about doing business). One area that is receiving considerable attention from the restructuring movement is assessment. Educators are becoming increasingly concerned that the way children are evaluated has a negative impact on curriculum, on instructional strategies, and on the children themselves.

The roots of formal education and the evaluation of achievement are in apprenticeship systems. Apprenticeships were conducted in real-life settings, and assessments were performance-

based with meaningful consequences. For example, a canoe builder taught his child to build canoes. Assessment of the work was straightforward: either the canoe met the requirements of floating, remaining upright in the water, staying watertight, and being easy to navigate, or it did not. Likewise, the child of an island fisherman, whose livelihood depended on the sea, was taught to navigate by the stars. Here again, the assessment was straightforward: either the child could use the positions of the stars to get from island to island and home again or he could not. Similarly, girls were taught to cook and sew by performing those tasks until they achieved mastery-- that is, until the family was satisfied with the product.

The Medieval Guild system provides another example of meaningful education and performance-based assessment. A child was apprenticed to a master craftsman and worked closely under him until it was determined that the apprentice had learned enough through observation and practice to become a journeyman. After a period of time in which the journeyman perfected his skills, in order to become a master, he designed and created a "master piece" that was evaluated by experts in his field.

Performance-based assessment is, in fact, ancient. An early example of it can be found in the Bible in Judges 12:5-6. When a group of Gilead guards wanted to ascertain whether or not those trying to cross the Jordan River after they were escaped fugitives from the tribe of Ephraim, they asked each, "Art thou an Ephraimite?" If the respondent answered, "Nay," he was told, "Say

now Shibboleth." If he was unable to pronounce the word correctly and said "Sibboleth," he was slain.

As the nations of Europe and North America became industrialized, education became more industrialized too and moved out of the hands of family members and master craftsmen and into the hands of general practitioners in separate buildings called schools. Formal education became increasingly abstract and removed from practical concerns--from building boats, for example, to reading about how other people built boats. Likewise, assessment changed. It went from complex, hands-on, performance-based, personally meaningful activities that involved immediate feedback and important consequences to activities that tend to be more simplistic, abstract, indirect, remote, and in many cases, inert (i.e., no action results). In other words, the assessment of learning has evolved from situations in which children produced objects or performed actions in natural settings, to situations in which children are evaluated in artificial settings. Instead of being required to produce original ideas, writing, products, or performances, they are evaluated on their knowledge of the ideas, writing, products, and performances of others.

Educators are questioning this system and the impact of this type of assessment. They are expressing concern that current curricula and the way achievement is measured in schools today may be contributing to the poor performance of U.S. students in both national and cross-national research studies (Wiggins, 1991).

On the kindergarten level, assessment issues have centered on the inappropriateness of standardized testing for young children; the questionable validity and reliability of some developmental tests commonly used for kindergarten screening; and the inappropriate use of tests for exclusion from kindergarten, tracking into two-year kindergarten programs, or retention/promotion decisions. The authors of Here They Come, Ready or Not! The Report of the School Readiness Task Force (California State Department of Education, 1988) recommend that the assessment of young children rely heavily on teacher observation and parental input, that standardized tests be used only rarely, that on those rare occasions that they are used they meet American Psychological Association standards of acceptability, and that the law in regard to using tests to place minority students be rigorously observed. Developmental screening (as opposed to readiness testing) is recommended only for the purposes of identifying children with special needs, such as those with poor vision or hearing, so that intervention may begin early.

The National Association for the Education of Young Children, in a publication entitled Developmentally Appropriate Practice in Early Childhood Programs Serving Children from Birth through Age 8 (1987), stresses that no important decision about a child should be made solely on the basis of one test, that children's progress should be assessed primarily through observations recorded at regular intervals, and that results should be used to improve and individualize instruction.

As traditional forms of classroom assessment, particularly standardized testing, have come under fire, new forms of assessment, termed "authentic" and "alternative," have appeared. Authentic assessment proponents such as Wiggins (1991) draw sharp contrasts between standardized testing and authentic assessment.

Whereas standardized testing is done to the student by someone else for someone else's purposes, he says, authentic assessment is done by the student in collaboration with interested others for the purpose of improving his or her own learning. The validity of standardized tests is dependent upon the secrecy of test content; authentic assessment goals are up front and public.

Testing is a "one-shot deal." Authentic assessment is on-going. Testing often involves discrete, lower-level thinking skills such as the memorization of facts and surface level involvement with subject matter. Authentic assessment presupposes the integration of curricular areas, deep involvement with subject matter, and higher-level thinking skills such as synthesizing, analyzing, and evaluating.

Standardized testing involves "filling in the bubble" and reporting someone else's knowledge. Authentic assessment involves the production of new knowledge and displaying it through a performance, an exhibition, or original discourse.

Standardized testing involves one-way "conversation" and inauthentic questioning (i.e., asking a question for which one already has the answer). Authentic assessment utilizes authentic questioning and substantive conversation (i.e., lengthy joint truth-seeking in

which each person is genuinely interested in what the other has to say) (Wiggins, 1991).

With standardized testing, one form of reporting is used (often inappropriately) as feedback for different audiences. Proponents of authentic assessment recognize that different audiences need different forms of feedback.

Standardized testing deals with the knowledge of facts, but ignores other important components of learning. Authentic assessment attempts to chronicle level of interest, motivation, effort, learning style, perseverance, and transfer of learning to times and places outside the school setting.

Standardized testing may or may not be aligned with the curriculum being taught and, once a standardized test has been taken and sent off for scoring, it is "out of sight, out of mind." Often, results come back too late to inform classroom decisions. Authentic assessment is an integral part of the curriculum being taught and involves repeated reflection.

Standardized testing may either give teachers a false sense of security or be insensitive to the progress students are making. Authentic assessment shows what students can do as opposed to what they can not do.

The purpose of this thesis is to examine the current literature on the impact of assessment practices on school students; to consider the pros and cons of standardized testing and of alternative forms of

assessment; and, finally, to explore the use of specific alternative assessment techniques with kindergarten students.

The following questions are explored:

1. Can kindergarten students' "writing" growth be reliably and validly documented using a portfolio system that allows children to choose their own best journal writing?

2. Does entry length in experience-story production provide a useful indicator of growth in literacy?

3. Does working in groups provide a valid and reliable format for assessment of individual growth in mathematical problem solving?

4. Can assessment of individual children's nursery rhyme reconstruction provide a teacher with useful data on the link between oral memory skills and reading and on how children go about beginning to decode words?

5. Does the quality and usefulness of the information gained in alternative assessment situations justify the time it takes to gather and analyze the data?

These questions will be examined in the following manner:

Children will write in journals daily. Once a week, the teacher will meet individually with each child to help the child choose a "best" page for his or her "keeper" journal. The teacher will reflect on the time it takes to choose "keeper" journal pages and on the usefulness of the information gained.

The teacher will write down a story dictated by each child each week. The stories will be typed into a Macintosh SE/30 computer so that a running record of each child's story dictation is kept. Each child's words will be counted using the SE/30's word count feature and results graphed for individuals and for the class. The teacher will reflect on the usefulness of assessment information gained in this manner.

Individually, children will attempt to color sub-divided squares as many different ways as possible using only two colors. The results will be tabulated, after which the children will be randomly assigned to groups of four or five children each. Each group will work together to create a product displaying as many different squares as possible, after which the groups will be videotaped presenting their products to the class. Out of view of the class, the teacher will also make a Squares product and will be videotaped presenting her product to the class as a means of scaffolding between pre- and posttest phases of the project. Then a second group of Squares products will be created by the groups and their presentations of their second products will be videotaped. Products will be examined to ascertain whether children are able to find more of the Squares possibilities the second time they do the project over the first, whether they are better able to follow directions the second time, and whether they are better able to persevere the second time over the first. In addition, the videotapes will be examined to ascertain whether the videotaping of the presentations of these

products allows for the documentation of growth in understanding, in audience involvement, and in presenter competence. A People Pieces/Toads and Rabbits project will follow a similar format with a similar videotaping procedure.

The teacher will discover a nursery rhyme that each child can say from memory. Then the teacher will watch each child reconstruct in a pocket chart the rhyme he or she knows by heart. The teacher will ascertain whether this technique allows her to gain helpful assessment information on the link between oral memory skills and reading and on how children go about beginning to decode words. At the conclusion of these projects, the teacher will reflect on whether the quality and usefulness of the information gained through the use of these alternative assessment techniques justified the time it took to gather and analyze the data.

Chapter 2

REVIEW OF THE LITERATURE

Whether one considers the impact of traditional classroom assessment practices on American students (Crooks, 1988) or the ways they have been affected by standardized testing (Ellwein, et al., 1988; Fredericksen, 1984; Natriello, 1987), it becomes readily apparent that there is a growing concern that traditional assessment practices have been detrimental to students. American educators are beginning to realize that they have been caught up in an accountability system which measured what was easy to measure--discrete, low-level skills--instead of complex intellectual performance (Valencia, 1991; Wiggins, 1991). Ultimately, what was assessed determined in large part what was taught, and both elementary and secondary curricula came to be driven largely by standardized tests. As a result, students and their parents have come to accept an incomplete view of achievement and of what it means to be an educated person (Eisner, 1991; Gardner & Hatch, 1989; Wiggins, 1991).

Gardner (1983) pointed out that educators need to be much broader about what they assess and much more flexible about how they assess it. Assessments should recognize the complex, longterm, non-linear nature of learning; they should respect the cultural backgrounds, varied learning styles, and individual differences among learners; and they should assess multiple intelligences rather

than relying solely on the measuring of linguistic and logico-mathematical capabilities. Assessment and evaluation should be integral parts of instruction, interwoven with it so intimately as to sometimes become unobtrusive (Wiggins, 1991). At its best, assessment should be something that students do for themselves (or in concert with interested others) to improve their own learning. This view of assessment, however, is a far cry from the kinds of assessments that have often been practiced on students in American classrooms.

How Classroom Evaluation Practices Impact Students

Educational researchers have been interested for a long time in how evaluation impacts students. It is interesting to note that a criticism of the way students are evaluated which is widely acknowledged today was virtually ignored when voiced by Miller (1978) sixteen years ago:

We say we want sensitive, analytic, independent scholars, then treat them like Belgian geese being stuffed for pate' de foie gras. We reward them for compliance rather than independence, for giving the answer we have taught them rather than challenging the conclusions we have reached...

(Miller, cited in Crooks, 1988).

In 1988, Crooks reviewed the literature on student evaluation and found that evaluation impacted students in numerous ways, both positive and negative. On the positive side, he reported that

evaluation can sometimes help students consolidate prior knowledge and focus on important aspects of a subject; it can provide corrective feedback and give students a feeling of accomplishment. However, he also reported that there are many ways in which evaluation can affect students negatively (Crooks, 1988).

For example, researchers have investigated both the cognitive levels of questions teachers ask and the influence question level has on student learning. The use of higher-level questions enhances interest, development of learning skills, retention, and transfer (Crooks, 1988); but for the most part, teachers in elementary through university level ask questions at Bloom's lowest ("knowledge") level --that is, questions that depend on the memorization of facts (Stiggins, Griswold, Green et al., cited in Crooks, 1988).

Marton and Säljö found that students tend to use either deep or surface approaches to their learning. Deep approaches involve an active search for meaning and underlying principles and the connecting of ideas; surface level approaches involve memorizing facts as if they have no relation to each other (1976a). Many students study primarily to do well on examinations; if they perceive that the teacher expects them to memorize facts in order to do well, they are willing to memorize, even when they recognize that this approach interferes with understanding the subject (1976b).

Crooks (1988) stated that several researchers in the studies he reviewed differentiated between task goals and ego goals. The definition of a task goal is that it has clear criteria for success;

students working toward task goals believe that they are responsible for their own success and that the result is not preordained. Ego goals, on the other hand, depend on doing better than someone else. Crooks (1988) concluded that if all students are to be encouraged to learn, conditions that favor task goals over ego goals are desirable. Norm-referencing discourages weaker students and discourages students from helping each other (Deutsch, 1979).

Quality of evaluation appears to be bound up with kind of learning environment. Johnson, et al. (1981) conducted a meta-analysis of 122 studies that examined the comparative effects on student achievement of any two or more of the following structures: competitive, individualistic, and cooperative. They came to the conclusion that cooperative structures generally contribute to higher student achievement than either of the other two, that they lead to increased cohesiveness and interpersonal attraction among group members (Johnson, Johnson, & Maruyama, 1983), and that they are more enjoyable than individual learning for most students (Johnson & Johnson, 1985). Corno and others (Corno & Mandinach, 1983; Corno & Rohrkemper, 1985) found that self-regulated learning activities fostered intrinsic motivation, and that intrinsic motivation in turn encouraged students to be more independent learners. Maehr & Stallings (1972) found extrinsically motivated students to be more answer-oriented than intrinsically motivated students. Whereas the intrinsically motivated tended to use deeper, more meaningful approaches to understanding tasks, the extrinsically

motivated tried to take shortcuts to get to the answers. Deci (1975) noted that the key factor in dampening intrinsic motivation appears to be whether evaluation is used to control behavior or to give helpful feedback on progress in learning. Friedman (1987) speculated that repeated opportunities to attain standards may increase motivation and the feeling of self-efficacy while reducing test anxiety. Repeated opportunities to meet standards also encourage students to help each other, success on a task is more likely to be attributed to effort rather than to ability, and intrinsic motivation is fostered (Ames, 1984).

Several studies have shown that students with high test anxiety do much better on the same cognitive tasks under less stressful conditions (Hill, 1984; Hill & Wigfield, 1984). Hill and Wigfield recommend that students be given very generous time limits, that tasks be given that allow for success by all students, and that letter grades not be given in elementary school.

Schunk (1984, 1985) recommends that evaluation emphasize performance rather than task engagement; that credit be given for quality of work, not merely for handing it in; and that feedback involve informing students of their progress toward mastery, not comparison with other students.

According to the above researchers' conclusions, the following would appear to constitute good learning environments that foster good assessment: conditions that involve self-regulated, intrinsically motivated students working in co-operative groups; repeated

opportunities for attaining standards; generous and flexible time limits; feedback on progress toward mastery rather than credit merely for handing assignments in; self-comparisons rather than comparison with other students; and success for all students.

That view of optimal learning and evaluation conditions, however, is at variance with what Natriello (1987) perceived when he reviewed the literature on the impact of evaluation on students. He found that because most classroom tasks are complex, teachers tend to avoid open-ended tasks. This, he says, simplifies the evaluation process and makes life in the classroom more secure for both teachers and students (since both teachers and students must justify their grades), but it also makes it less likely that students will engage in discovery-oriented or creative tasks. Crooks (1988) notes:

Many observers have commented on the contrast between the broad enthusiasm for learning demonstrated by most children in the first year or two of schooling and the jaded approach of many older students. Although some of this difference may relate to developmental factors, it is hard to escape the conclusion that for many students schooling tends to lower rather than increase interest in learning (p. 464).

Crooks made the following recommendations based on their being supported by several different areas of the research he reviewed: Deep learning should be a central goal of education, which

requires that educators emphasize understanding, thinking skills, and transfer of learning to unfamiliar situations. Too much emphasis is placed on grading, and too little on helping students learn. Crooks says, "It is hard to see any justification before the final year or so of high school for placing much emphasis on using classroom evaluation for normative grading...[given the] undesirable consequences for most students" (p. 468). Greater emphasis should be placed on giving useful feedback to students and on final competence; less should be placed on mistakes made along the way. For optimal learning, standards and tasks should be individualized and there should be considerable flexibility in learning routes.

Too often, students in U.S. classrooms are teacher-, grade-, and test regulated rather than self-regulated; they graduate as a ritual rather than because they are motivated to learn; and they work independently instead of in cooperative groups. Too often, a subject is "covered" and teacher and class move on instead of students having repeated opportunities for attaining standards; time limits are neither flexible nor generous; and requirements for success and test content are kept secret rather than being given to students at the outset. If the discrepancy between optimal evaluation conditions and what actually occurs in many American classrooms is as wide as it appears, it is not difficult to see how classroom evaluation practices may be harming students.

How Standardized Testing Impacts Students

Standardized testing has contributed to our educational problems as well. Although the impact of evaluation on students has been studied for many years, most of the research has dealt with testing, often with how two tests (not necessarily designed for the same purpose) have compared with each other (Natriello, 1987). Traditionally, this research has been based on a psychometric model that has relied heavily on statistics and has been conducted by people other than classroom teachers (Gibboney, 1991; Maier, 1991). Perhaps partially for these reasons, teachers as well as school administrators have been found to have woefully inadequate assessment backgrounds (Hills, 1991; Stiggins, 1991). According to Stiggins, even those who have been trained in assessment are ill-prepared with respect to assessment concepts and procedures needed to address the ongoing assessment demands of the classroom. He also says that traditionally, assessment training has been based on a very narrow definition of assessment that has failed historically to meet the needs of teachers.

It has been only comparatively recently that the concept of testing as a means to evaluate student progress has been called into question. Concurrently, there is an increasing call for educational research that is more qualitative in nature and that is done by educational practitioners taking into consideration the complexity of the classroom situation (Gibboney, 1989; Maier, 1991). A third, rather anomolous recent development has been the concurrent rise

of two seemingly contradictory thrusts: the increase in standardized testing and the alternative assessment movement.

The growth of educational testing over the past 40 years or so has been explosive (Frederiksen, 1984). Standardized test scores have been used to determine who should be promoted to the next grade, who should graduate from high school, who should be admitted to college, and who should be certified as a teacher (Ellwein, Glass & Smith, 1988). They have determined which academic track a student would follow (Marzano & Costa, 1988) and which students should be admitted to gifted programs (Hoge, 1988). And, as Popham has put it, "Tests [have been] employed to make keep-or-kill decisions about educational programs. Big dollars...[have been] riding on the results of achievement tests" (Popham, 1983, cited in Frederiksen, 1984, p. 194).

The usefulness of objectively scorable standardized tests for mass testing was first demonstrated during World War I when the Army Alpha was used to classify military personnel. During the 1920s, the College Board began investigating the use of standardized tests for college admissions. In 1926, the first Scholastic Aptitude test (SAT) was administered. During the 1950s, Lindquist pioneered the development of highly-sophisticated test-scoring equipment. By 1984, the Educational Testing Service had machines that could score upwards of 10,000 answer sheets an hour, and almost all 50 states had legislation requiring some form of testing (Frederiksen, 1984).

During the 1980s, the trend was toward educational measurement on a national level. With increasing public outcry for school accountability, state school officials turned increasingly to national assessment to compare their students to students in other states (Linn, 1988), and as American workers have begun to lose their competitive edge to the workers of Japan and Germany, the comparison of students on an international level has increased. Americans test their children more than anyone else in the world: 46 million students from kindergarten through high school are given more than 150 million tests a year (Allis, 1991).

With the proliferation of standardized testing, however, a growing body of criticism of standardized tests has also developed. Critics claim that they are based on faulty underlying constructs (Gardner and Hatch, 1989; Hoge, 1988; Bussis & Chittenden, 1987; Cross & Paris, 1987) and outdated norms and psychometric models (Bock & Wood, 1971; and Lumsden, 1976, cited in Cross & Paris, 1987); that they are culturally biased (Boyle, 1989); that they lack validity because too many uncontrollable variables influence test scores (Boyle, 1989; Cross & Paris, 1987); that there has been an over-reliance on using tests to validate each other in a circular fashion (Cross & Paris, 1987); that they are poor predictors of either academic or occupational success (McCall, 1977); that they are questionable for use with young children (Cannella & Reiff, 1989; Elkind, 1989; Glickman & Pellegrini, 1989) or the learning disabled (Elrod & Sorgenfrei, 1988); that there is a mismatch between test

purposes and the uses to which they are put (Cross & Paris, 1987; Boyle, 1989); that they fail to take recent research on literacy acquisition and the importance of prior knowledge into account (Buckley, 1987; Bussis & Chittenden, 1987; Samway, 1987; Searle and Stevenson, 1987; Teale, Hiebert, & Chittenden, 1987); and that their results are used inappropriately in ways that are morally, socially, and politically untenable (Boyle, 1989; Hoge, 1988).

Gardner & Hatch (1989); Hoge (1988); Bussis & Chittenden (1987), Cross & Paris (1987), Schell (1988), and Wixson & Peters (1987) all believe that standardized tests are based on faulty underlying constructs, in the areas of intelligence, giftedness, and reading comprehension respectively.

Logico-mathematical intelligence, as its name implies, has to do with a propensity for logic and mathematics. Linguistic intelligence involves "sensitivity to the sounds, rhythms, meanings and functions of language." Whereas intelligence tests are based solely on verbal and logico-mathematical abilities, Gardner (1983) believes that in reality there are at least five other intelligences: musical, spatial, bodily-kinesthetic, interpersonal, and intrapersonal. Musical intelligence encompasses "the ability to produce and appreciate rhythm, pitch, and timbre and appreciation of the forms of musical expressiveness." Spatial intelligence has to do with "the capacity to perceive the visual-spatial world accurately and to perform transformations on those perceptions." Bodily-kinesthetic intelligence involves "the ability to control one's body movements

and to handle objects skillfully." Interpersonal intelligence includes "the capacity to discern and respond appropriately to the moods, temperaments, motivations, and desires of other people." And intrapersonal intelligence refers to "access to one's own feelings and the ability to discriminate among them and draw upon them to guide behavior" (Gardner & Hatch, 1989, p. 6). Not surprisingly, Gardner considers traditional intelligence tests inadequate for measuring his broader definition of "intelligence." (For Gardner's criteria for defining an intelligence, see Gardner, 1983.)

In a vein similar to Gardner's, Hoge (1988) finds fault with the use of intelligence tests for placing children in gifted programs. In the first place, he maintains, there are problems with the way giftedness is defined:

Many school boards have only a vaguely defined global concept of giftedness. Those that have an official definition usually incorporate such attributes as superior levels of intellectual potential, academic aptitudes, motivation levels, creativity, and leadership potential, but the actual selection of pupils for the gifted program may be based solely on the results of an individual I.Q. test such as the WISC-R, an instrument whose scores carry no connotations respecting motivation, leadership, or creativity (p. 13).

In short, traditional standardized measures made poor sorters for gifted programs even when educators believed that an optimal plan involved sorting students by ability and tracking them.

Bussis & Chittenden (1987), Cross & Paris (1987), Schell (1988), and Wixson & Peters (1987) all consider traditional standardized reading tests poor tests of reading comprehension. They claim that there is little correspondence between current theories of the reading process and assumptions implicit in the tests. Current theory views the reading process as involving the construction of meaning through the dynamic interaction of reader and text; tests, on the other hand, assume that comprehension can be measured by short passages unlike those the reader encounters in texts and through discrete details that fail to take the reader's prior knowledge into account.

Toward the end of a study of beginning readers, Bussis & Chittenden (1987) administered an oral reading inventory that was comparable to a test of reading comprehension and observed how children dealt with it. They found that some children attempted to link one item to the next in order to discover a story in the material, that prior knowledge interfered with the correct answering of some test items, and that children seemed to have an intuitive feel for their need to practice, an assumption totally ignored by reading tests.

Criticisms have also been made that have to do basically with honesty--honesty in the gathering of test data and honesty in reporting the scores to constituents. In 1988, in a study of five sites

(three state agencies that oversaw testing for high school graduation, college admission, and teacher certification respectively and two school districts that tested for elementary promotion and for high school graduation), Ellwein, Glass, and Smith (1988) found that when students were in danger of failing competency tests, a number of safety nets were strung. All five sites allowed students multiple chances to take the test; one allowed as many as eleven tries to pass. Three sites allowed alternate tests to be substituted, two allowed certain students to be exempt from taking the test, two allowed staff members to overrule test results, and two allowed standards to be lowered. And at two sites, when it was thought that raw scores needed to pass might be perceived by the public as being too low, standard scores instead of raw scores were reported. In short, these researchers found that competency tests function more as political gestures than as reforms (Ellwein, Glass, & Smith, 1988).

Perhaps the criticisms of standardized tests of most concern to educators are that they concentrate on a narrow range of discrete lower-level thinking skills (i.e., those skills that depend on the memorization of isolated facts); they drive curriculum and instruction, diverting teacher and student time and effort from higher-level thinking skills (those that require students to analyze, synthesize, and evaluate); and they lack penetration (i.e., they do not tell us very much about student strengths and weaknesses, about the processes students use to arrive at their conclusions, or about differences between students getting the same score); they tell us

even less about student interests, motivation, learning styles, perseverance, or tendency to transfer what they learn in school to their out-of-school hours. In short, they give us very little information that would help us to improve student learning.

Frederiksen (1984) tells one rather amusing story that exemplifies how testing can influence teacher and student behavior. During WWII, he was a staff member of a project that found that the best predictors of grades in gunner's mate school were scores on verbal and reading comprehension tests; given that a gunner's mate's tasks were to dismantle, repair, and reassemble guns, this finding did not seem logical to him. When Frederiksen and his group were sent to a gunner's mate school in Maryland to improve grading practices, they discovered that teachers were using the lecture method and tests were based on the lectures and on the gunnery manual. Frederiksen's group instituted performance tests that involved removing and replacing the extractor plunger on an anti-aircraft gun and adjusting the oil buffer on a Browning machine gun for maximum rate of fire. The instructors complained that the tests were too hard, and, under the circumstances, they were; most of the prospective gunner's mates failed. However, Frederiksen's group persisted with the performance tests, and eventually the word got around. The men started learning to take guns apart and put them back together, and the instructors took the lecturns out of their classrooms and brought guns in. Eventually, the performance tests became better predictors of gunner's mate grades than the tests of

verbal ability. Frederiksen pointed out that no overt attempt was made to change either teacher behavior or student behavior, but that when the tests changed, the behavior changed too.

When the stakes become high, measurement drives instruction. If teachers and programs are evaluated on the basis of test scores, and especially if program funding depends on them, some teachers are tempted to "teach to the test," perhaps with some justification. As Le Mahieu and Wallace (1986, cited in Linn, 1988, p. 8) have noted,

It is untenable to agree that
achievement is the product,
and that test scores are its measure,
and then assert "please don't
pay too much attention to the scores."

There are two main schools of thought on "teaching to the test." One school finds it morally reprehensible and sees it as cheating. The other sees it as the logical thing to do if the test reflects what is valuable to learn. As A. Lawrence Lowell said in *The Atlantic Monthly* in 1926 (cited in Frederiksen, 1984, p. 193),

To chide a tennis player for training
himself with a view to winning a match,
instead of acquiring skill in the game,
would be absurd, because the two things
are the same...if marks are not an adequate
measure of what the course is intended to

impart, then the examination is defective...

New Jersey has taken 'teaching to the test' seriously (Lichota, 1981). Detailed lists of specific skills measured by each test have been issued to teachers with encouragement to use the lists in their basic-skills teaching. According to the *New York Times*, (Lichota, 1981, cited in Frederiksen, 1984, p. 195), a school official in New Jersey said the schools "conducted workshops, distributed special work sheets, instructed students how to take the tests, and geared curriculum to the test content." Teaching to the test has gone too far. On one test, all fifty states reported being above the 50th percentile. This phenomenon has been dubbed "the Lake Wobegon effect," named after Garrison Keillor's fictional town in which "all the children are above average" (Beck, et al., 1991). If teaching to the test does not account for such a situation, then outdated norms do. Either way, standardized tests project an unrealistic view of student learning.

Elrod and Sorgenfrei (1988) feel that traditional assessment practices for mildly-handicapped youth need to be re-examined. According to these researchers, mildly handicapped youth meet their real challenges when they leave high school. Research has shown that young adults with learning disabilities have less proficiency than their non-disabled peers at such job-related skills as interviewing, explaining problems to supervisors, accepting criticism, and providing constructive criticism (Matthews, Whang, & Fawcett,

1980, cited in Elrod & Sorgenfrei, 1988). What mildly-handicapped adolescents need, claim Elrod and Sorgenfrei, is assessment of their career interests and skills, not standardized tests of academic achievement (Elrod & Sorgenfrei, 1988).

According to Glickman and Pellegrini (1989), standardized testing is also inappropriate for young children. These researchers found that how kindergartners interacted with their peers on the playground was a better indicator of first-grade readiness than a standardized test was. According to their study, the Metropolitan Readiness Test accounted for only 34 per cent of the variance in first-grade performance, whereas measures of social competence, popularity, and teacher ratings accounted for 58 per cent. The Association for the Education of Young Children has also come out with a position paper urging extreme caution in the use of standardized tests with children aged three to eight (*NAEYC Position Statement on Standardized Testing of Young Children 3 through 8 Years of Age*, 1988).

Standardized testing appears not to work well for the young child, the learning disabled, the gifted, or the "normal" student. Probably it does work, however, at least to an extent, for policymakers whose goal is to obtain a broad overview of student learning and for college admissions officers who are faced with sorting students for college eligibility; it may also work, if only as a subconscious survival technique, for teachers faced with large student/teacher ratios and/or multi-subject teaching situations.

Standardized tests are relatively easy to administer and to score; in some cases they can be given to large groups at once; they may be cost-effective when compared to establishing the conditions necessary for making alternative assessment viable; and they have the reputation, at least, for being objective. It may be partly for these reasons that the use of standardized testing has proliferated. Some educators, on the other hand, are beginning to suspect that, for the learner, standardized testing works poorly.

Frederiksen (1984) contends that part of the reason standardized testing works as poorly as it does is that problems found in standardized tests are well-structured, whereas problems in real life are ill-structured. In well-structured problems we know what the problem is, all the information needed to solve it is available within the problem, and there is an algorithm that, if applied correctly, guarantees a correct solution. In ill-structured problems, however, we do not know what the problem is; we do not have the information needed to solve it; we do not know what steps to take; and we do not know when we have the problem solved. Unfortunately, as Frederiksen points out, all the really important social, political, and scientific problems in the world are ill-structured.

Educators are beginning to realize that traditional assessment practices, especially an over-reliance on standardized testing, give children a false notion of the learning process and discourage the provision of complex, high-quality tasks that could produce good

thinkers. According to Wiggins (1991), "As a result of our grading practices and methods of evaluation, many children come to feel that they are, due to their genetic codes, either A students or C students." And, as a result, many feel they are incapable of high quality work and either disengage from school or drop out. Wiggins claims that the result of the failure to provide students with complex, high quality classroom tasks is that educators fail to provide opportunities for producing responsible, self-reliant, creative thinkers and problem solvers who have the capability to extend the knowledge base rather than merely parroting back the knowledge of others. As a result of these considerations, an increasing number of educators are calling for a shift away from standardized testing and toward forms of assessment that are more authentic.

What Is Authentic Assessment?

As traditional classroom evaluation practices and standardized testing have come under criticism and as the nature of student assessment has begun to change, "new" terms are appearing throughout the assessment and evaluation literature: authentic assessment, performance-based assessment, portfolio assessment, enhanced multiple-choice questions, rubrics, and holistic scoring.

Authentic assessment has to do with assessment done in the context of activities initiated by the student for the student's own purposes. Probably truly authentic assessment is possible only in real-life settings with immediate real-life natural consequences; it is

unlikely that authentic assessment is possible for all students in a school setting as schools are now constituted.

Wiggins (1991) and Valencia (1991) claim that authentic assessment [as the term is used in the context of the school setting] presupposes giving students complex, involving, high-quality tasks that replicate those done by competent adults. High standards must either be developed collaboratively with the students or given to them at the outset. Quality output must be demanded of students, and progress must be documented over time. Proponents of authentic assessment recognize that there are multiple models of excellence and that there are different entry points and numerous routes to the solution of a problem. They recognize the complexity both of the learning process and of the nature of most work that students will be expected to engage in as adults. Authentic assessment is flexible enough to allow for creative and original student work that no-one may have anticipated, and it encompasses a number of facets of learning that traditional assessment neglects, such as student interests, effort, motivation, perseverance, and the need to practice.

Traditionally, the coaching of sports and the teaching of music have involved authentic assessment. In both these cases, the subject matter is inherently valuable, motivating, and emotionally involving to the student; the task is complex and often collaborative; the student knows at the outset the goals and criteria for success; there is an adult end-state toward which the student is striving to become

increasingly competent; tasks are closely bound with ongoing assessment; the student is responsible for his or her own learning; teachers and students reflect collaboratively on progress; and progress is shown concretely and documented over time.

According to Wiggins (1991), if educators want to use authentic assessment (and he says they must if they are to educate students), they must decide what they value; create complex, collaborative tasks that incorporate the competencies those values imply; and support students in their efforts to make increasingly competent approximations toward those competencies. Educators, he says, must be unwilling to settle for simplistic, proxy tasks that waste student and teacher time, hide the complexity of the real task, and mislead stakeholders about student learning.

According to Wiggins (1991), authentic assessment is not an instrument; it is a concept that embraces the following assumptions:

- 1) the purpose of evaluation is to support learning
 - 2) all students can (and must) do high-quality work
 - 3) students must be responsible for their own learning and involved in the evaluation of their work
 - 4) real learning takes time
 - 5) evaluation is a normal part of the learning process and, in the school setting, should be intricately interwoven with and inseparable from curriculum and instruction
 - 6) the process one goes through to learn is as important as the product one produces
-

- 7) growth in learning occurs over time and can be documented
- 8) collaboration with others and reflection on one's work are necessary components of the learning process
- 9) assessment should be multi-faceted; not only should there be many ways to document the same ability, but many different facets of learning should be documented
- 10) good authentic assessment is impossible without good, rich curricula and learning environments

(for this list I am indebted to Grant Wiggins, *Bridge to Change* conference, January 19, 1991)

Performance-based Assessment

Performance-based assessment refers to the assessment of learning based on the actions of the student, preferably in the context of a complex, high-quality learning task. When a violinist auditions for a part in a symphony orchestra, either she is able to play the chosen piece well enough to be accepted by the orchestra or she is not. This assessment is authentic because it is done in the context of real work and because it bears real and important consequences. It is also performance-based, because it depends on the actions of the performer.

An assessment can be performance-based without being authentic (writing, for example, that is viewed as process but that is done under timed test conditions). Educators should also be aware

1) that "new" terms are often confused and interchanged in the literature and that care must be taken to ascertain what an author means by authentic, alternative, performance-based or portfolio assessment and 2) that an educator's own assessments should be, as much as possible, alternative to standardized testing, performance-based, and as close to authentic as is possible in a school setting. Such assessments will involve giving students choices, providing high quality tasks, and breaking down classroom walls.

Enhanced Multiple-Choice Questions

Enhanced multiple choice questions ask students to do something more than just make a choice. They are more challenging than straight multiple choice questions and require students to solve a problem, infer from a graph, or synthesize information rather than merely recognizing a correct answer. Sometimes students are also asked to write down what they were thinking as they were making their choices.

Holistic Scoring

Holistic scoring has to do with looking at the overall effect of a piece of writing or the performance of a task rather than giving points for discrete details. Rubrics are established procedures and sets of criteria for scoring student performance (Freedman, 1994). According to Wiggins, rubrics should not be developed for a task until one has collected a number of student products and ascertained what attributes characterize various levels of student performance (Wiggins, 1993). Benchmarks are performance samples that serve as

standards against which other samples may be judged. The benchmarks most familiar to most elementary school teachers are those of the old penmanship books, in which comparison samples were given for handwriting that was below average, about what could be expected, and above average for the grade level.

Portfolio Assessment

Portfolio assessment has to do with the documentation of progress over time and, like authentic assessment, is a concept, not a particular instrument. Thinking of a portfolio as simply a folder of student work misses both the point of the concept and the richness of its potential. The portfolio takes its name from the notion of an artist's, model's, or architect's portfolio. Whereas a professional portfolio, however, contains representative samples of the artist's best work and is designed to show a prospective employer the artist's range of accomplishments, the portfolio in an educational setting includes work that represents an on-going process. Therefore, it may contain not just the student's best work (until, perhaps, the end of the year), but rough drafts and work that the student does not like as well as work that he or she does. Portfolio assessment necessitates a classroom atmosphere that encourages the role of student as constructor of his or her own knowledge and the role of teacher as collaborator and co-learner. Student choice and involvement are important, because the ultimate goal is to make students thoughtful evaluators of their own work. Therefore the teacher will not be the sole arbiter of what goes into the portfolio.

Often the student will decide; sometimes the teacher will decide; and sometimes the two will decide together. Measurement over time is also a salient characteristic of portfolio assessment, so student work should be sampled at the beginning of the year and at regular intervals thereafter.

Portfolio assessment uses a variety of techniques. There is no one "right" way to create a portfolio. The literature is replete with ideas for what might be included in one. Tape recordings; videotapes; checklists; samples of student writing; student comments and self-assessments; literary logs; photocopies of what children are able to read at various times in the year; teacher comments and anecdotal records; interest inventories; children's drawings; parent input; write-ups of student projects or open-ended investigations; photographs; enhanced multiple-choice questions; and laser disks are all candidates for inclusion in a student's portfolio.

Since reflection is such an important part of portfolio assessment, regular times should be set aside for students to review the contents of their portfolios--by themselves, with peers, with parents, or with teachers. The idea is for students to reflect on what they have done in the past and to project what they might do in the future, and in so doing become increasingly competent both at evaluating their own work and in taking responsibility for their own learning.

Portfolio assessment is based on the idea that children show progress over time, and that that progress can be documented. It

concentrates on what students can do as opposed to what they can not do, and it involves process as well as product. Portfolio assessment may have particular implications for children who score poorly on tests, giving them a better chance than traditional testing to show what they can do. Simmons (1989) compared writing samples taken from portfolios to samples collected from a large-scale writing test. He found that portfolio assessment provided a more complete and accurate picture of student writing growth. In his study, average and above-average children scored about the same on their portfolio papers as on their tests, but students who scored the lowest on the test scored significantly better on the writings in their portfolios (cited in Jongsma, 1989).

Portfolio assessment for the lower grades is still in its embryo stage. More research is needed to determine how to make portfolio assessment work for young children in self-contained classrooms. One major problem that has already come to light is the length of time it takes. Terry Johnson (1990) and his colleagues, who have been working for some time on a checklist of emergent literacy behaviors, postulate that the greatest number of items a teacher can be expected to monitor consistently is six--and that for only fifteen students and only in the area of language arts. Another deterrent to portfolio assessment and to other forms of authentic assessment is that they will necessitate a shift in the way teachers "teach." Teachers will have to give up the comparatively comfortable and

familiar role of "sage on the stage" and take on the less familiar and more demanding role of "guide on the side" (Fitzpatrick, 1991, p. 20).

Ill-Structured Problems

In addition, if the idea is to concentrate on higher-order thinking skills, it must be presupposed that teachers know what the higher order thinking skills are and that there is a way to teach them. In the words of Frederiksen (1984), "Parents are not likely to insist that their children be taught to solve ill-structured problems if they have not heard of such problems. Even if they did, schools would not know what to do about it, because they have not heard of them either" (Frederiksen, 1984, p. 199). While that statement may be less true in 1994 than it was in 1984, it is still true that creating good learning tasks and good assessments--and especially the conditions and environments that foster them--is difficult, expensive, labor-intensive, and time-consuming.

Authentic Assessment and Reporting to the Wider Community

There is also the problem of reporting to parents and the wider community. At first, educators will be feeling their way with authentic assessment, which will necessitate the willingness to take risks and a climate that allows them to try and fail and try again. In an era in which parents are crying for accountability and school boards are facing budgetary crises, such atmospheres are hard to come by. As Buckley has said, "The process of change involves a constant consideration of theory in light of reality" (Buckley, 1987, p. 746).

New Forms of Assessment Being Tried

In spite of the difficulties involved, new forms of assessment are currently being tried in various formats and in different parts of the country. One innovative authentic assessment project for young children is Project Spectrum, initiated at Tufts and Harvard Universities in 1984. Based partially on Gardner's (1983) theory of multiple intelligences, Project Spectrum attempts to identify young children's distinctive cognitive and stylistic strengths through various activities in a rich classroom environment. Spectrum's premise is that every child has the potential to develop strengths in one or more content areas, and that it is the educational system's job to discover those strengths and nurture them. Children are assessed throughout the year. In addition to linguistic and mathematical abilities, Spectrum's battery assesses musical, scientific, social, spatial and mechanical skills and stylistic features such as curiosity, level of confidence, and persistence. Assessments range from relatively unstructured to quite structured, and documentation takes a variety of forms, including portfolios, tape recordings, and observational checklists. Children are involved in their own evaluations and given time to reflect. Krechevsky (1991), who is project director, describes Spectrum as "a developmentally appropriate alternative based on a broad view of the mind" (p. 44).

In "Assessment as Theatre: Staging an Exposition," Thomas Barone describes the efforts of the Rural Educational Alliance for Collaborative Humanities (REACH) program to assess South Carolina

high school students through involving them in researching the history of their county. Students go out into the community to gather artifacts, do library research, and interview residents. They also have access to computer networks that allow them to talk to university professors, state and local agencies, community experts, and students and parents across South Carolina. Students work collaboratively, and time is flexible. After doing their research, they showcase what they have learned by creating their own original media presentations, short stories, plays, poems, portfolios, and musical performances, which they present to the public at an exposition (Barone, 1991).

In the future, technology may help make the documentation of authentic student learning possible. At Conestoga Elementary School in rural Wyoming, students use a multi-media system consisting of a CD-ROM drive, computer, scanner, optical drive, and laser printer to document their work and transfer it onto their own individual laser disks. The scanner can be used to transfer student writing quickly; videos of children performing such actions as climbing a rope in PE or throwing a pot on the potter's wheel can be transferred to the child's disk as well and kept as a permanent part of achievement records. Large amounts of information can be added to or retrieved from a disk as often as necessary, yet the disk is small enough to fit easily into a student's permanent file (Campbell, 1992).

Many schools, however, have neither the requisite equipment nor the expertise to institute an assessment program like Conestoga's.

Perhaps the best way for most classroom teachers to embrace the authentic assessment movement is to start slowly, choosing one aspect of authentic assessment and attempting to implement it.

Regardless of the version of alternative assessment that schools attempt to implement, writers of recent assessment literature caution that problems are to be expected. Alternative forms of assessment are time-consuming and labor-intensive. They involve new attitudes, new paradigms, and changes in the roles of both teachers and students. Teachers will be trying to institute new forms of assessment while being judged publicly by standardized tests. As educational systems attempt to assess students authentically, based on their performance on complex, high quality tasks, it is likely that the delivery system will have to change to ensure that students are routinely given complex, high quality tasks to do in all areas. Wiggins goes so far as to call authentic assessment "the Trojan horse of school reform" (Wiggins, 1991).

Conclusion

The number and range of well-documented examples of how both traditional classroom assessment practices and standardized testing can negatively impact students is thought-provoking. It is becoming increasingly apparent that traditional classroom evaluation practices and standardized testing are often detrimental to student learning and have been contributing to poor American educational results. Instead of encouraging the kinds of tasks that contribute to

producing students who can infer, analyze, synthesize, and evaluate, they encourage students to memorize facts. Instead of creating an open climate that encourages students to collaborate with each other, they isolate and pit one student against another. Instead of presenting an authentic view of the complexity of the learning process, they give a simplistic and biased view.

Authentic assessment, on the other hand, appears to hold promise for enhancing learning and for giving stakeholders an authentic view of the complexity of the learning process. Educators attempting to implement alternative forms of assessment recognize the importance of prior knowledge and respect the differences among learners. They use assessment techniques that take into consideration recent research on the development of cognition and the acquisition of literacy, techniques that respect both teacher and student judgment. They view students as the constructors of their own knowledge and the logical evaluators of their own work. Authentic assessment gives legitimacy to the view that the purpose of evaluation should be to improve learning and should therefore be in the hands of the principal shareholders.

This thesis explores the feasibility of using alternative assessment techniques in a kindergarten classroom. The following research questions are explored:

1. Can kindergarten students' "writing" growth be reliably and validly documented using a portfolio system that allows children to choose their own best journal writing?

2. Does entry length in experience-story production provide a useful indicator of growth in literacy?

3. Does working in groups provide a valid and reliable format for assessment of individual growth in mathematical problem solving?

4. Can assessment of individual children's nursery rhyme reconstruction provide a teacher with useful data on the link between oral memory skills and reading and on how kindergarten children go about beginning to decode words?

5. Does the quality and usefulness of the information gained in alternative assessment situations justify the time it takes to gather and analyze the data?

Chapter 3
METHODOLOGY
Introduction

Traditional assessment has placed a heavy emphasis on those aspects of learning that could be easily quantified, often ignoring higher level, complex skills, attitudes, and outcomes that are more difficult to measure but that are of crucial importance in the eyes of teachers (Sternberg, 1990). There appears to be a need for assessment of learning that involves complex, high-level tasks and that is used to inform classroom decisions (Wiggins, 1991). An increasing number of concerns have arisen recently about the harmful effects of evaluation practices on students (Crooks, 1988). Evaluation practices have dampened motivation, promoted surface-level memorization over deep understanding, discouraged students at all levels from doing quality work, and misled stakeholders about the complexity of the learning process (Crooks, 1988; Wiggins, 1991). Educators are beginning to move away from norm-referenced measures that discourage cooperation among students, give a false picture of the complex work that students will be expected to do as adults, and fail to inform classroom practice (Wiggins, 1991; Valencia, 1991). This study was undertaken in an attempt to explore the feasibility of using alternative assessment techniques in a

kindergarten classroom within the context of a half-day kindergarten day.

The following research questions were explored:

1. Can kindergarten students' "writing" growth be reliably and validly documented using a portfolio system that allows children to choose their own best journal writing?

2. Does entry length in experience-story production provide a useful indicator of growth in literacy?

3. Does working in groups provide a valid and reliable format for assessment of individual growth in mathematical problem solving?

4. Can assessment of individual children's nursery rhyme reconstruction provide a teacher with useful data on the link between oral memory skills and reading and on how children go about beginning to decode words?

5. Does the quality and usefulness of the information gained in alternative assessment situations justify the time it takes to gather and analyze the data?

Method

Environment

This study was conducted in a small suburban elementary school district in the greater San Francisco Bay area in a classroom that contained 23 kindergarten children. The school was a kindergarten through fifth-grade school with three classes at each grade level. Most students in the school were from middle to high

socio-economic backgrounds. The environment of the study involved a mostly self-contained, half-day kindergarten class that met from 8:40 a.m. to 12:05 p.m. five days a week. Music and physical education were taught on alternate days by specialists (i.e., music was taught on days with even-numbered dates and PE was taught on days with odd-numbered dates). At the beginning of the study, there were 11 females and 10 males in the class. One female joined the class the week of January 28; one male left the class during the Easter break (first week of April); and one female joined the class right after Easter vacation. At the beginning of the year, children in the class ranged in age from five years one month to six years three months. At the end of the year, they ranged in age from five years ten months to seven years. (See Appendices A and B for a complete breakdown of ages.) Of the 23 children who participated in any phase of the study, nineteen were Caucasian, one was half Japanese-American and half Caucasian, one was Hispanic, one was part Native American, and one was of Middle Eastern background. All of the children spoke fluent English except for the child of Middle Eastern background, who was very close to fluent, and the Hispanic child, who spoke very little English at the beginning of the year.

Procedures

The intent of this research was to explore the feasibility and usefulness of employing alternative assessment techniques in a

kindergarten classroom. The techniques used had the following characteristics:

- 1) They were conducted over extended periods of time.
- 2) They involved active student participation.
- 3) They considered process as well as product.
- 4) They were embedded in the curriculum of the class.
- 5) They involved student reflection.
- 6) They employed a variety of techniques.
- 7) They emphasized what children could do rather than what they could not do.
- 8) They viewed the role of the teacher as that of facilitator and co-learner rather than that of evaluator or judge.
- 9) They were used to inform classroom decisions.

Five alternative assessment procedures were undertaken. The first two, Journals and Very Important Person Stories, were initiated at the beginning of the school year. The other three, People Pieces Project, Squares Project, and Nursery Rhyme Project, occurred between the end of Easter vacation (April 5) and the end of the school year. The three end-of-the year tasks were conducted one right after the other, without overlapping, in the order stated above. Of the two tasks initiated at the beginning of the year, Journal Writing was discontinued in early March, and Very Important Person Stories continued until the first week in June, by which time each

child in the class had been the Very Important Person. Specific procedures for each project are outlined below.

Journals

During September and October, children "wrote" in their journals sporadically, partly because other activities seemed more appropriate for kindergarten children at this time of year. During November and December, journals were more or less dropped due to the encroachment of holiday projects. Starting right after Christmas vacation, children and teacher began writing in their journals for ten minutes each day. The teacher and sometimes students suggested topics to write about, but everyone was free to write on a topic of his or her own choosing. The ground rules were that only pencils were to be used in journals and that the ten minute period must be silent. Invented spelling was encouraged, and children who were not yet at the invented-spelling stage were encouraged to draw. The teacher and aide had journals as well. Journals were made from folder covers with randomly arranged sheets of colored copy machine paper inside. The covers, designed to hold binder paper, had three brads that could be folded over, so that paper could be added or removed. At first the journal pages were not dated, but when the teacher noticed that some children were skipping around in their journals, working back to front, or using numerous pages at one sitting, the teacher began dating pages on the copy machine, paper

clipping the journals open to the right page, and requiring that only the front and back of the page dated for the day be used.

The teacher's intent was to meet individually with each child each week, at which time the child would choose his or her "best" page to go into a "keeper" journal. Within a matter of weeks, it became apparent that it was going to be difficult to get around to all children to have them choose their "keeper" journal pages, so the teacher replaced the system of using randomly-arranged colored journal pages with one that involved color-coding the pages by week --all white pages one week, all blue pages the next, and so on. The rationale for color coding was that when the class and teacher got behind on choosing the selections to be kept for a given week, it would expedite matters to be able to lay out quickly the five entries on blue paper, for example, and say, "Choose the best of these."

Very Important Person Experience Stories

Very Important Person experience stories were chosen as one area of assessment because they integrated reading, writing, speaking, listening, and self-esteem; because there was a built-in plan (at least in theory) of insuring consistency and measurement over time; and because the teacher already had several years' experience doing Very Important Person books and had worked some of the kinks out of the system. Also, the Very Important Person books were congruent with whole language principles and had been found to be popular with students and parents alike; not

only did students keep their books from year to year and read them periodically, but parents mentioned frequently how much they enjoyed them; thus, the books constituted good public relations between home and school. For all the above reasons, they had become a regular part of the kindergarten curriculum.

At the beginning of the year, the dates of all five-day weeks were put into a box and each child drew out a week to be Very Important Person. Dates and VIPs were the following:

Week of September 17--Teacher (took the first turn to model the routine)

Week of September 24--Student F-1

Week of October 1--Student F-2

Week of October 15--Student F-3

Week of October 22--Student F-4

Week of November 5--Student M-5

Week of November 26--Student F-6

Week of December 3--Student F-7

Week of December 10--Student F-8

Week of December 17--Student M-9

Week of January 7--Student M-10

Week of January 28--Student M-11

Week of February 4--Student M-12

Week of February 25--Student M-13

Week of March 11--Student F-14

Week of March 18--Student F-15

Week of March 25--Student F-16

Week of April 15--Student M-17

Week of April 22--Student F-18

Week of April 29--Student M-19

Week of May 6--Student F-20

Week of May 13--Student M-21

Week of May 20--Student M-22

Week of June 3--Student F-23, who entered the class
after VIP weeks were drawn.

The Monday of each VIP week, the VIP for that week was interviewed in front of the class and asked the following questions:

What is your favorite color?

What is your favorite food?

What is your favorite sport?

What is your favorite book or story?

What is your favorite T.V. show?

Do you have any brothers or sisters? What are their names
and how old are they?

Do you have any pets? What are their names and what do they
look like?

Who are your best friends?

What do you like?

What do you not like?

What are your plans for the future? (If no response was given
on this item, two prompts were given: "Do you know yet what you

want to be when you grow up?" and "Is there anywhere you would like to travel?" If still no answer was given, the teacher wrote No plans yet.)

All answers were recorded on the chalkboard, in the presence of the class. The aide copied them onto a sheet of paper, and as soon as possible, usually after school that day, the teacher wrote them on sentence strips and put them up on the Very Important Person bulletin board at the front of the classroom.

On Tuesday, the teacher asked what the class remembered about the Very Important Person. When all items had been reviewed, the children went to their seats, where 12" by 18" white drawing paper and their crayons had already been set out. The children were reminded that their pictures had to include the VIP doing something.

At the earliest viable opportunity, usually during the long snack recess on Tuesday, the teacher began calling children in individually to take down their explanations of their pictures. A routine was established whereby each child sent the next. At the beginning of the year, recess time was utilized so that both teacher and child could concentrate without distraction. The child was seated on the teacher's left, so he or she could watch as the teacher wrote. The teacher pronounced each word aloud as it was written. Later in the year, when the routine had been firmly established and children had been trained to be more independent during Free Choice time, it

was sometimes possible to take down stories during Free Choice as well.

During the week, the teacher made a cover for the book and copied the child's "words" (preferences) onto a piece of 12" by 18" drawing paper that became the first page of the book. When all drawings had been completed and all experience stories taken, the pages were bound together using Oriental bookbinding. The book was read aloud to the class, after which it was given to the VIP, along with the child's "words" from the bulletin board, to take home and keep.

Before the book was taken home, the teacher copied the experience stories into her computer (Macintosh SE/30) so that a running record of each child's weekly production was created. The computer's word count feature was used to count the number of words each child "wrote" each week; the teacher graphed the results, both on individual graphs and on a whole-class graph, and photographed the graphs with a Ricoh Shotmaster camera.

People Pieces Project

Each child was called into the room individually to work with the teacher with a set of Creative Publications People Pieces. The children had not used the People Pieces previously in this classroom. The pieces were spread out face up so that all could be seen. The teacher told the child, "These are People Pieces. There are lots of ways they can be sorted. For example, you could put all the red ones

here and all the blue ones here." (The teacher moved them into two separate piles, sorting by color in front of the child to demonstrate.) "But there are lots of other ways do it. Can you sort these People Pieces a different way?"

After each time the child sorted, the tiles were flipped over (each had a small sticker with an alphabet letter on the back so the teacher could quickly record how the child had sorted and later, if necessary, could reconstruct what the child had done.) Then the tiles were flipped back and the child was asked, "Can you sort another way?" An attempt was made to do all individual sorting at the same time of day and in the same location, but in order to get the sorting done within a reasonable length of time, keeping time and location constant proved impossible. The teacher soon began using all times available to work with children alone. These included 1) before school, when the first child arrived 2) during snack recess, when the other children were on the yard 3) during Free Choice, at a table outside while the rest of the class was inside with the aide, and, toward the end of this part of the project, when the teacher could see that time was getting tight, in a couple of cases, 4) behind the pocket chart at the drainboard while the aide took the rest of the class for sharing. In these last cases, children were asked to do their sorting silently, so that they would not disturb the rest of the class and so the rest of the class would not hear how they were sorting.

After all children had sorted individually, children were placed in randomly-selected groups of four to five children. The first day of

this part of the project, they worked in their groups and each child made a "recording sheet" by coloring his or her duplicated People Pieces sheet to match the one the teacher had made. Then the children cut apart their "recording sheets" and placed their paper People Pieces tiles into individual envelopes labeled with each child's name and group number.

The following day, each group was given a large sheet of white butcher paper, the envelopes containing their paper People Pieces, glue sticks, and a set of real People Pieces to share. (Blank "recording sheets" and red and blue marking pens were also available in case any group wanted to make additional paper People Piece sets; since there were a maximum of five children in a group and six ways to sort People Pieces, and since at least one child had sorted more than five ways when they had sorted individually, it was conceivable that at least one group might.) Children were told to sort as a group as many ways as they could, recording their answers by gluing their paper People Pieces to the butcher paper.

Over the weekend, at home and consequently out of sight of the children, the teacher made her own butcher paper recording sheet showing the six ways the People Pieces could be sorted (by color, by height [or adult/child], by girth, by gender, by shoe color, and by hair color.)

Each group was videotaped making a presentation to the class, showing the project and telling how many different ways their group had found to sort, after which the class discussed how some of the

problems encountered could have been avoided (such as having half the work upside down when the group project was pinned to the wall). As a means of scaffolding, after all groups had been videotaped making their presentations, the teacher was also videotaped making her presentation to the class.

Between the first time and the second time the People Pieces project was done, the teacher decided to do additional scaffolding for the children by giving them additional sorting practice that would be comparable to the kind of sorting they would have to do with People Pieces. Using a set of stylized chickens that had been run off for a patterning project, the teacher used Japanese brush pens to create five sets of "Chicken Pieces" that were all different from each other, each of which could be sorted by color of comb, head, eye, beak, cheek, wing, tail, body, and feet. (See Appendix C for an example of a set of Chicken Pieces.) A colored dot was placed on the bottom of each card, a different color dot for each set, so that if the chickens got mixed up, the teacher could readily sort them into workable sets again. Children seated in a large circle on the rug sorted by some attribute, reported to the class how they had sorted, and passed the cards until all children had had a chance to sort.

On the following day, the children worked in their groups (the same groups as for the People Pieces) to sort Toads and Rabbits. The Toads and Rabbits project followed the same format as the People Pieces project; the children had never seen the Toad and Rabbit tiles before. Because the Toads and Rabbits were more difficult to color

than the People Pieces (due to the fact that they were wearing stripes or polka-dots), the teacher had colored all the recording sheets. Otherwise, the project proceeded as the People Pieces project had. Again, the children were videotaped presenting their products to the class. Products were, again, taken home by the teacher and photographed.

Squares Project

The Squares Project involved having the children try to figure out, first individually and then in groups, how many different ways a square that had been divided into four equal smaller squares could be colored using only two colors.

First the teacher explained the task. Four large squares had been drawn on the chalkboard, each divided into four smaller squares. The teacher told the class, "Our job is to figure out how many different ways we can color our squares using only two colors. One way we might color would be to color the top two squares purple and the bottom two green." The teacher did so, demonstrating in front of the class.

Going on to the second large square, the teacher said, "Or, we could color the first little square green and the other three purple." Again the teacher demonstrated by coloring in the little squares in front of the class. Then the teacher asked, "What else could we do?" If no-one had suggested coloring all four little squares the same color, the teacher was prepared to demonstrate that as a possibility;

fortunately, the next child did. The teacher had the child making the suggestion come up and color in the example as the child had suggested, then called on another child for still another example. Then the teacher drew a fifth divided square on the board, colored it the same way as one that had already been demonstrated, and asked, "Would it be a good idea to color the next square like this?" The children decided that, since they were trying to get as many different squares as possible, it would not be a good idea to have "repeats." The teacher concurred, then said, "What you may not do is divide the little squares like this." (The teacher drew a line diagonally from one corner to another on one of the small squares and colored half the small square green and the other half purple.) "Each little square must be one solid color." The children were told that they did not need to rush, that they would have all the time they needed to figure out all the possibilities. If they did not have time to finish that day, they could continue the following day. They were also told that they had more than enough squares on their papers to make mistakes and still get all the possibilities, and that in any event there were no more papers, so that if they made a mistake, they were to cross out the square with the mistake and just go on to the next one.

After the children appeared to understand the assignment, they were sent to "private offices" that had been built out of large kindergarten blocks placed on tables so that they could not see each other's work. Each "office" contained a sheet and a half of outlined

two-inch by two-inch squares. (See Appendix D for an example.) Each of the large squares had been divided into four smaller squares one-inch by one-inch. (12 divided squares fit on an 8 1/2 by 11 inch piece of copy paper; an additional six fit on the half sheet stapled to it.) Each "office" also contained the child's box of crayons, with the purple and green crayons removed from the box and the rest of the crayons, in their closed box, pushed to the back of the office area. The children's names were already on their papers. They knew where to sit because their desk name tags had been randomly placed on the tops of the blocks. Before the children began work, the teacher checked to see that everyone was sitting in the right spot.

Most children finished this part of the project within one day (i.e., within what was left of the forty-minute work period after explanations). However, some opted to continue the following day. The teacher collected the papers of those who said they were finished and paperclipped them together. Students who needed more time were instructed to turn their papers face down and leave them in their "offices." During the supervised recess that occurred immediately thereafter, the teacher collected the papers from the "offices," noted in pencil at the top of each that more time was needed, and paperclipped them together. The two sets of papers, finished and unfinished, were placed in separate stacks on the tops of two tall file cabinets, out of reach of the children.

The following day, all children returned to the same "offices." Those who had finished the previous day were given a simple

coloring job to do until the others had finished their Squares projects.

After the students had taken as much time as they needed to figure out as many possibilities as they could, the teacher took their work home and made copies of what they had done (that is, the teacher colored a duplicate class set of squares, making the work as qualitatively like each child's as possible). The teacher also tabulated the number of different ways each child had found, noted the number of duplicates, and took notes on any problems with direction following.

The next day, students were randomly assigned to groups of four or five children each. The children in each group were to combine their ideas and see how many different two-color squares they could come up with as a group. Each group glued its squares onto an 18" by 24" sheet of yellow construction paper.

Over the weekend previous to the start of this project, the teacher also created a product, gluing the sixteen different ways the squares could be colored onto an 18" by 24" sheet of yellow construction paper. The teacher's product was not shown to the children until theirs had been completed and presented to the class.

When all products had been completed, the children were videotaped presenting their products to the class. At the end of the students' presentations, the teacher was videotaped presenting her product to the class. When student products were presented, the teacher asked the presenting group a set of structured questions to

probe their thinking and to elicit information. Questions asked were the following:

- 1) What were you supposed to do?
- 2) How did you do it?
- 3) Did your group have any problems?
- 4) Did you work as a group, or did each of you work by yourself?
- 5) How many different ways did you get?
- 6) Did you learn anything from doing this?

A parent volunteer did the videotaping. When the teacher made her presentation, the same questions were asked of her by a student who could read the prepared questions from a card. At the conclusion of the project, the teacher took the student group products home and photographed them.

The project was done a second time, using two different colors for coloring in the squares (purple and green the first time, blue and orange the second). A different-colored background was also used (white instead of yellow). Children remained with the same groups; the same videotaping procedure was followed. The teacher did not do a second Squares project. Children's projects were, again, taken home and photographed.

Nursery Rhyme Project

The teacher worked with children individually to discover a nursery rhyme that each child could say from memory. When she

had a rhyme for each child, each rhyme was printed in D'Nealian in wideline black marking pen, cut into individual words, and laminated. Then each rhyme was placed in a large Ziploc baggy, which was labeled with the name of the rhyme; the baggy was placed in a larger transparent (McCracken) bag with a handle, which was hung on the back of the pocket chart.

Each child was called into the room individually to work with the teacher. When the child came into the room, the teacher already had all the words of the child's rhyme laid out on the drainboard, oriented correctly (i.e., with the words both face up and topside up), but mixed up. The teacher said, "These are the words to 'Jack and Jill' (or whatever the rhyme was that the teacher knew the child could say), but they're all mixed up. Can you put them in the pocket chart the way they're supposed to go?" The teacher had a checklist she had made out that she thought anticipated most of the things the children were likely to do, but they moved their words around so much as they worked that it was easier just to flip the checklist over on the back and take notes. When the child had done the best with the rhyme that he or she could, his or her picture was taken with the words in the pocket chart. Each child was sent for the next. During the time one child was going out to the playground to get the next, the teacher put away (or mixed up if the rhyme was the same) the word cards from the first child's rhyme and got out and arranged the words for the second child's rhyme.

Analysis of Data

Journals were discontinued before the conclusion of the project due to their taking more time than was available, both for the writing and for the choosing of "keeper" journal pages. The teacher, bothered by giving up the project, analyzed those journal pages that were available in order to ascertain what percentage of each child's pages contained pictures; pretend writing; pictures and symbols; pictures, symbols, and words; pictures and words; and words only.

Very Important Person data consisted of numbers of words dictated per child per week. Each child's words were counted by the Macintosh SE/30, graphed both for individual children and for the class as a whole, and compared over the course of a school year.

Data for both People Pieces projects and Squares projects consisted of numbers of possibilities found by groups and the comparison of post products to pre products. For the Squares Project, individual data are also reported. Videotaping of the presentations of group products also allowed for the documenting of student progress in the areas of presenter competence, audience behavior, understanding of what each assignment involved, and ability to work as a member of a group.

Data for the Nursery Rhyme Project consisted of photographs showing children with their reconstructed nursery rhymes. Data for this project are incomplete, due to the fact that the end of the school year came before the project was completed.

A limitation of this research design is that the results found may not be generalizable to other parts of the country or state, other types of schools, other school districts, other socio-economic levels, or even other classes. They also may not be generalizable to full-day kindergartens or to kindergartens with different daily schedules or different levels of adult support.

Chapter 4

RESULTS

The purpose of this project was to explore the feasibility of using alternative assessment techniques with kindergarten children to inform teaching. Five techniques covering a range of curricular areas were chosen:

1) Portfolio Assessment of Journal Writing

Individual students and the teacher worked together to choose pages from each child's journal that would go into a 'keeper' journal to show progress in writing over time.

2) Very Important Person (VIP) Stories

Patterns of verbal output in story writing were tracked over a year's time to identify patterns and assess progress.

3) People Pieces Project

This project was designed to assess kindergarten children's sorting ability, as well as their ability to collaborate with other members of a small group, to display their combined efforts sensibly, and to present their work to the class.

4) Squares Project

The Squares Project was a thinking-skills and mathematics project designed to assess kindergarten children's ability to work out the number of possibilities involved in coloring subdivided squares, combine their work with that of other small-group members, and present their group work to the class.

5) Nursery Rhyme Project

The Nursery Rhyme Project was designed to assess how kindergartners go about decoding and recoding material that is accessible to them on an oral level.

The results from each technique will be treated separately in the narrative that follows.

Portfolio Assessment of Journal Writing

Journal writing was attempted from the beginning of the year, but other kindergarten needs encroached to the extent that it was done only sporadically until January. All but two of the children were able to adhere to the requirement to use only the front and back of the journal page designated for the day, with the result that the teacher could tell when work had been done.

Having the opportunity to listen to children carefully and individually during the choosing of "keeper" journal pages gave the teacher insight into children's feelings and allowed the teacher to know the children better. In one case a boy drew a cemetery and when he chose the page for his "keeper" journal commented, "It reminds me of my mom's mom. She's dead."

Having their opinions sought and respected when they chose their "keeper" journal pages had several results: it put the burden of evaluation on the learner; it provided a baseline for student evaluation competencies; and it impressed upon the teacher the need to establish with the children standards for evaluating student

writing. The experience of evaluating their own work also appeared to have possibilities for building self-esteem: Many children smiled, displayed enthusiasm, or made positive comments as they chose their 'keeper' pages. Some children, as noted above, also talked about their pictures as they made their choices, giving the teacher added information, both about themselves and about their interests and concerns. Many children appeared to have difficulty giving what the teacher considered good reasons for choosing the keeper pages they chose. Often, when asked why a page had been chosen for the "keeper" journal, the child responded, "I don't know" or "I just like it."

Attempting the project forced the teacher to confront her values both as to the kinds of activities she deemed appropriate for kindergartners and as to the use of kindergarten time. Journal writing was discontinued in the Spring when the teacher decided that it was prohibitively time-consuming to meet individually with all kindergarten children twice each week, once to take their VIP experience stories and once to choose their "keeper" journal pages. Kindergarten children took anywhere from several seconds to twenty minutes to dictate their stories for VIP books; choosing the pages for their "keeper" journals took from five minutes to fifteen, and time also had to be allowed for each child to find the next child on the yard and send him or her in. Doing both these projects within the time available was not possible. Since one of the two projects had to

be dropped, journal writing was chosen.

Very Important Person Project

Results of the VIP experience story project were the following:

The VIP project allowed the teacher to work within a whole language framework, in that reading and writing were viewed as interactive processes and concentrated on the supremacy of meaning. Each child's point of entry was necessarily at his or her own level of language acquisition and children proceeded at their own paces. In addition, the VIP books were involving for both student and teacher. Each child's "words" from the bulletin board were personally meaningful, and there was the possibility of enhancing self-esteem, as each child was the main character of his or her own book. In addition, the teacher was able to get to know her students better, as each week she concentrated on the preferences of one of them.

Consistency in regularly scheduled assessment was built into the project. Putting the children's writing into the computer allowed for the comparison of each child's entries to previous and subsequent entries; the teacher could see errors common to most of the class; the computer's word count feature allowed for the (relatively) easy documentation of growth in quantity of student writing over the course of a year, and graphing student output allowed the teacher to see patterns of student production as they related to school holidays.

These children, most of whom were probably at the preoperational level, did not "write" more about themselves than they did about other members of their class. Boys "wrote" more than girls did. Range in production was very wide, from a low of two words to a high of 415 words for individual entries over the course of the year.

During the first, second, and third eight-week periods of the year respectively, children dictated an average of approximately 19, 25, and 33 words per entry (See Figure 1). Children dictated an average of 210 words during the first half of the school year and an average of 353 words during the second half.

Fourteen children out of twenty or 70% of the children who were enrolled all year wrote more the second half of the year than they did the first half. The other six children, or 30% of those enrolled all year, wrote less the second half of the year than they did the first (although two of them dictated almost as much the second semester as they did the first [279 words to 263 and 256 to 242 respectively]).

Girls enrolled all year dictated an average of 21 words per entry, whereas boys enrolled all year dictated an average of 37 words per entry. If the one exceptionally verbose boy is omitted from the count, boys enrolled all year dictated an average of 31 words per entry. The average number of words dictated by boys dropped below the average number of words dictated by girls during only two weeks of the year, weeks 16 and 19. Data were also

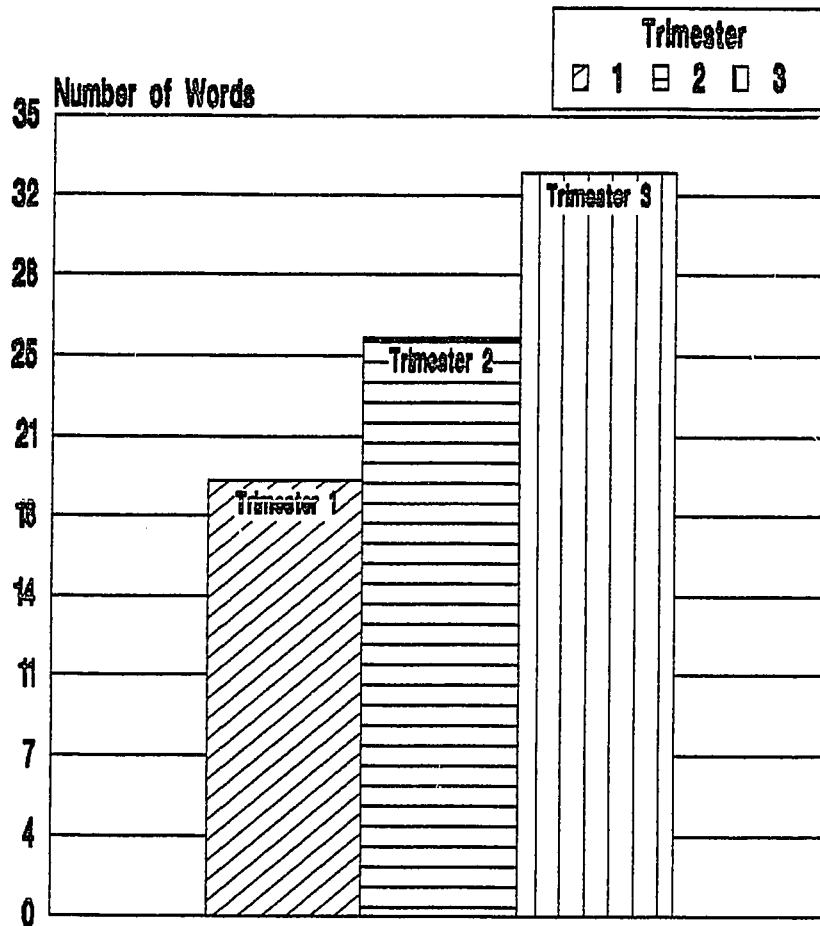


Figure 1. Very important person word count by trimester.

Average number of words per entry.

Atypical verbose male excluded

extrapolated to determine whether boys dictated more about boys, and girls dictated more about girls, or vice versa. Regardless of whether the VIP was a boy or a girl, boys dictated more than girls in 22 out of the 24 weeks of the project. In week 16, when a girl was VIP, and in week 19, when a boy was, girls dictated more than boys (See Figure 2).

Most children appeared to have an up-and-down pattern to their writing, up one week (or period of weeks), down the next, up the next, and so on. Some children wrote the same thing or the same thing with minor changes for weeks in a row (something the teacher had not noticed in previous years when VIP books had been done without putting them into the computer). Lines of production flattened out over holidays that lasted two weeks. That is, whereas production generally went up as the school year progressed, when writing was discontinued for two weeks, production remained constant for a number of children (i.e., they were writing about as much when they came back after vacation as they had been before they left [See Figure 3.]). No similar result was shown for one-week holidays. The three children who wrote a great deal toward the end of the school year were all children who exhibited emotional and behavioral problems at school.

The VIP project was both time- and labor- intensive. Taking story dictation for every child for every week that the books were done (excluding entries for two late entrants, one early leaver, absences, and a few inadvertently lost entries) amounted to the

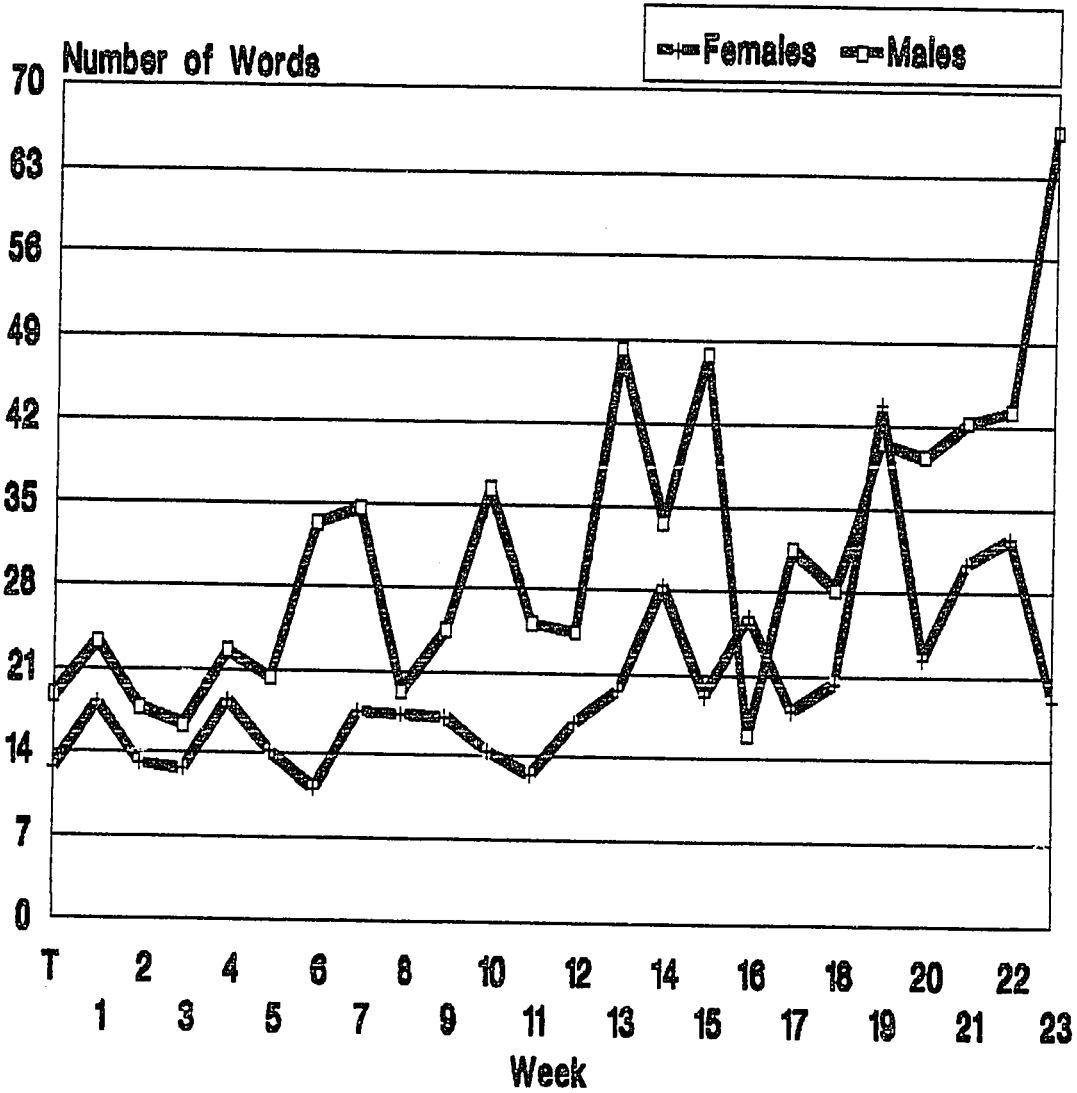
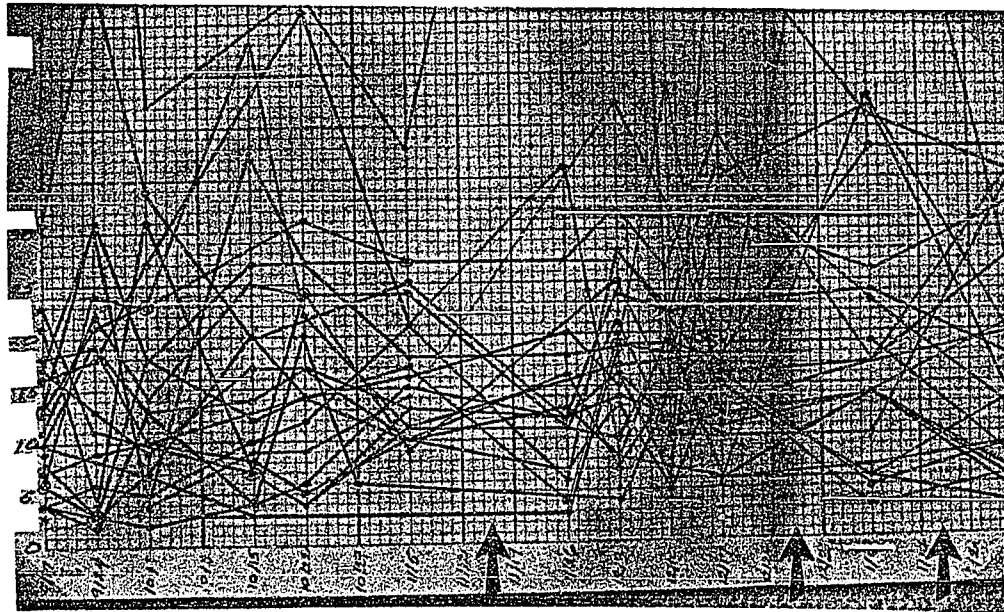
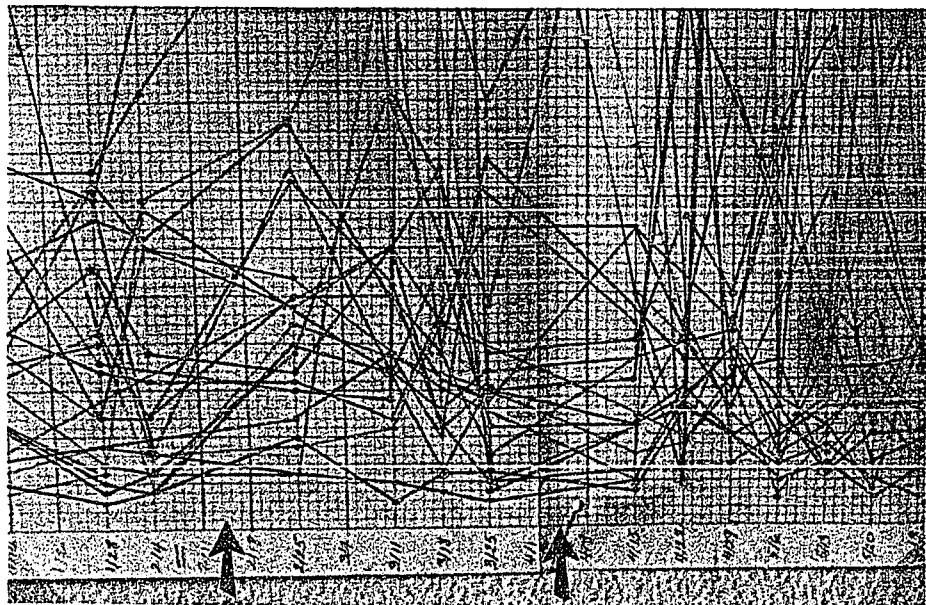


Figure 2. Very important person average word production by gender (atypical verbose male excluded).



September 17 - January 21



January 28 - June 3

Figure 3. Year's Very Important Person word production graphs showing the flattening out of lines over two-week vacation periods (arrows indicate two-week breaks).

dictating of 476 stories. Finding time to get stories from everybody, including children absent for only part of a week, necessitated a high degree of consistency and organization on the part of the teacher and help from an aide to watch the rest of the class while stories were being dictated. When the teacher was absent, problems also resulted. For example, in the case of this project, the teacher was ill in October and the substitute completed F-4's book and sent it home before it had been put into the computer, necessitating its retrieval when the teacher returned.

The VIP project forced the teacher to make difficult choices about the use of teacher time. When the teacher was inside taking dictation during recess, she was unable either to watch children interacting on the yard or to do set-up and clean-up jobs that it might otherwise have been possible to do. Likewise, care had to be taken to see that all children had time to eat their snacks and to spend time on the yard during recess. Another drawback to the project from the teacher's point of view was that she felt locked into the VIP books. The children were given breaks from VIP stories (none were taken on weeks that contained school holidays), but the teacher was not, as in addition to taking the stories from the children originally, the teacher also had to find time to type them into the computer, and "vacation" days provided needed time to catch up.

Putting all the children's dictation into the computer took an inordinate amount of teacher time. In fact, the teacher got behind the very first week and was forced to change the going-home date of

the VIP book from Friday to Monday in order to have the weekend for putting stories into the computer. Even then the length of time required to do the job was prohibitive, and for awhile the teacher had a double filing system going in order to get books out on time without losing data: Bob's VIP was a running record of all the experience stories Bob had written; Bob As VIP was a record of what everybody in the class had written about Bob. This system involved extra work, but it did allow the teacher to get the VIP books out on time and to keep track of most entries.

People Pieces Project

Results of the People Pieces Project were the following:

The project was time- and labor-intensive and expensive.

Since the teacher owned only one set of People Pieces, she attempted to find four other sets to borrow. Two colleagues had them, but they were older versions that had different attributes from the teacher's set. None were available either from surrounding districts or from the County Office of Education. Ultimately, the teacher made sets (reproduced facsimiles mounted on tagboard) so each group would have a set to use while making their products. The teacher also made five prototype recording sheets for the children to copy in making their own. The teacher bought Toad and Rabbit Pieces and then made a recording sheet for each child. (It was thought that it would be too difficult and therefore frustrating for the children to

try to color their own recording sheets, since the striped and polka-dotted detail on the toads' and rabbits' clothing was quite small.)

The project was also expensive. It involved not only the cost of the Toad and Rabbit Pieces, but also Japanese brush pens to make the Chicken Pieces, videotape for the filming part of the project, a good camera, and film for recording the children's products.

Some children had trouble making their People Pieces recording sheets. In addition, most children had trouble working as collaborating members of a group. They ruled off their own territory and labeled their parts of the group product with their individual names rather than thinking of the product as a group product. They did not collaborate; either they copied another group member's work or did not pay attention to what other members of the group were doing. Either way, the result was that they duplicated each other's work rather than getting a number of different possibilities as a group. They also worked on opposite sides of the product rather than viewing it from the same orientation. Then they argued about what side of the product was "up."

The videotaping part of the project was problematic for a number of reasons. The teacher, who was inexperienced at videotaping, had to rely on parent volunteers to do the taping while she interviewed the groups. That meant that videotaping could be done only on days when parent volunteers were present, and because the taping had to be done within the constraints of the classroom schedule (i.e., P.E./music taught by specialists), timing was

difficult. In addition, there were a number of interruptions, including noise from children going by outside, the kindergarten toilet flushing, bells that applied to other classes but not to the kindergarten, an aide coming into the classroom to check heads for lice, calls over the intercom system, and the interruptions of the kindergartners themselves. The teacher had trouble keeping her questions consistent from group to group, and the children being interviewed, who were having their first experience at making presentations, did an inordinate amount of moving around, especially during the early phases of the videotaped projects. Some talked to the product rather than to the audience. They turned their backs on the camera and consequently were hard to hear. The school video camera was not compatible with the school video machines, so the film had to be "cooked" (i.e., converted) before it could be shown.

Nevertheless, videotaping allowed observers to see how children behaved and provided a window into how they thought. In addition, progress over time could be documented. On the early tapes, it is obvious that there is no substantive conversation occurring between audience and presenters. In the later tapes, however, growth both as presenters and as audience members is evident. In one of the later tapes, one little boy starts to say, "When you thought..." and then interrupts himself to ask instead, "*Did you think...?*"

The videotape could be revisited many times. The teacher gained a clearer understanding of how the children were thinking.

In one case, due to the inexperience of kindergarten children in producing products, it was not clear from the product how one little girl had sorted; but it was very clear when she explained her work on the videotape: "These are the skinny ones and these are the fat ones..." In addition, the teacher made two errors in presenting her work to the class and was not aware she had done so until she viewed the videotape.

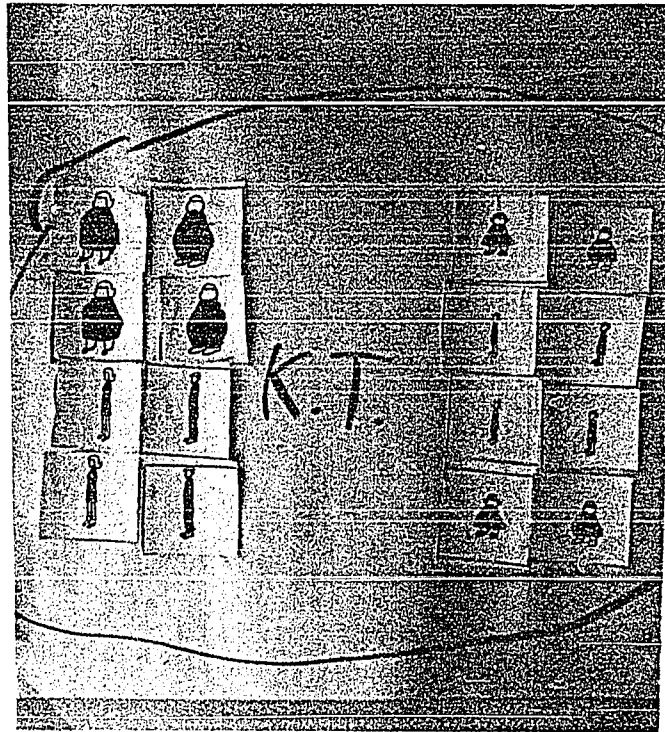
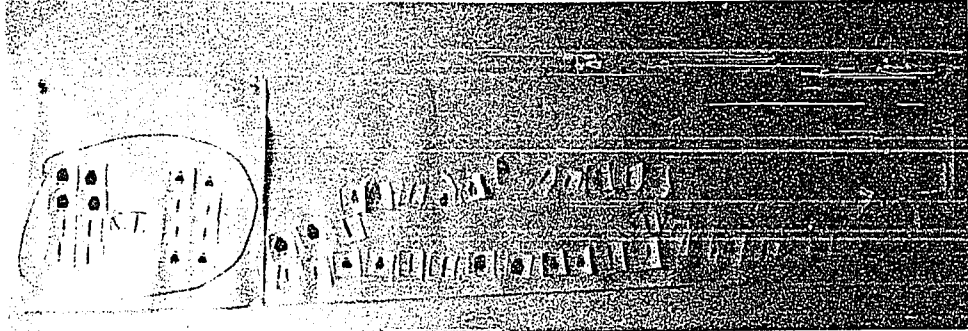
Over the course of the project, children showed progress both in working as group members and at finding more of the possibilities. During the People Pieces phase of the project, one little girl sorted all six different ways during the individual sort; her group product, however, showed only one way to sort, by color. The individual was brought down to the level of the group.

During the People Pieces sort (pretest), Groups 1, 3, 4, and 5 found only one way each to sort; Group 2 found only two ways. There was a lot of misunderstanding about what the assignment was and much encroaching of patterning (which did, however, allow the teacher to see what some of the children knew about patterning). (See Figure 4 for an example of a People Pieces [pretest] group sorting product and detail from that product showing how many kindergarteners rule off their own territory. See Figures 5.1 and 5.2 for the teacher's scaffolding project. See Figures 6.1, 6.2, and 6.3 for the above group's Toad and Rabbit [posttest] product showing how one group's work changed over the course of the project.)

During the Toad and Rabbit sort (posttest), Group 5 found two

Group IV's People Pieces (Pre-Test) Product

Only one person sorted accurately- and she ruled off her own work



Detail From Group Product Above
(Shows how accurate sorter sorted - by Adult/Child)

Figure 4. Detail from People Pieces group product.

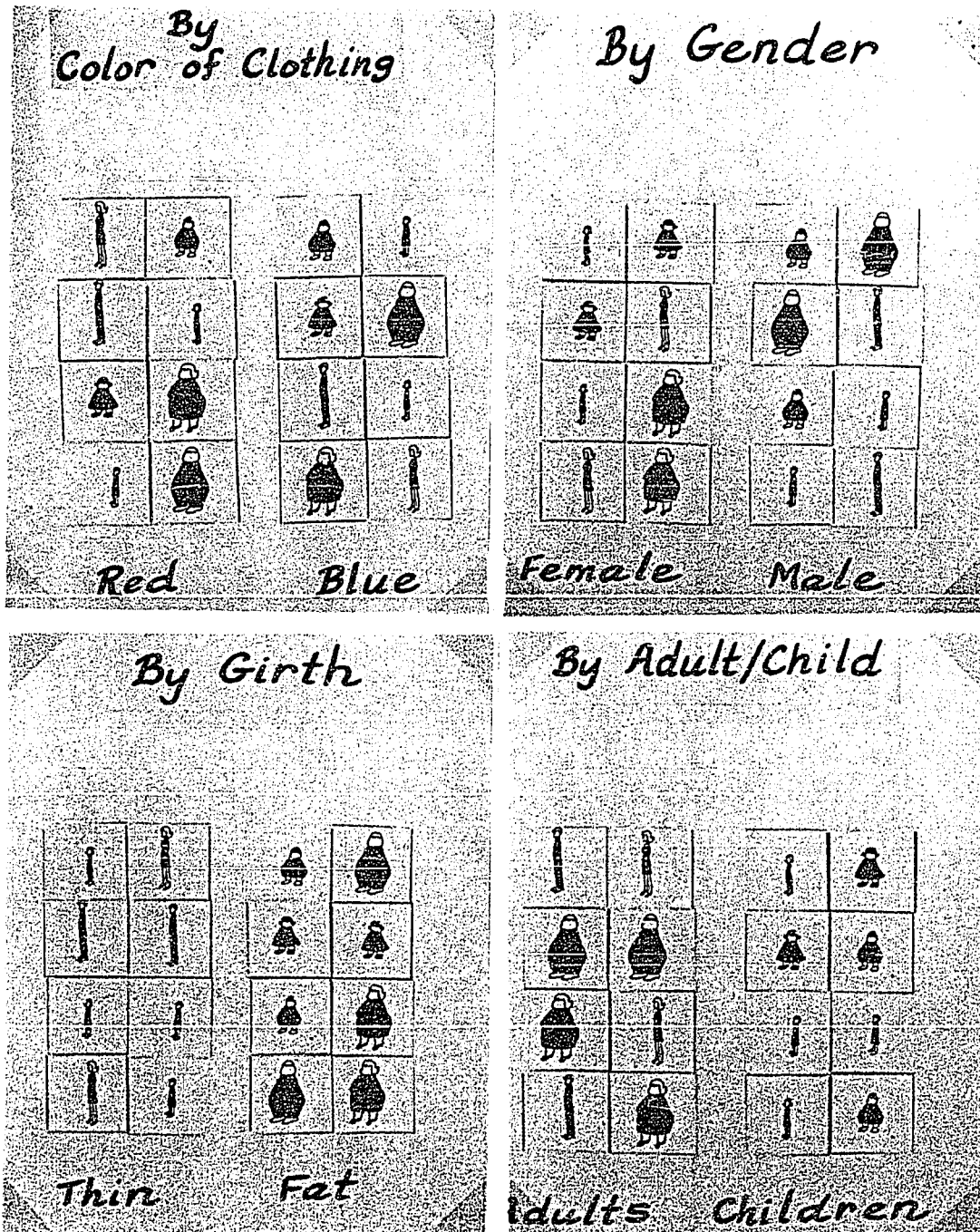


Figure 5.1. Detail from teacher's People Pieces product used to scaffold between People Pieces products and Toad and Rabbit products (shows four of the six ways to sort People Pieces.)

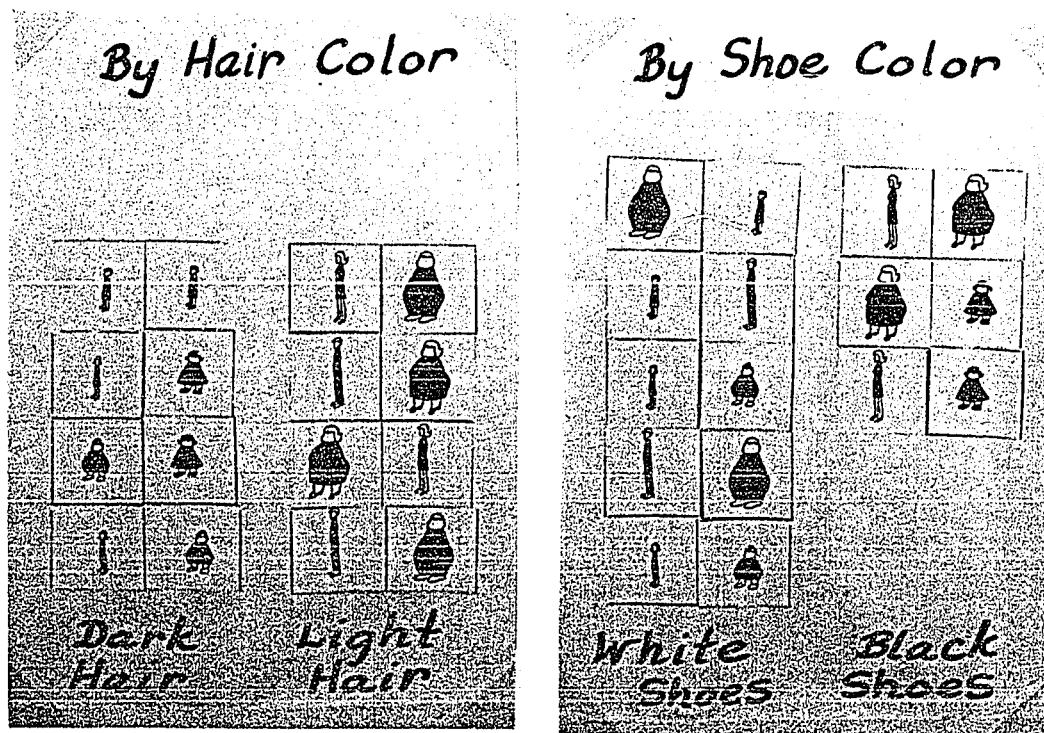
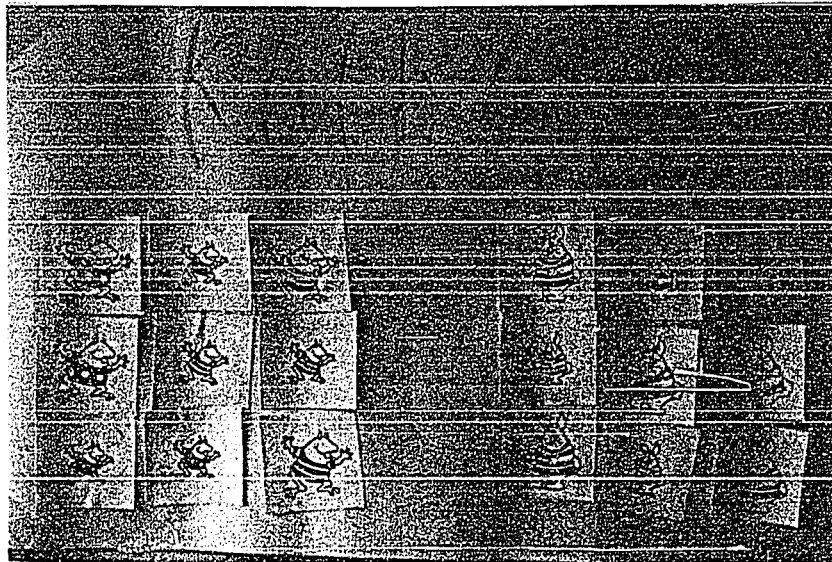
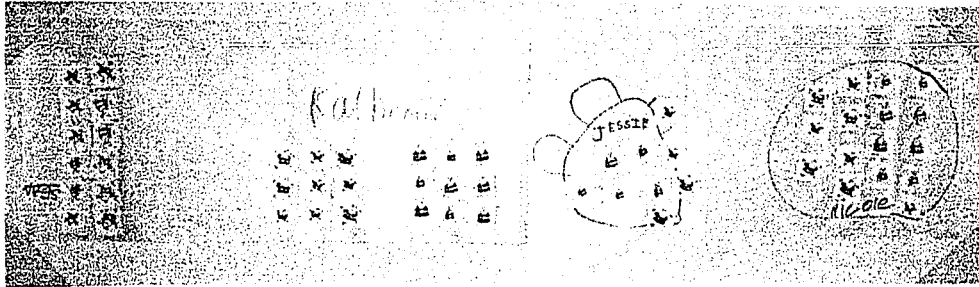


Figure 5.2. Detail from teacher's People Pieces product used to scaffold between People Pieces products and Toad and Rabbit products (shows two of the six ways to sort People Pieces).

Group IV's Toad and Rabbit (Post-Test) Product

Three out of three of the group's original members have sorted accurately by three different attributes. The fourth member entered the class late and took the place of a group member who left early. The 4th member sorted the same as another group member.



Detail from Gr. IV's Toad and Rabbit (Post-Test) Product

(Shows how original accurate sorter sorted)

Figure 6.1. Group IV's Toad and Rabbit (post-test) product and a detail from the group product.

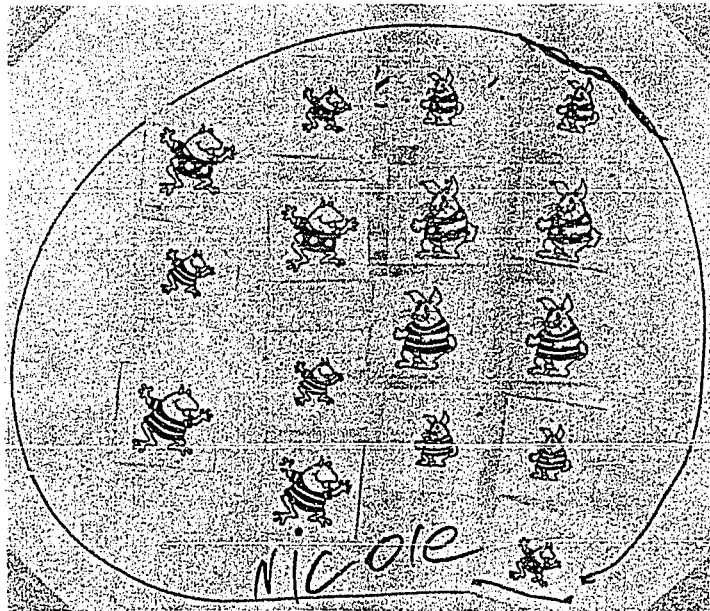
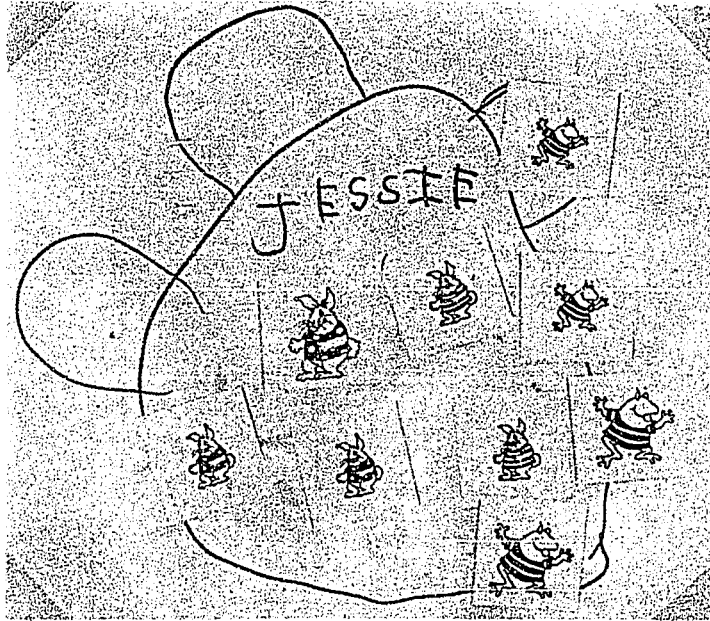


Figure 6.2. Detail from Group IV's Toad and Rabbit (post-test product showing how members of the group other than original accurate sorter sorted (continued next page).

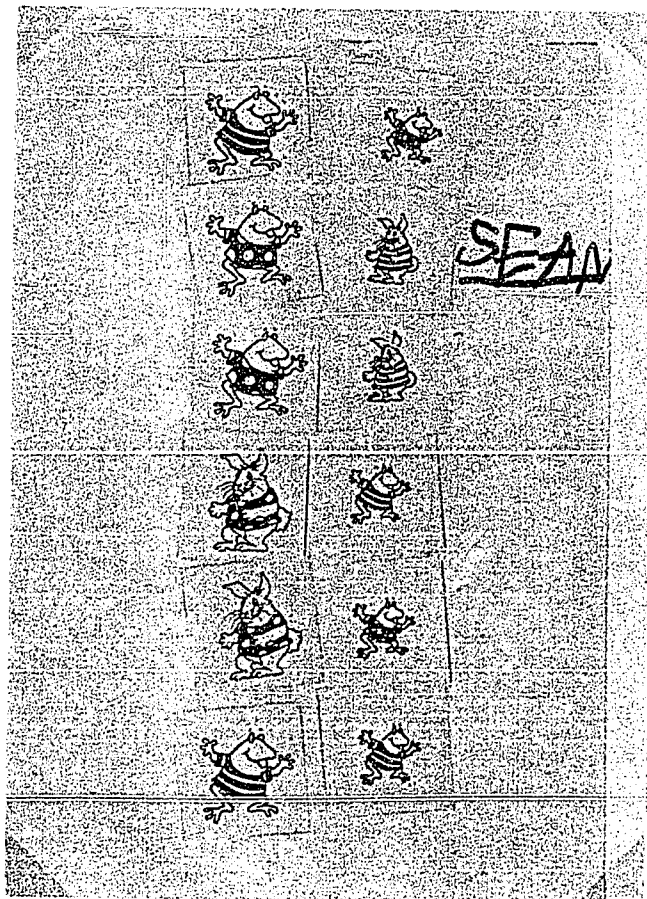


Figure 6.3. Detail from Group IV's Toad and Rabbit (post-test) product showing how one member of the group other than the original accurate sorter sorted.

ways to sort; Groups 1, 3, and 4 each found three ways; and Group 2 found four ways (See Table 1). All groups did a better job of getting more of the possibilities on the posttest phase of the project than they did on the pretest phase, which may have indicated that at least some children were attempting to work together and pay more attention to what other group members were doing. However, the many "I" statements on the video showed that many children were still thinking of themselves as individuals rather than as group members.

Squares Project

If the children had problems with the orientation of their People Pieces products, they had even more with their Squares products. Not only did children not collaborate and not work on the same side of the product, but once the squares were cut out, they were easily turned so that it was impossible to see how they had been colored in the first place.

On the individual sort, where the turning of squares was not a factor, a large percentage of the class improved in the number of different squares they were able to find the second time they did the project over the first: 18 out of 21 improved, one stayed the same both times, and two found fewer of the possibilities the second time than they did the first. On average, the class as a whole found twice as many possibilities the second time as the first, from an average of approximately a third of the possibilities to an average of

Table 1

Comparison of Ways Individuals Sorted on Group Sorting Products

People Pieces (Pre-Test) and Toads and Rabbits (Post-Test)

<u>From Group 1</u>	<u>People Pieces</u>	<u>Toads and Rabbits</u>
M-9	None (Patterned)	By type of animal
F-8	(Absent)	By color
M-19	By color	By color
M-12	By color	By design on clothing
M-11	None (Patterned)	By color (confirmed by video - looks as if he may have switched to [or from] by design on clothes midstream)
 <u>From Group 2</u>		
F-14	None	By type of animal
M-5	(Absent)	By type of animal
F-7	By girth	By design on clothing
F-1	By girth	By color
F-6	None	By size
 <u>From Group 3</u>		
F-16	None	By type of animal
M-21	None (Patterned)	By color
F-20	None	By design on clothing
F-2	None	By size
F-3	None visible (video says by color)	By design on clothing

Table 1 (Continued)

Comparison of Ways Individuals Sorted on Group Sorting Products

People Pieces (Pre-Test) and Toads and Rabbits (Post-Test)

From Group 4*

M-17	None	(No longer in class)
F-23	(Not yet enrolled)	By type of animal
F-18	By size	By type of animal
F-3	None visible (video says by color)	By design on clothing
M-10	None	By size

From Group 5

People Pieces

Toads and Rabbits

F-15	By color	None
M-13	By color	(Absent)
F-4	By color	By color
M-22	None (Patterned)	By design on clothing

* Please note that there are four (not five) people in Group 4. F-23, who entered the class after the pre-test, replaced M-17, who left before the post-test. Post-test total may or may not have been different if M-17 had stayed. M-17 patterned rather than sorted on the pre-test; F-23 duplicated another team member's way of sorting on the post-test.

approximately two-thirds of the possibilities. The first time the project was done, no children were able to find all the possibilities; the second time, two were. In fact, the first time the project was done, only one child, the highest scoring, found as many as three-quarters of the possibilities. The second time, eight children (or slightly more than a third of the class) found three-quarters of the possibilities.

During the first (pretest) part of the Squares Project, of those working above the class median, seven were females and four were males. On the posttest, six were females and five were males. In addition, on the pretest, the six highest scoring children were all females, but on the posttest, two of the highest scoring were males. Whereas females got the three top scores on the pretest, on the posttest two of the top three scorers were males.

The top scorer on the pretest made exactly the same score on the posttest two months later. Her errors were analyzed on both tests to see which of the possibilities she was unable to find. Seven of the eight dealt with the diagonal, so it is likely that she was working at the top of her developmental potential on both phases of the project.

During the pretest phase of the project, seven children, or almost a third of the class, were unable to follow the directions. In one case the child followed a negative example; that is, she did what was illustrated as not being allowed, subdividing small squares. Four developmentally young boys in the class went to the extent of

putting their green and purple crayons away and taking out other colors. In all four cases, they used more than two colors. In the posttest phase, all children were able to follow the directions.

Children showed a marked ability to persevere the second time over the first. During the pretest stage, only one child indicated the need for more time at the end of the first session and then actually wanted to continue to work on the project the second day. During the posttest stage, eight children asked for more time and all eight used the time for their projects the following day. As a class, the children made approximately the same number of self-corrections on the pretest as they did on the posttest, and on both tests in most cases the self-corrections were viable ones. That is, most self-corrections involved the crossing out of duplicates, not the crossing out of new examples that would have added to the child's score. Children made on average fewer duplications on the second phase of the project than on the first. Many children had no duplications either time (16 out of 22 on the pretest, 11 out of 22 on the posttest). Of those who had the highest number of duplications on the pretest, F-3 dropped from eleven duplications on the pretest to one on the posttest; M-12 dropped from 10 to zero, and M-13 dropped from 16 to zero. Of these three, F-3 and M-12 took more time with their work the second time than the first. On the other hand, eight children had more duplications on the posttest than they had on the pretest. (See Table 2 for how children compared to themselves across individual Squares pre- and posttests.)

Table 2

Children Compared to Themselves across Individual Squares Pre- and Post-tests

<u>Student</u>	<u>Ways</u> ¹	<u>Self-Cor</u> ²	<u>Repeats</u> ³	<u>Fol. Dir.?</u> ⁴	<u>More Time?</u> ⁵
M-12	2-16	0-0	10-0	Yes-Yes	No-Yes
F-18	10-16	(1) ⁶ -1	1-1	Yes-Yes	No-No
M-22	5-15	0-0	0-1	Yes-Yes	No-Yes
F-20	10-14	3-0	2-3	Yes-Yes	No-No
F-8	4-14	1-0	0-4	Yes-Yes	No-Yes
F-3	7-13	0-0	11-1	Yes-Yes	No-Yes
F-1	12-12	0-0	0-0	Yes-Yes	No-No
F-16	9-12	0-1	0-0	Yes-Yes	No-Yes

¹ First number indicates number of different ways child found to color squares on pretest. Second number indicates number of different ways child found to color squares on posttest.

² First number indicates how many self-corrections child made on pretest. Second number indicates how many self-corrections child made on posttest.

³ First word indicates how many duplications child had on pretest. Second word indicates how many duplications child had on posttest.

⁴ First word indicates whether child followed directions on pretest. Second word indicates whether child followed directions on posttest.

⁵ First word indicates whether child used more time on pretest. Second word indicates whether child used more time on posttest.

⁶ Number in parentheses indicates that it is a non-viable correction; i.e., that child crossed out a viable alternative. Number not in parentheses indicates that child had discovered and deleted a "repeat."

Table 2 (continued)

Children Compared to Themselves across Individual Squares Pre- and Post-Tests

<u>Student</u>	<u>Ways</u>	<u>Self-Cor</u>	<u>Repeats</u>	<u>Fol. Dir.?</u>	<u>More Time?</u>
M-5	3-11	0-0	0-1	No-Yes	No-No
M-21	0-11	0-0	0-1	No-Yes	No-No
M-19	1-11	0-0	0-7	No-Yes	No-Yes
F-15	0-10	0-0	1-0	No-Yes	No-Yes
F-6	3-10	0-0	0-2	Yes-Yes	No-Yes
M-11	7-9	0-0	0-0	Yes-Yes	No-No
F-7	9-8	0-0	0-0	No-Yes	No-No
F-23	--8	--0	--1	--Yes	--No ⁷
F-4	3-7 (1) 4-0		0-0	No-Yes	Yes-No
M-9	6-7	1-0	0-1	Yes-Yes	No-No
M-10	4 (6?)-7	1-7	0-2	Yes-Yes	No-No
M-13	2-6	0-0	16-0	Yes-Yes	No-No
F-2	5-6	0-1	0-0	Yes-Yes	No-No
F-14	8-5	(1)-(2)	0-0	Yes-Yes	No-No
M-17	1--	0--	0--	No--	No--

⁷ This child asked for more time at the end of the first day, but by the beginning of the second day decided she was finished.

If children were more successful on the posttest than on the pretest for the individual sort, however, the same could not be said for the Squares group sort. This project was beyond the capabilities of all but one group, which contained one child, the leader of the group, who was an older child who had been to a transitional kindergarten for a year before coming to this kindergarten. It appeared from the products that all groups but one continued to work without consulting each other, work on opposite sides of the product, and produce work that was virtually undecipherable. However, the videotape indicates that some children, while not there yet, were moving in the direction of attempting to work together.

The one group that succeeded chose a leader, worked together as a group, worked on the same side of the product, and got all sixteen of the possibilities with no duplications. In addition, group members put their names together at the top of their product and, when they were interviewed in front of the video camera, used "we" instead of "I" in describing their work. Also, when an audience member asked, "Why didn't you put a line around your own group?" the group leader replied, "We weren't supposed to. This (indicating all the group members) is a group." (See Figures 7.1 - 7.3 for Squares group pre-test products, Figure 8 for teacher's scaffolding product, and Figures 9.1 - 9.4 for Squares group post-test products.)

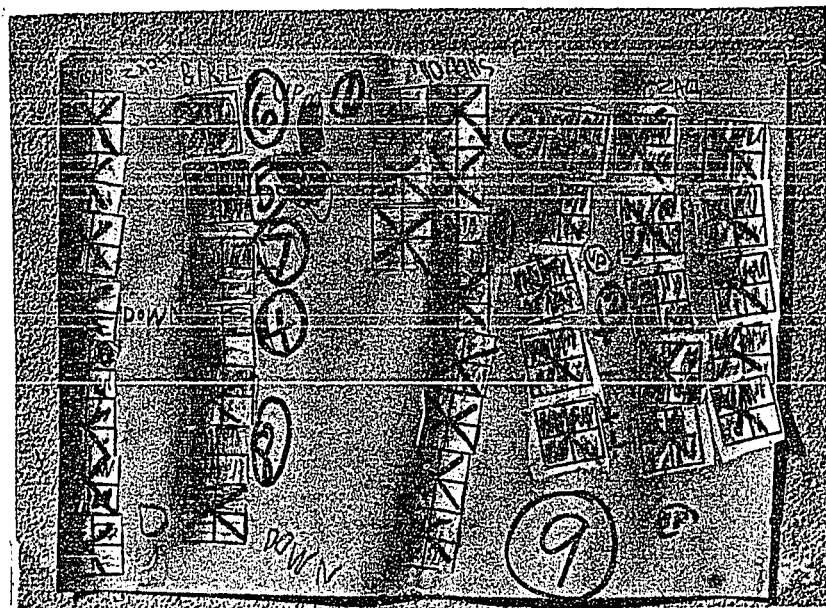
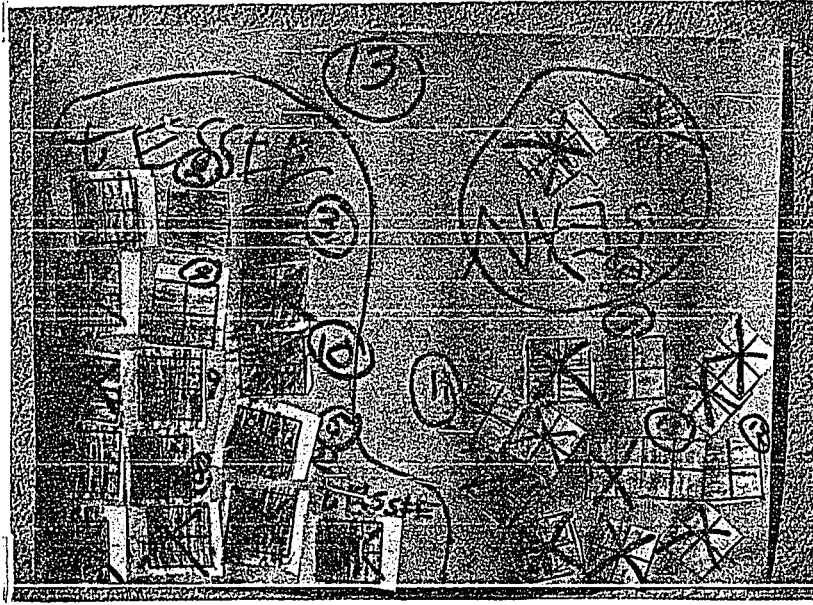


Figure 7.1. Purple and green squares (pre-test) products (Nobody got it).

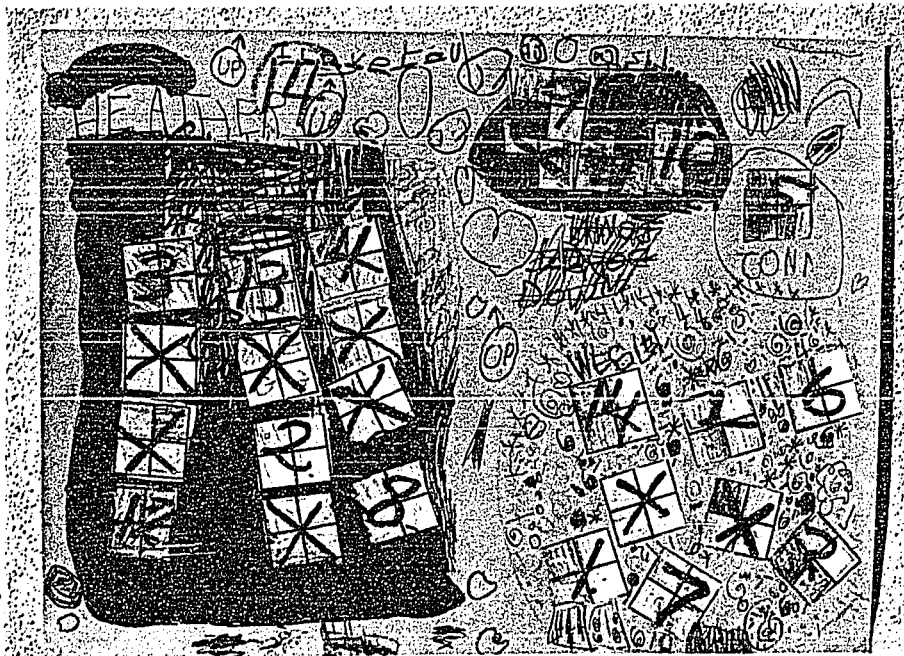
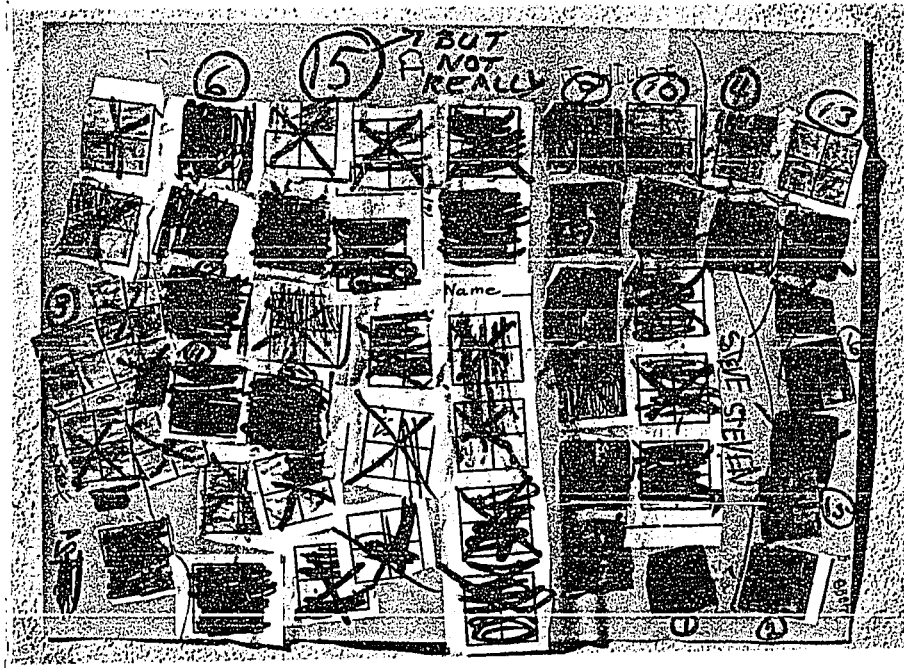


Figure 7.2. Purple and green squares (pre-test) products (Nobody got it).

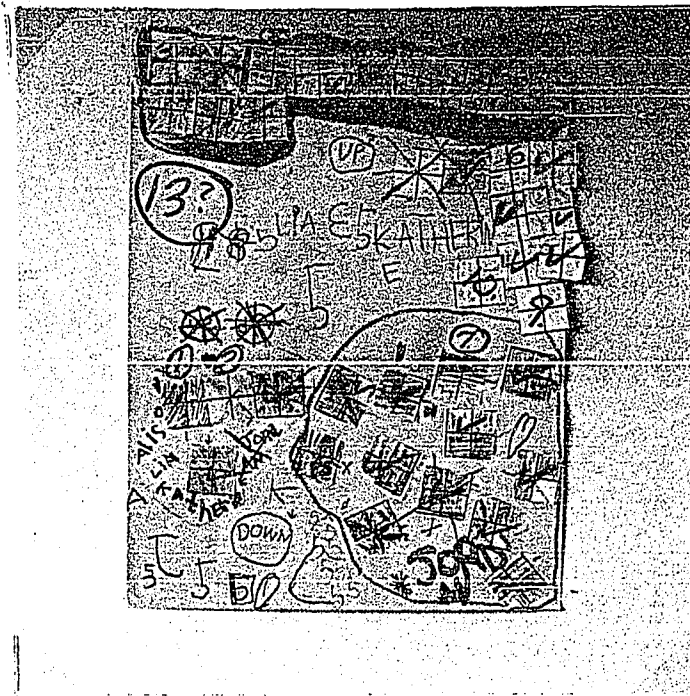


Figure 7.3. Purple and green squares (pre-test) products (Nobody got it).

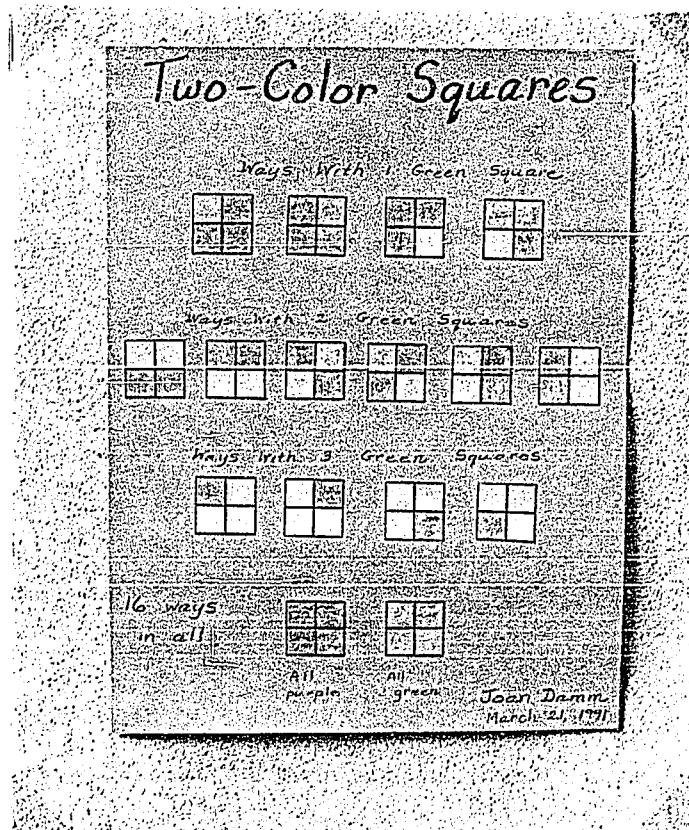


Figure 8. Teacher's two-color squares example used for scaffolding between student attempts.

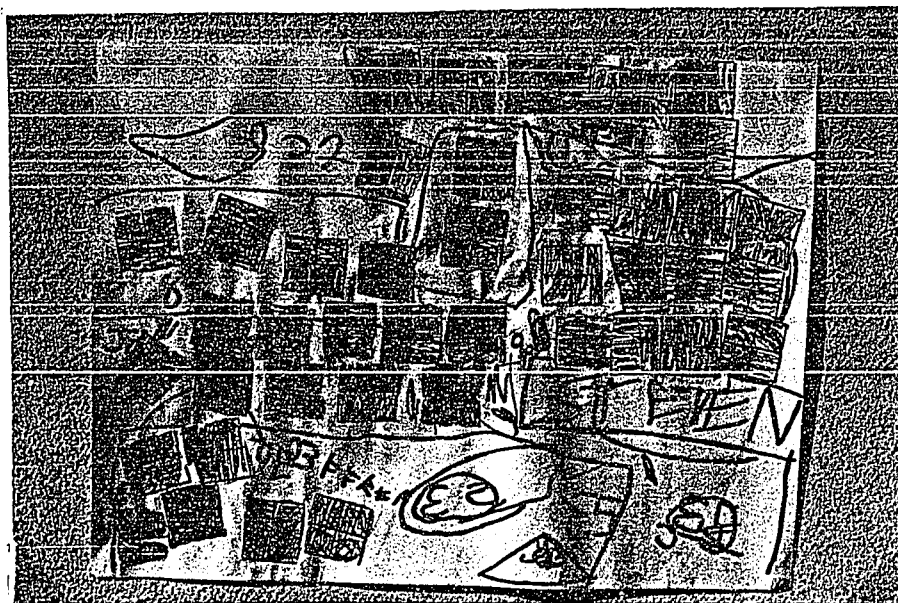
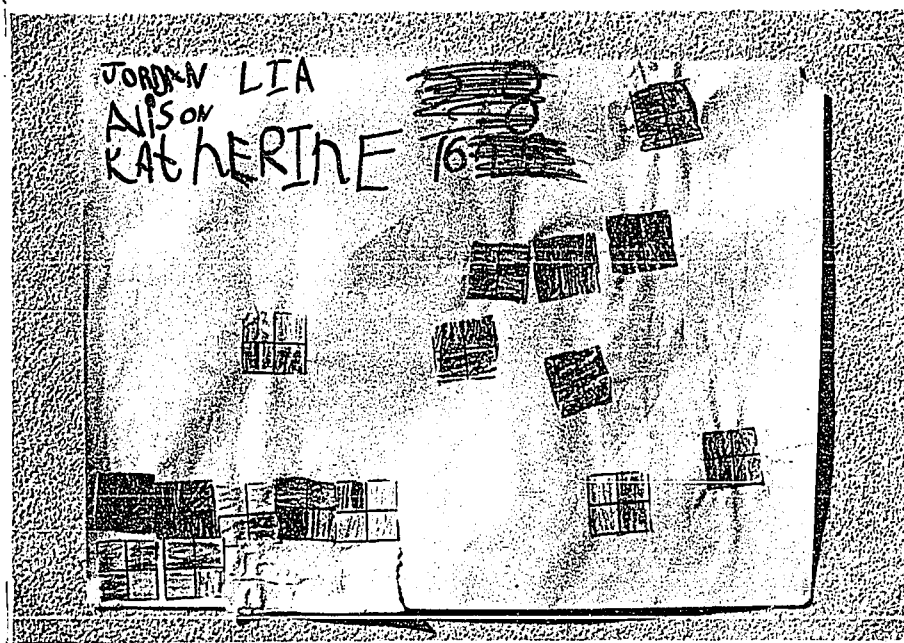


Figure 9.1. Blue and orange squares (post-test) products: Best (one) to worst (all the rest) (continued on next three pages).

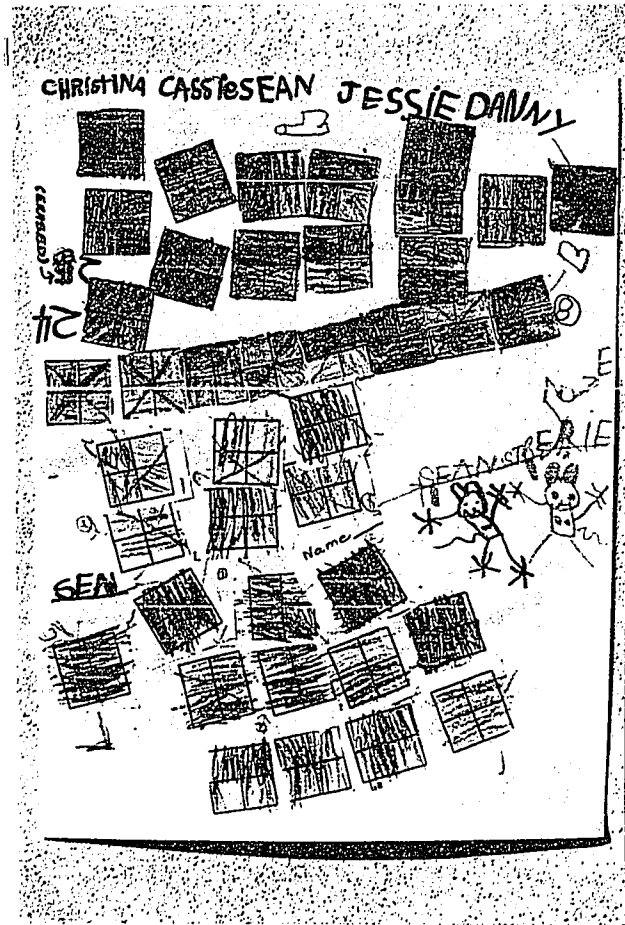


Figure 9.2. Blue and orange squares (post-test) products.

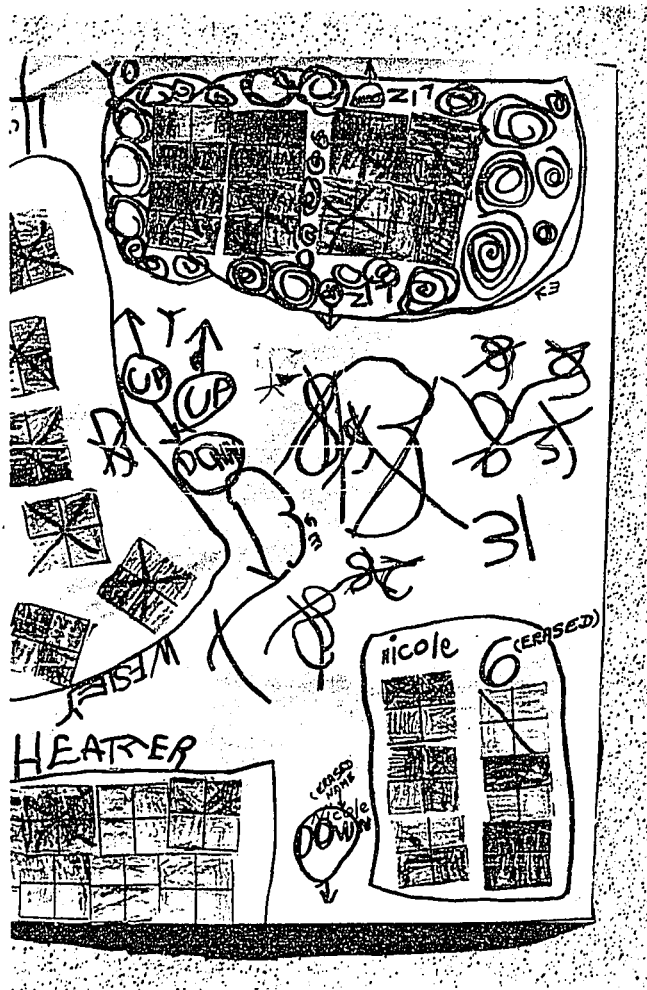


Figure 9.3. Blue and orange squares (post-test) products.

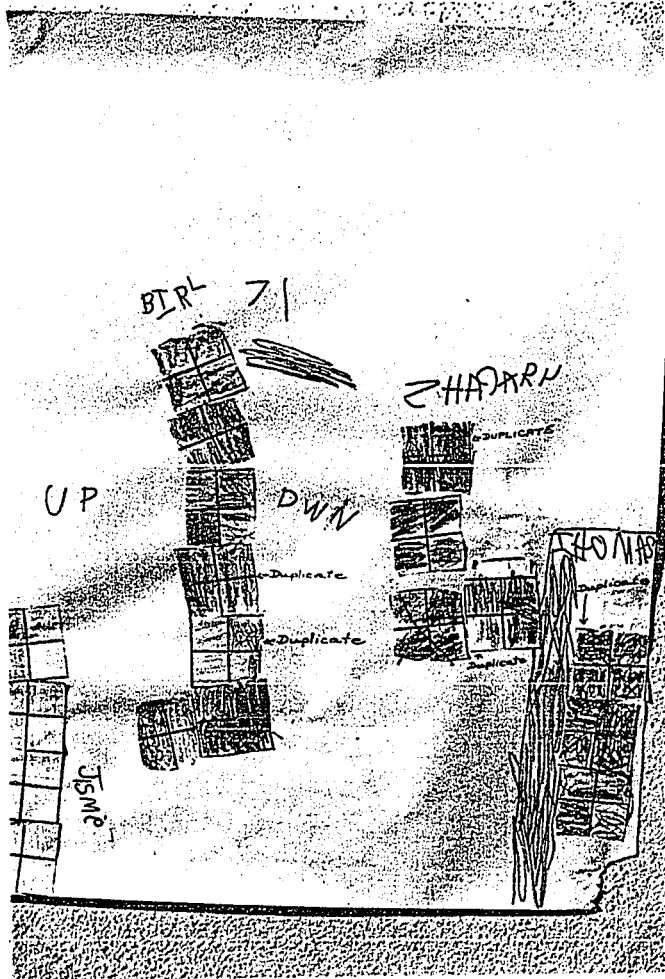


Figure 9.4. Blue and orange squares (post-test) products.

Nursery Rhyme Project

The Nursery Rhyme project, too, was time-consuming and labor-intensive. First the teacher had to meet individually with each child to ascertain which nursery rhymes the child could say by heart. (During the "search" phase of this project, the teacher was able to discover which children had learned which nursery rhymes, ranging from one child who knew none to another who could recite a great many.) Then the teacher had to print all the applicable rhymes on sentence strips and have them laminated. After that, the teacher met individually with each child to watch as the nursery rhyme was reconstructed in the pocket chart. The reconstructions took a long time, up to twenty minutes each, because almost every child went back to the beginning of the rhyme and said it again aloud each time he or she placed a word. Recording what the children were doing was difficult as well; in many cases, about the time the teacher had written down what a child had done, the child rearranged his or her work.

This project was still in progress at the end of the school year; the teacher ran out of time to complete it. Consequently, pictures are available for only part of the class. Those that are available show what children are able to do--or not do--in the way of reconstructing known nursery rhymes. The project also gave the teacher insight into each child's level of motivation, tendency to persevere, and level of confidence.

Of those children who completed the project, except for one

fluent reader, all appeared to be using the initial consonant to determine which word was which; when an incorrect word was chosen, it often had the same initial consonant as the correct word. The fluent reader was the only one who was able to leave blanks in her construction and fill them later. When she had all but two words placed and had left blanks where thes should go, she turned to the teacher with identical "the" cards, one in each hand, and said, "I know that *this* says the and *this* says the, but which goes in which blank?"

The photographs "froze" moments in time in each child's level of literacy acquisition. By reviewing the photographs, the teacher was able to see both individual achievement levels and the range in the class as a whole. (A series of such photographs for each child could document progress over time.) For the students who were able to reconstruct their nursery rhymes well, the photographs showing them with their work confirmed both their progress and their feelings of self-esteem. Most of the children who were unable to do the project well showed their distress in their expressions. (See Figures 10.1 - 10.8 for student generated Nursery Rhyme products.)

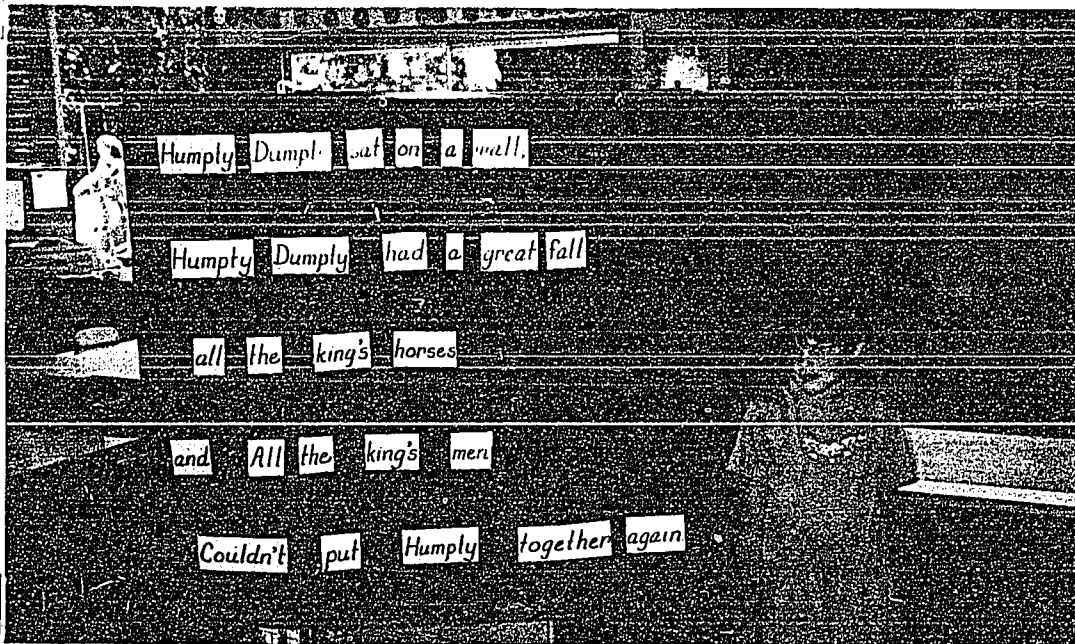
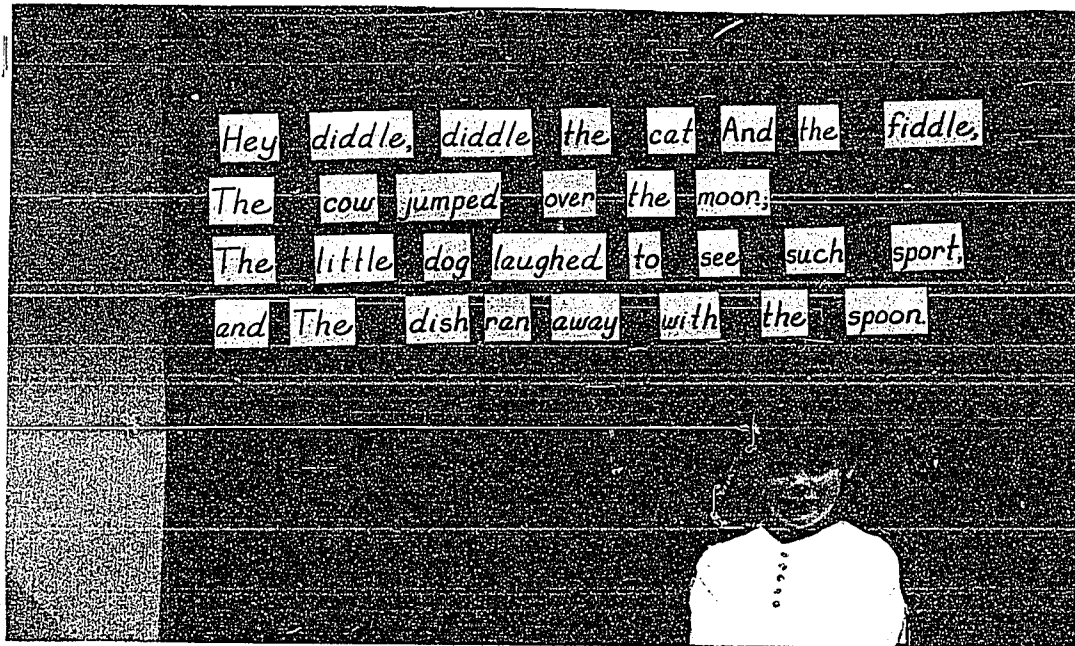


Figure 10.1. Student-generated nursery rhyme reconstructions (best products to worst). See also next seven pages.

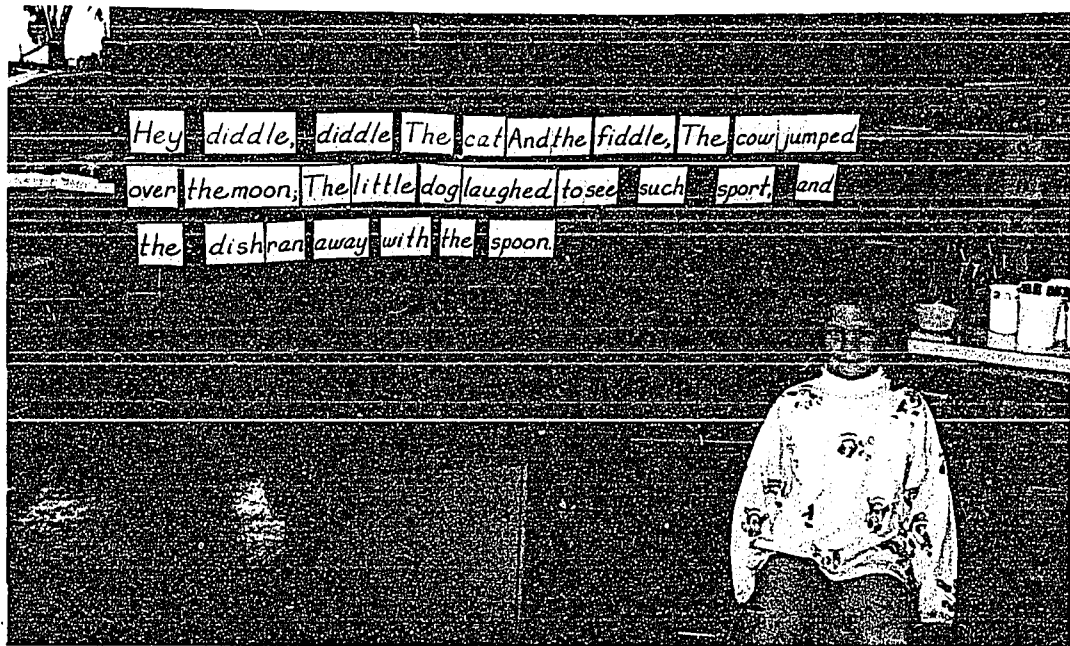
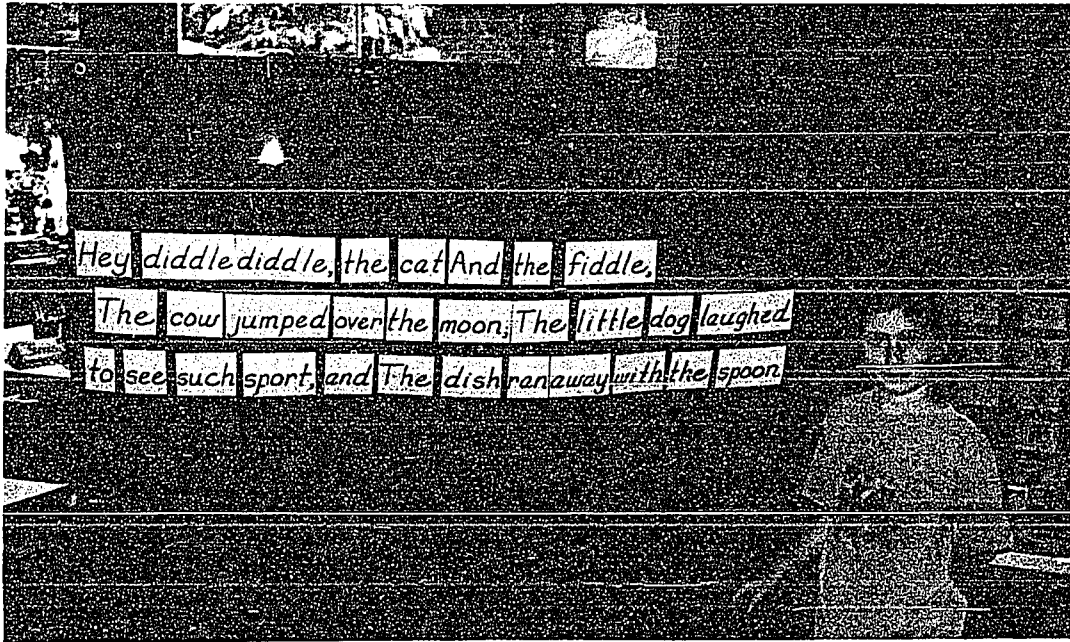


Figure 10.2. Student-generated nursery rhyme reconstructions (best products to worst).

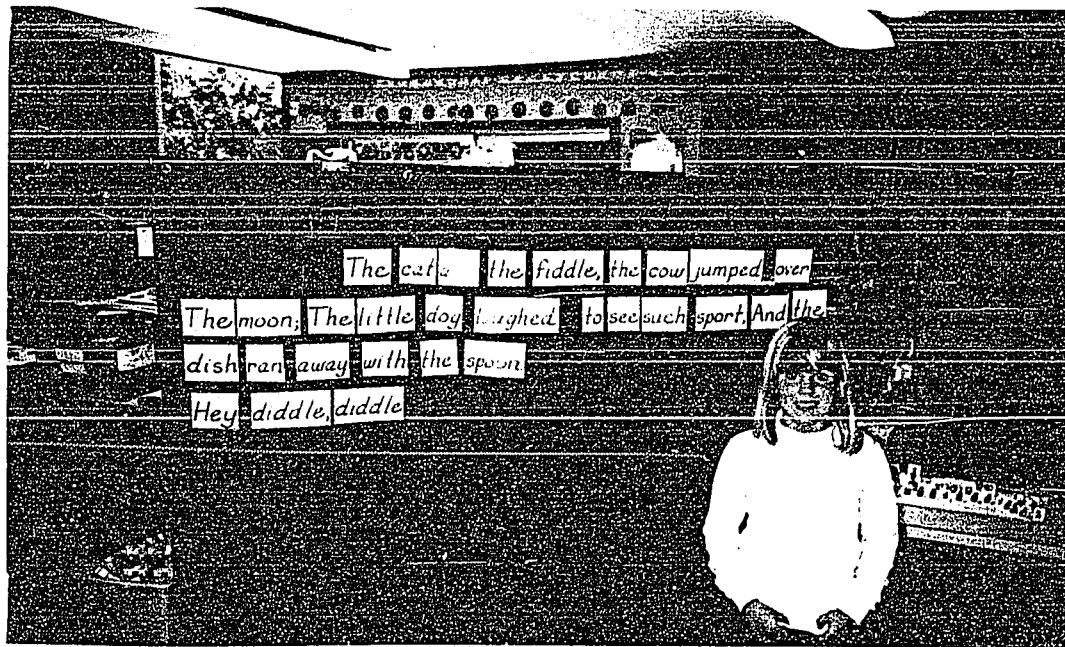
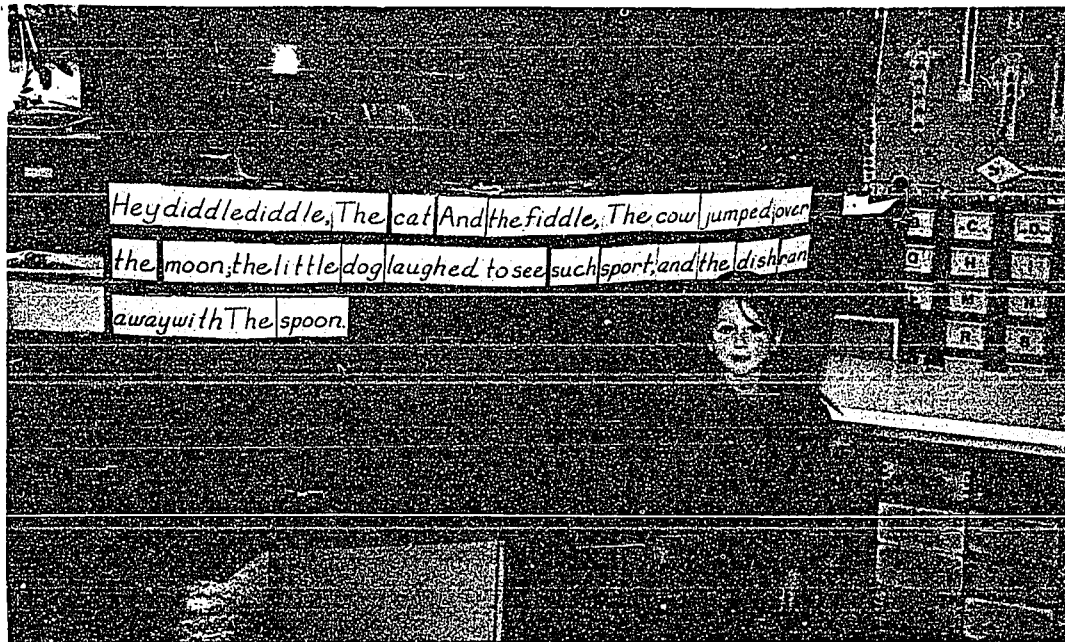


Figure 10.3. Student-generated nursery rhyme reconstructions (best products to worst).

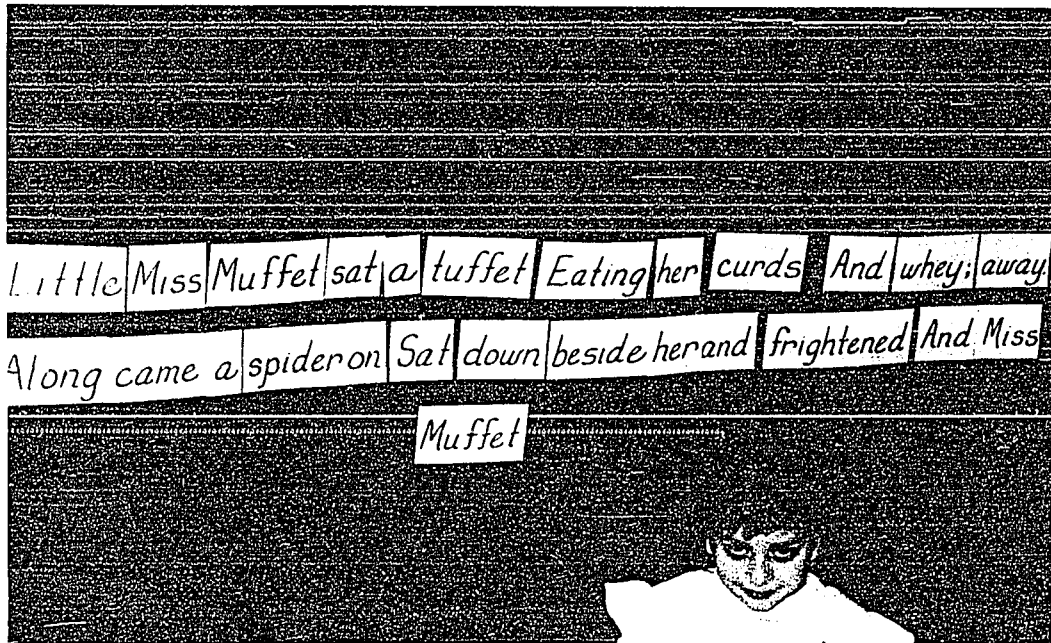
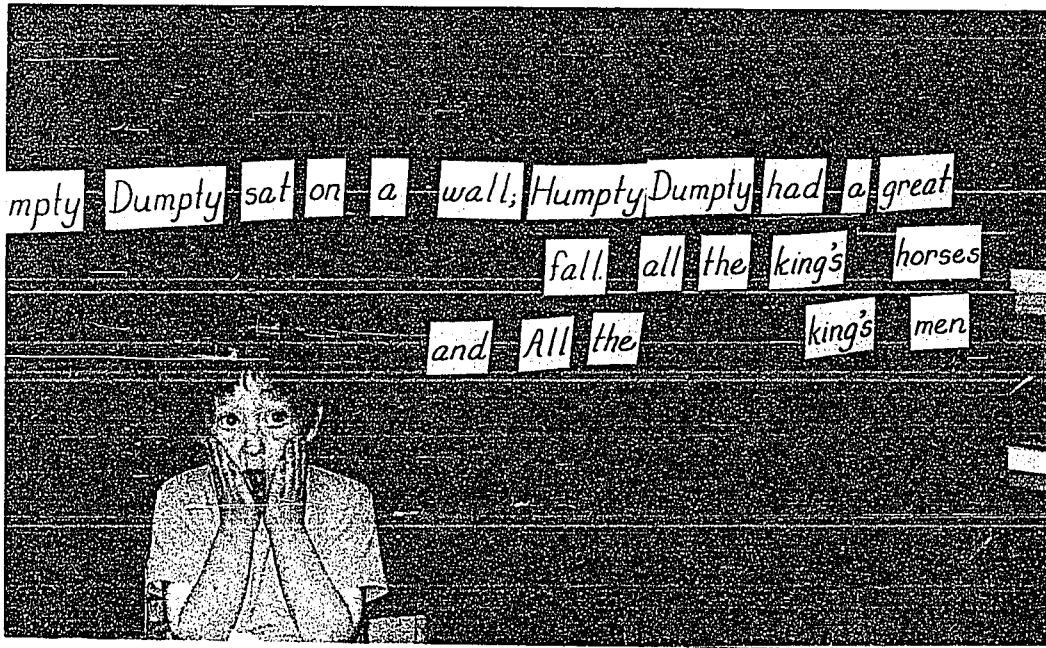


Figure 10.4. Student-generated nursery rhyme reconstructions (best products to worst).

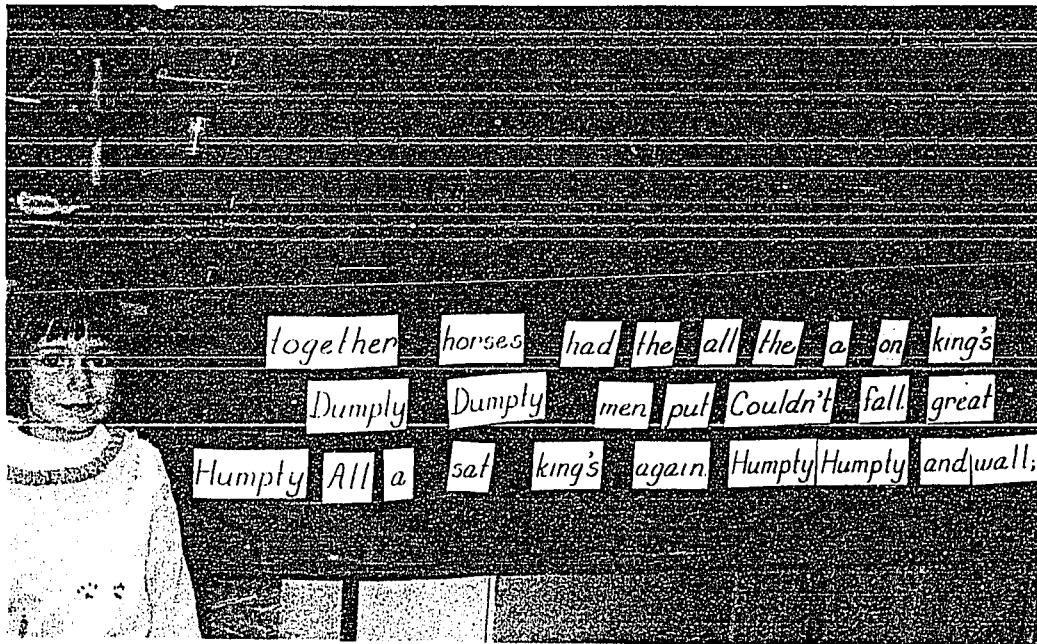
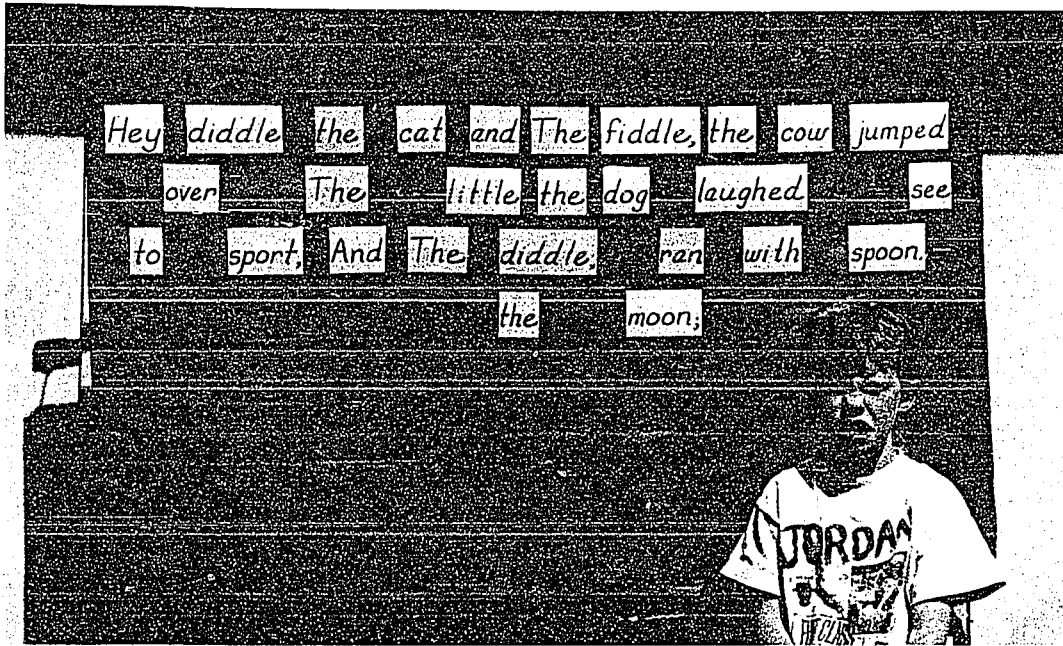


Figure 10.5. Student-generated nursery rhyme reconstructions (best products to worst).

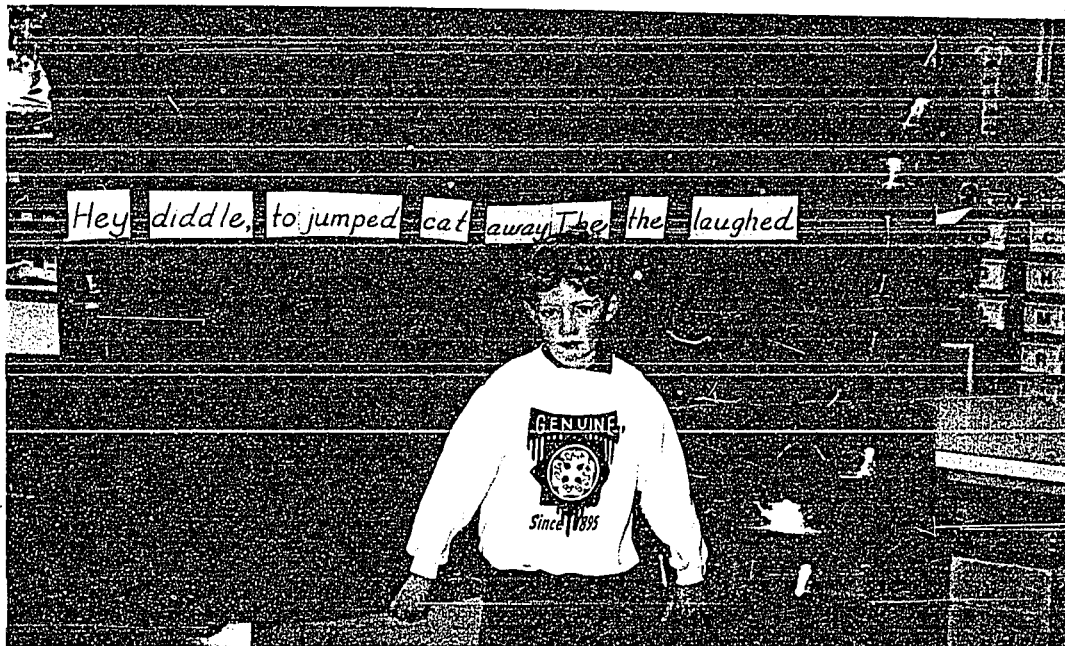
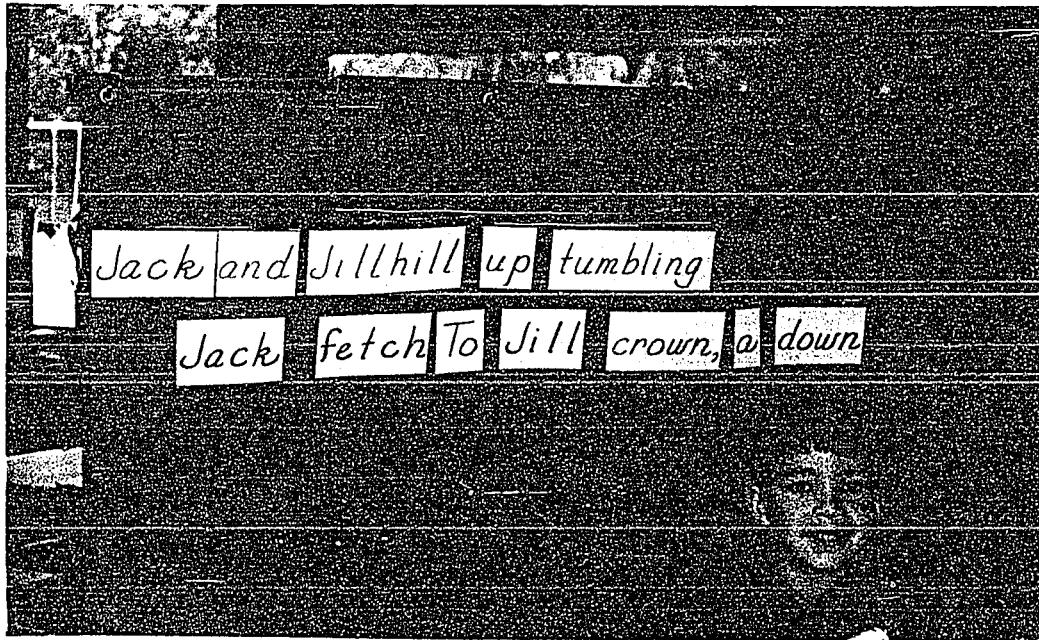


Figure 10.6. Student-generated nursery rhyme reconstructions (best products to worst).

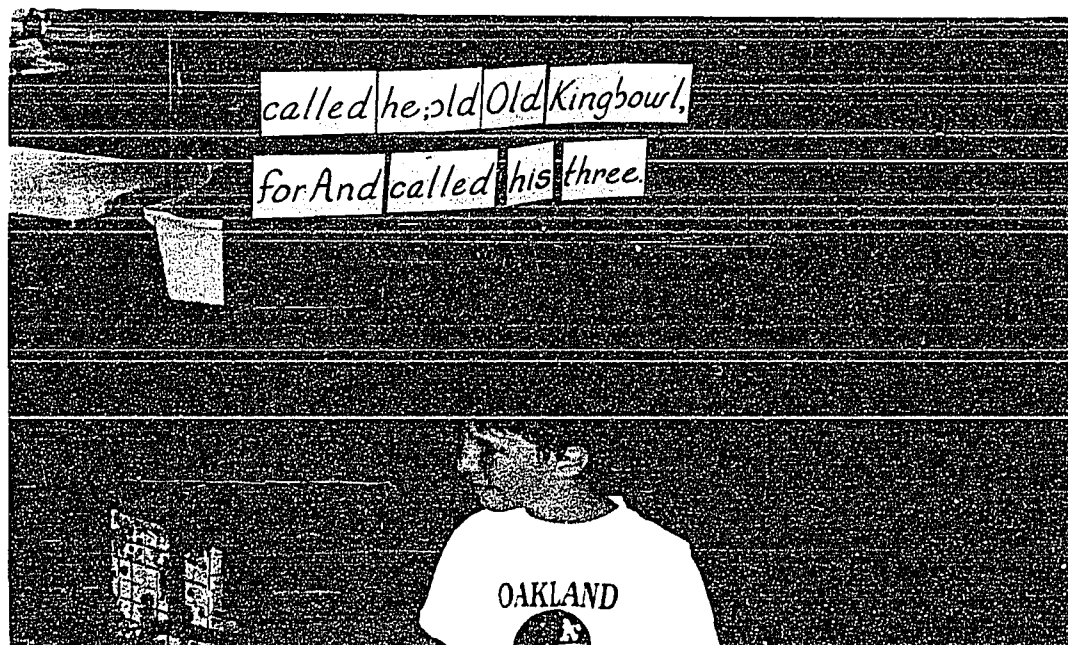
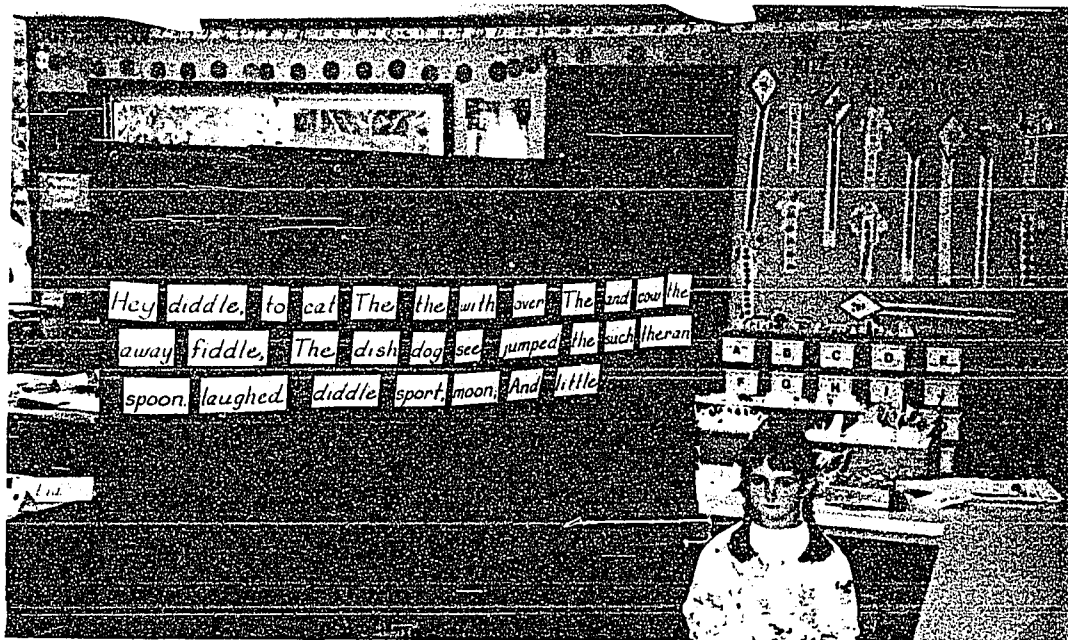


Figure 10.7. Student-generated nursery rhyme reconstructions (best products to worst).

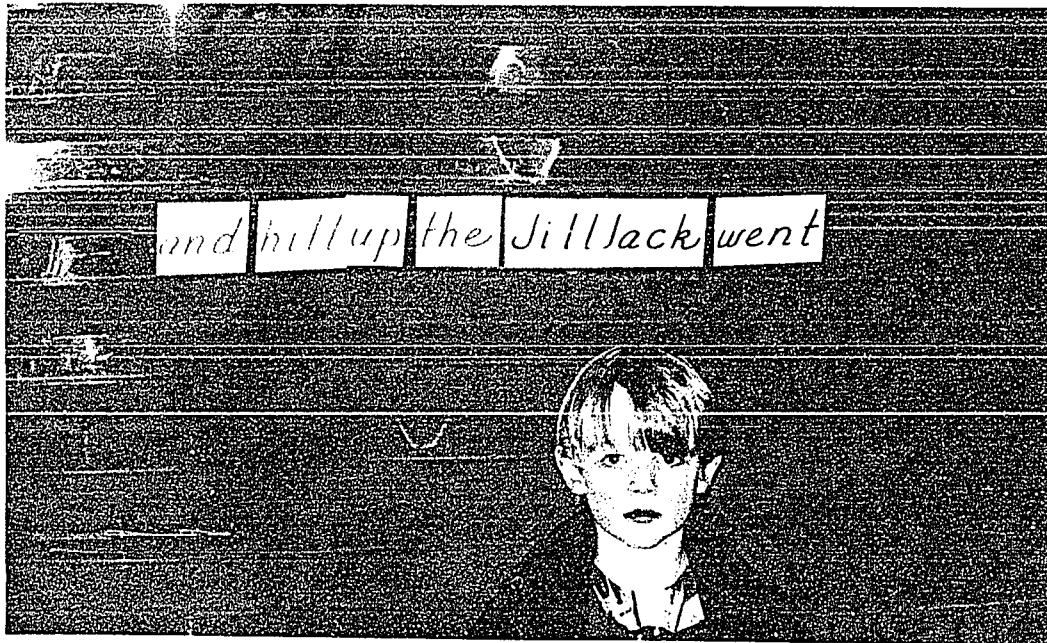
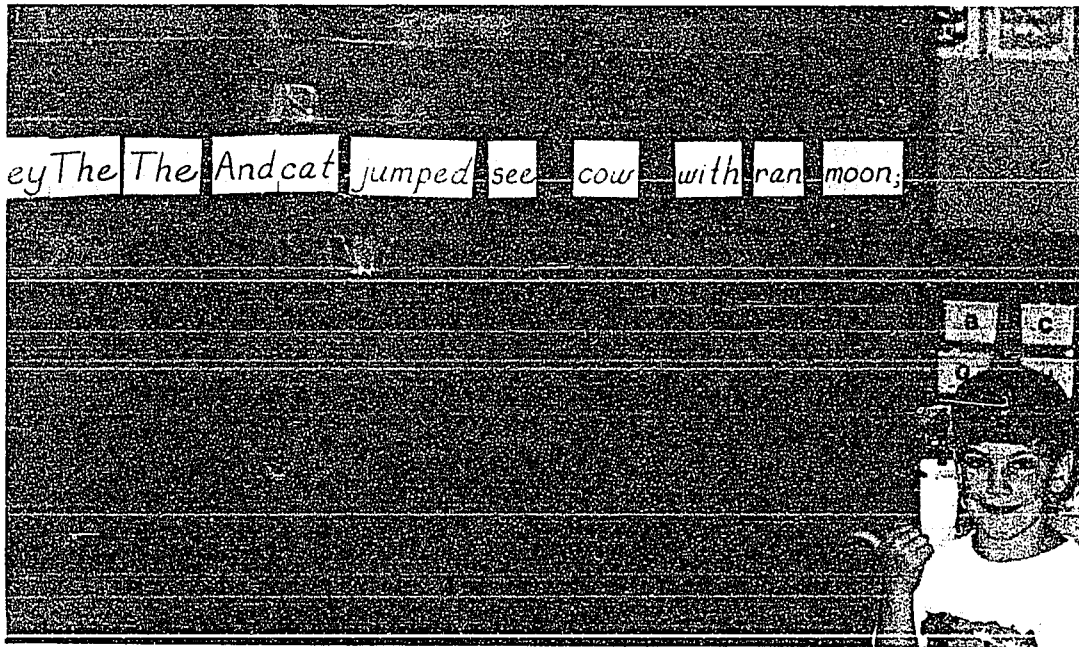


Figure 10.8. Student-generated nursery rhyme reconstructions (best products to worst).

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

In Chapter 5, each of the five projects will be discussed separately, interactions between People Pieces and Squares projects will be discussed, and suggestions for improving the assessment of kindergarten students will be offered.

Portfolio Assessment of Journals

The procedure that evolved for journal writing was somewhat valuable from both a teaching/learning perspective and from an assessment perspective. Working together to decipher topics from the board gave children practice in being members of a collaborating team under teacher direction, and the teacher was able to see which children were beginning to decode and how they went about using context clues to do so. However, the activity appeared not to be appropriate for all children involved, and it was felt that reading aloud to the children, doing pocket chart activities with them, and reading Big Books with them would have constituted better use of kindergarten time.

Likewise, there was some value for both teacher and student in working together to choose "best" pages for "keeper" journals. Students were given the idea that it was both logical and necessary for them to evaluate their own work, and their self-esteem was enhanced by having their opinions sought and respected; the teacher

learned more about children's interests and concerns and was able to get to know them better than would have been the case had there been no opportunity to talk to them individually and privately.

However, working individually with each child to help him or her choose a "keeper" journal page a week was unwieldy and took an inordinate amount of time and organization. In addition, the children did not appear mature enough to have adequate reasons, in the teacher's opinion, for choosing their "best" pages. They "just liked them." (This activity did constitute a first step, however, toward developing the "maturity" to judge.) Exacerbating the problem of evaluating writing was the fact that most of these kindergartners were not yet writing much; most of their journal pages contained drawn pictures or pictures with a few words. On balance, the teacher felt that the benefits derived from choosing "keeper" journal pages were not worth the amount of time and effort it took to do the choosing. In a full-day kindergarten with a fulltime aide, after children had had a good deal of oral whole-language input, the plan might have been a viable one, but in a half-day kindergarten with a partially-mandated schedule and a part-time aide, both the writing of journals and the choosing of "keeper" journal pages constituted a less-than-optimal use of kindergarten time. A better plan would be to write in journals just two days a week and to keep everything.

For older children, however, portfolio assessment of journal writing might be one good way to document writing progress over time. Older children are able to write more than younger children

and are also capable of working much more independently, so conceivably less teacher time would be required for the choosing of "keeper" pages. Also, after the children had had more experience at journal writing and after they were a little older, perhaps seven, it would be easier to set standards with them for the choosing of "keeper" journal pages. On the other hand, older children might be interested in keeping all of their journal pages, and if only certain pages were kept, valuable information might be lost.

VIP Experience Stories: Tracking via Computer

Gibboney (1989) intimated in regard to educational research done by measurement specialists from outside the classroom that visual acuity does not compensate for the misdirected eye. Ironically, it appears that, at least to a degree, that statement could also be made about this project. Whereas it was confirming to find that most of these kindergartners did indeed write more as the year progressed, it was felt that the great deal of time involved in typing 476 experience stories into the computer and then using the word count feature to count them was time that could have been better spent any number of other ways. In the first place, it seems reasonable to assume that children would gain better command over the written word as they got older and consequently would write more. They do it naturally with spoken language; it seems reasonable to expect that they would do it naturally with written language as well.

Complexity of sentence structure, which is sometimes shown partially by length of utterance, could indicate growth in writing ability; but kindergarten experience story entries can be long without being complex. The teacher felt that some children were going on and on to hold the teacher's attention; others stated that they wanted to keep going until they got "all the way to the bottom of the page." In addition, concentrating on quantity precluded interrupting children and changed teacher behavior; in an effort to capture all of what children had to say, the teacher sometimes allowed them to continue when under ordinary circumstances they would have been interrupted; in addition, (possibly) "teachable moments" were allowed to slip by, as when children repeatedly said, "Me and Jim did thus and so."

The two results from this project the teacher found interesting were that boys, on average, wrote more than girls; and that lines of production flattened out over two-week vacation periods. The first result may be idiosyncratic to this class, a result of the interactions between this particular teacher and this particular group of children, or result from small sample size; the second may have implications for the way school years are structured. At first it was thought that perhaps the lines flattened out because children needed "breathing space" every so often to consolidate their learning (and that may also be true), but close examination of the class graph revealed that lines of production flattened out over all two week vacations, even though the vacations were not evenly spaced.

More research would have to be done to ascertain whether this marking of time is a usual occurrence and whether it occurs at other grade levels. Should this project be replicated, it is strongly recommended that a scanner be used. It is also recommended that graphing be done on a computer and that computer proficiency be a requirement in teacher training programs, including graduate programs.

Whereas counting and graphing the children's words was somewhat counterproductive (i.e., not worth the time it took), recording the children's stories in the computer was not; it made the researching teacher aware of things that previously had gone unnoticed, such as that some children repeat the same thing or the same thing with slight variations each week. It was consciousness raising to do this project once. As an alternative assessment technique for kindergartners, however, it falls short of the mark.

People Pieces Project

The People Pieces Project has possibilities as an alternative assessment technique for use with kindergartners, as it does allow a teacher to assess both individual sorting ability and progress toward working as a member of a group. As a precursor to it, however, it might be a good idea to have the children paint group murals. Kindergartners do the same thing with murals that they did with recording the People Pieces; that is, they work on opposite sides of the paper so that two children working across from each other have

the sky on opposite edges of the paper. That might be a gentler and more concrete way to convince kindergartners that group members producing a group product designed to go on a wall need to work on the same side of the product. Also, if this project were to be done again with kindergartners, the teacher should make all the recording sheets--preferably on a color copier (provided that the school had a color copier, which is highly unlikely, or that doing so was not excessively expensive). Having kindergartners produce their own recording sheets is both time consuming and, in some cases, inaccurate.

Videotaping was a good idea, in spite of all the problems that went with it, not only because the videotapes provided information that was not available merely from looking at the products, but also because the products eventually began to come unglued and the videotapes constituted the only permanent record. In addition, it was evident in comparing the earlier tapes and the later ones how much progress the children had made, both as audience members and as presenters. In the early tapes, neither audience members nor presenters could pay attention to the task at hand. By the last tapes, audience members were having substantive conversations with presenters and showed by their thoughtful questions that they had listened to what the presenters had had to say. It might be viable to use this project as a baseline in kindergarten and then to do the same project in second or third grade and videotape it again. It might also be viable to use the videotapes as a teaching tool, showing

them to the children and inviting discussion. In viewing the early videotapes, the researching teacher was somewhat disturbed by how hard the taping was on some of the children, as evidenced by the amount of moving around they did and the number of ways they found to distract themselves from the task at hand; she questions taping for such long periods of time with children so young. Another time, taping sessions might be done more often but for shorter periods of time.

It might be informative to do this project and a series of Piaget tasks with the same children and compare the results. An even better idea might be to do this project in second grade, when some of the children would be likely to be conservors and others would not. There again, Piaget tasks could be done with the same children and the results compared.

Squares Project

The Squares project was too difficult for most kindergarten children. It might be a useful alternative assessment technique for a slightly older grade level. Having a color copier available to record the children's individual products would have made the project quicker. One way to make alternative assessment more viable would be to be sure that schools had state-of-the-art equipment.

For the researching teacher, it was worth all the frustration over videotaping, undecipherable products, and arguments over what was "down" and what was "up" when Jordan's group (the group

that got all 16 squares on the posttest) was asked by Steven, "Why didn't you put a line around your own group?" (meaning the squares on the product) and Jordan answered, "We weren't supposed to. This is a group" (indicating herself and the people with her). Growth in learning could be provoked using this project, and that growth could be documented through videotaping.

If videotapes are to be shown to parents, however, (and presumably they would be if they are to constitute part of an alternate assessment reporting procedure), a few cautions should be considered. Whereas videotapes can show graphically what some people's children are capable of doing, they show at least as well what other people's children are not capable of doing. Videotapes have the potential for invading children's privacy and also for causing human relations problems with parents. A good deal of parent education would have to be done before using videotapes with parents as part of an alternative assessment reporting procedure. Another caution is that, whereas usually videotapes clarify what is actually occurring (as in the case of the child who reported, "These are the skinny ones and these are the fat ones..."), there are times when information given on videotapes can be convincing but misleading. For example, when the children were asked what they were supposed to do on the People Pieces project, one child answered, "'Vide 'em out [divide them out]. Nobody in our group did it." That was not true. One member of her group did "'vide 'em out." Another child went to great lengths to explain how

he had sorted his pieces, in spite of the fact that he had been absent the day the project was done.

Both the People Pieces and the Squares projects seemed too abstract for almost all kindergarten children. Kindergarten children should be working with real objects as much as possible. Before undertaking a project like the People Pieces Project, they should have a lot of experience sorting rocks, leaves, buttons, attribute blocks, and so on, all tasks that are more authentic. Note that over all groups and four products for each group (namely People Pieces, Toads and Rabbits, Purple and Green Squares, and Orange and Blue Squares), only one group actually managed to collaborate fully and work as a total group by the time they had completed their last product (although the videotapes do show that there were many more movements in that direction than one could tell by merely examining the products).

Nursery Rhyme Project

The Nursery Rhyme project was also too time consuming and labor intensive to constitute a viable alternative assessment technique, especially since all the nursery rhymes had to be written and laminated. Another time, with at least some of the materials already constructed, it would take less time--but it is still prohibitively time consuming to meet individually with a class of kindergartners in a half-day kindergarten class more than once a week, and it takes additional adult coverage. It is not possible for a

teacher to be a careful kid watcher while simultaneously watching twenty-one other five-year-olds. The researching teacher also had some concerns about photographing those children, with their work, who were not yet able to decode. Their distressed expressions bear out that they are already aware of what they are not (yet) able to do. This project might make a viable alternative assessment technique toward the end of first grade, when most children are able to read at least to an extent, but probably a better plan would be for a teacher to become proficient at both "kid watching" and miscue analysis. A teacher might be able to do some unobtrusive assessment in either kindergarten or first grade by leaving the nursery rhyme cards out for the children to play with in the pocket chart during Free Choice, but it would be difficult to use this technique for systematic assessment. It would also be more respectful for the teacher to ask children's permission before taking their pictures with their work and to do so only after they had chosen the task and had completed it at least somewhat successfully.

Transfer Between Projects

There appeared to be some transfer between the People Pieces and Squares projects. One child remembered that his group's People Pieces product had been partially upside down when it was displayed on the wall due to the fact that group members had worked on opposite sides of the table and, during the Squares

project, tried (unsuccessfully) to convince his group to work on the same side of the table.

More useful assessment information might be gleaned by trying the People Pieces and Squares projects in second grade; by that level the children would have more experience working as group members and, for at least some of them, developmental level would be less likely to preclude successful completion of the task. Consequently, there should be a wider range of products.

The researching teacher would recommend doing both projects and scaffolding between them, in the same order and in the same manner as was done in this instance. The reason for that is that the teacher felt there was transfer from the People Pieces Project to the Squares Project in addition to that mentioned above. The teacher was not at all sure that the one group who got all the Blue and Orange Squares possibilities would have done so had they not had the previous People Pieces exposure to producing products and displaying work as a group, or had the teacher not scaffolded within projects.

Regarding the projects undertaken for this study, the researching teacher felt that too many projects were undertaken simultaneously, and that too many of those undertaken were developmentally inappropriate for kindergarten children. It would be better to implement alternative assessment slowly, beginning with one ostensibly small project. And even that should not be undertaken until an environment and support system has been

created that allows for the provision of high-quality tasks in all curricular areas and for the possibility of good kid watching.

Conclusion

It seems appropriate here to attempt to answer the questions posed at the beginning of this research, and then to offer some suggestions for the alternative assessment of kindergarten children.

1) Can kindergarten students' "writing" growth be reliably and validly documented using a portfolio system that allows children to choose their own best journal writing? The plan attempted for this research might be a viable one in a full-day whole language kindergarten. In a half-day kindergarten class with partially mandated schedule and less than a fulltime aide, this technique was too time consuming and labor intensive to constitute optimal use of either the children's or the teacher's time. A better plan would have been to write in journals just two days a week and to keep everything. At some point, however, children should have the experience of choosing items for a portfolio and reflecting on the reasons for their choices. Perhaps first grade would be a good time to start, when the children attend school for a longer time and would have had some experience with journal writing. At that time, standards for the choosing of pages should be set, perhaps collaboratively by the teacher and the children. Also, it would be helpful to have people other than just the teacher work with children

on their portfolios. Possibilities would include the principal, peers, older buddies, and parents.

2) Does entry length in experience-story production provide a useful indicator of growth in literacy? In the case of this project, it was felt that the amount of time needed to type 476 experience stories into the computer and then graph the word count was far too much, and that far too much effort was required, for the results obtained. However, some useful assessment information might be collected, particularly with older children, by using a scanner for putting work into the computer and, if quantity were an issue, by using the computer to graph. The technique might also be used by a trained speech specialist to record the stories of speech students. Work could be analyzed to ascertain the sounds being mispronounced and their positions in words. Results could be used to plan work for individual students, and progress over time could be documented when the mispronounced sounds no longer appeared. Recording experience stories for those children learning English as a second language could also constitute a viable assessment technique; the most noticeable progress in quality of writing in this class was shown through vocabulary growth in English by the one non-English speaker in the class.

In retrospect, it seems this project was ill-conceived if the goal was to assess progress in literacy acquisition. Wiggins (1991) is right when he says educators must be very clear, both with their students

and with themselves, about what it is they want. Teachers contemplating instituting alternative assessment techniques should consider carefully before they begin collecting data what it is they value enough to assess, and how they will know when they have it.

In the case of this project, part of the problem may have been not only with what was done, but with the way it was done. Collecting examples of children's writing on a regular basis (including dictated stories) does have the potential for documenting progress over time. Perhaps in kindergarten, collecting it regularly and reviewing it periodically is enough for a start.

3) Does working in groups provide a valid and reliable format for assessment of individual growth in mathematical problem solving? Groups do constitute a valid and reliable format for assessing individual growth in mathematical problem solving, particularly if group assessment is coupled with individual assessment and if group product presentations are videotaped. However, for optimal assessment possibilities, group projects should be developmentally appropriate and implemented at the proper grade level.

In the cases of the People Pieces and Squares Projects, it was evident that kindergarten children, most of whom are at the preoperational level, have an inordinate amount of trouble working in groups and thinking of themselves as members of groups. The

projects tried here would probably be more informative if tried in second grade or in an ungraded primary classroom.

4) Can assessment of nursery rhyme reconstruction by individual kindergarten children provide a teacher with useful data on the link between oral memory and reading and on how kindergarten children go about beginning to decode words? At its inception, the Nursery Rhyme project looked as if it might have potential as an assessment technique because it was in line with whole language principles, but, here again, the project proved much too time consuming and labor intensive for the results obtained. Most kindergarten teachers probably already know that children look only at the beginning consonant when they are first learning to read. If "to understand is to invent," then the project was worthwhile for convincing the researching teacher both of that and of the linearity of kindergarten thinking, but as a technique for assessing beginning reading, it fell short of the mark. Better ways on the kindergarten level would include becoming an unobtrusive kid watcher during Free Choice time and, for those kindergartners already reading, allowing children to read into a tape recorder and using miscue analysis techniques.

At the kindergarten level, particularly in a half-day class where time is at such a premium, perhaps the best reading assessment of all would be none at all. Perhaps the teacher should concentrate on reading aloud to the children from a wide variety of

genres and giving them plenty of oral language input through the use of Big Books, songs, chants, and pocket chart work and let the formal assessment of reading wait until first grade.

The researching teacher found it thought-provoking that a fluent kindergarten reader did not realize that written thes are interchangeable. What, if anything, it means, however, is a subject for someone else's study.

5) Does the quality and usefulness of the information gained in alternative assessment situations justify the time it takes to gather and analyze the data? Whether the quality and usefulness of the information gained in alternative assessment situations justify the time it takes to gather and analyze the data depends, probably, on what kind of data is gathered for what purpose, on how long it takes, on how many people are available to do it, on the classroom situation, on the alternative assessment technique, on the appropriateness of the assessment situation for the group of children involved, and on how many different alternative assessment techniques are attempted at once.

In the case of the projects attempted here, the People Pieces Project and the Squares Project were interesting and involving for both the teacher and the children. The teacher learned a great deal about how much trouble kindergarteners have working in groups and was able to document the growth of learning, audience behavior, and presenter behavior over time. The researching teacher also

made a start in trying out alternative assessment techniques. She will be able to build on those techniques in the future.

During the portfolio assessment phase of the Journal Project, the teacher learned that the interests and concerns of children could be ascertained through the use of a portfolio system; she also learned about the need for establishing standards for the choosing of portfolio entries with children.

During the Nursery Rhyme project the linearity of preoperational thinking was confirmed and the researching teacher was touched by how hard children as young as kindergarten age will persevere in an attempt to do well on a task.

Overall, however, as alternative assessment techniques, these projects were mostly off the mark for kindergarten. They were also too unwieldy, frustrating, time-consuming, labor-intensive, and expensive to be worth the effort and time they took.

Having said that, however, the researching teacher does not want to leave the impression that alternative assessment is a waste of time and should not be considered. Coming to such a conclusion would, in her opinion, be analogous to the conclusion reached by the preoperational child who does not conserve length because he or she looks at only one end of the stick.

In looking at alternative assessment, it would appear that the educational community is, for the most part, looking not only at just "one end of the stick," but at the wrong end of the stick. Alternative assessment is, in some instances, being imposed from the top down

without consideration of the conditions necessary to make it viable. Perhaps alternative assessment appears too time-consuming because there is not enough time in a teacher's day given the enormity of his or her task. Perhaps it appears too labor intensive because there are too few people for the amount of labor involved. Perhaps it appears too expensive because schools do not have the requisite equipment; because funds are limited; and because budgets are juggled, available monies spent on the wrong "priorities," and the budgeting process convoluted.

It may not be that "standardized testing [works], at least to an extent, for teachers faced with large student/teacher ratios and/or multi-subject teaching situations," but rather that standardized testing does not work well for either learners or teachers and that student/teacher ratios are too high and multi-subject teaching positions unworkable. If implementing alternative assessment techniques in a "small", basically monolingual class is as difficult as this research would indicate, how much more so must it be in the overcrowded multi-lingual classroom that is now the California norm?

Suggestions for Improving Kindergarten Assessment

Before educators can improve the assessment of any children, including kindergartners, communities need to come to agreement on what they value in the way of education. Then they need to decide what conditions must obtain in order to insure that all children are

given high-quality tasks to do in all those curricular areas that the community values. After that, they must provide those conditions. Only then will alternative assessment become a possibility--not as something superimposed on the schools from without, and not as a boot-straps exercise by individual teachers; but as an integral part of the education of children, conducted to see whether children truly are making progress in their efforts toward replicating tasks engaged in by competent adults in those areas that the community values.

Goodman (1992, p. 36) has said, "Perhaps those involved in developing alternative, authentic tests will do us all a good service by honestly reporting that they have taken on an impossible task." The task is impossible because the task of the classroom teacher has itself become impossible. Teachers are trying to do too much, in too little time, with too little support. Too many teachers lead lives of quiet desperation. Too many others work impossibly long hours under trying conditions in an attempt to get the job done--until they burn out. Like the proverbial caterpillar, they "lie distracted in a ditch, wondering which leg comes after which." And too many are unwilling spectators at the Emperor's parade, knowing full well that the Emperor wears no clothes (or at least is scantily clad) but are afraid to say so lest they appear stupid or unfit for their jobs.

The projects attempted here were time-consuming, labor-intensive, expensive, and often frustrating. Yet the researching teacher was left with the feeling that authentic assessment is a far better route for students and teachers than either traditional

classroom assessment or standardized testing is; the sooner the educational community gets its assessment practices into line with what benefits learning, the better. In a kindergarten classroom, that means implementing all the recommendations of the School Readiness Task Force of 1988. There should be an appropriate, integrated, experiential curriculum that meets the needs of all children, including the exceptional and the linguistically and culturally diverse. Class size should be reduced. Early primary staff should receive education and training, parent involvement should be encouraged, and funding and support should be made available for early childhood programs. Facilities should be built or remodeled to meet the needs of early primary programs, the public should be made aware of what constitutes appropriate learning practices for children aged four to six, and full-day kindergarten programs linked to child care should be an option. And last, assessment practices should be drastically altered. Assessment appears last on the list of task force recommendations; it should be last on our list too.

The authors of Here They Come, Ready or Not recommend that teachers use extensive individualized assessments that rely heavily of teacher observation and parental input. Extensive individualized assessments that rely heavily on teacher observation will not be possible until primary teachers have enough adult support in the classroom to allow them to be good kid watchers. It is neither safe nor practicable to observe one child carefully enough to assess while being simultaneously responsible for twenty or more others.

Additionally, elementary teachers need to be provided with the training both to be good at what they do and to allow them to embrace the alternative assessment movement. Both pre-service and in-service training are in need of revision. Too much pre-service training is anachronistic and far removed from the exigencies of the classroom situation. Too much in-service training does not meet the needs of teachers.

Wiggins calls authentic assessment "The Trojan Horse of school reform." It is time we let down our defenses and brought the horse within the gates.

References

- Allis, S. (1991). Testing, testing, testing. *Time*, July 15, 62-63.
- Ames, C. (1984). Competitive, cooperative, and individualistic goal structures: A cognitive-motivational analysis. In R.E. Ames & C. Ames (Eds.). *Research on motivation in education: Vol 1. Student motivation*. New York: Academic Press.
- Barone, T. (1991). Assessment as theatre: Staging an exposition. *Educational Leadership*, February, 57-59.
- Beck, M. [with L. Wilson in New York & P. Wingert in Washington]. (1991). New York meets Lake Wobegon. *Newsweek*, July 8, 48-49.
- Bock, R.D., & Wood, R. (1971). Test theory. *Annual Review of Psychology*, 22, 193-224.
- Boyle, J. D. (1989). Perspective on evaluation. *Music Educators Journal*, December, 22-25.
- Buckley, R. (1987). A funny thing happened on the way to "reading readiness": A teacher learns from the learners. *Language Arts*, 64 (7), November, 743-746.
- Bussis, A.M. & Chittenden, E.A. (1987). Research currents: What the reading tests neglect. *Language Arts*, 64(3), March, 302-308.
- Butler, O. (1985). Why Johnny can't get a job. *Fortune*, October, 163-168.
-

- Campbell, J. (1992). Laser disk portfolios: Total child assessment. *Educational Leadership*, May, 69-70.
- Cannella, G.S. & Reiff, J.C. (1989). Mandating early childhood entrance/retention assessment: Practices in the United States. *Child Study Journal*, 19(2), 83-99.
- Corno, L., & Mandinach, E.B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, 18, 88-108.
- Corno, L., & Rohrkemper, M.M. (1985). The intrinsic motivation to learn in classrooms. In C. Ames & R. Ames (Eds.), *Research on motivation in education: Vol. 2. The classroom milieu*. New York: Academic Press.
- Cross, D.R. & Paris, S.G. (1987). Assessment of reading comprehension: Matching test purposes and test properties. *Educational Psychologist*, 22(3 & 4), 313-332.
- Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, Winter, 58(4), 438-481.
- Deci, E.L. (1975). *Intrinsic motivation and self-determination in human behavior*. New York: Irvington.
- Dentzer, S. (1991). The vanishing dream. *U.S. News & World Report*, April 22, 39-43.
- Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist*, 34, 391-401.
-

- Developmentally Appropriate Practice in Early Childhood Programs Serving Children From Birth Through Age 8.* (1987). S. Bredekamp (Ed.). Washington, DC: National Association for the Education of Young Children.
- Eisner, E.W. (1991). What really counts in schools. *Educational Leadership*, February, 10-11 & 14-17.
- Elkind, David. (1989). Developmentally appropriate practice: Philosophical and practical applications. *Phi Delta Kappan*, October, 113-177.
- Ellwein, M.C., Glass, G.V., & Smith, M.L. (1988). Standards of competence: Propositions on the nature of testing reforms. *Educational Researcher*, November, 4-9.
- Elrod, G.F., & Sorgenfrei, T. (1988). Toward an appropriate assessment model for adolescents who are mildly handicapped. *Career Development for Exceptional Individuals*, 11(2), Fall, 93-99.
- Fitzpatrick, K. (1991). Restructuring to achieve outcomes of significance for all students. *Educational Leadership*, May, 18-22.
- Fredericksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39(3), 193-202.
- Freedman, R.L.H. (1994). *Open-ended questioning; A handbook for educators* [Assessment Bookshelf series]. Menlo Park, California: Addison-Wesley.
-

- Friedman, H. (1987). Repeat examinations in introductory statistics. *Teaching of Psychology*, 14, 20-23.
- Gardner, H. (1983). *Frames of mind*. United States of America: Basic Books, a division of Harper Collins.
- Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school. *Educational Researcher*, 18(8), November, 4-10.
- Gibboney, R.A. (1989). The unscientific character of educational research. *Phi Delta Kappan*, November, 225-227.
- Gibboney, R.A. (1991). The killing field of reform. *Phi Delta Kappan*, May, 682-688.
- Glickman, C.D., & Pellegrini, A.D. (1989). The case against kindergarten testing for promotion. *Principal*, November, 34-35.
- Goodman, K.S., Bird, L.B., & Goodman, Y.M. (1992). The whole language catalog supplement on authentic assessment. USA: SRA division of Macmillan/McGraw-Hill.
- Here they come: Ready or not!* (1988). Report of the School Readiness Task Force, California State Department of Education, Bill Honig, Superintendent of Public Instruction.
- Hill, K.T. (1984). Debilitating motivation and testing: A major educational problem--Possible solutions and policy applications. In R.E. Ames & C. Ames (Eds.), *Research on motivation in education: Vol 1. Student motivation*. New York: Academic Press.
- Hill, K.T., & Wigfield, A. (1984). Test anxiety: A major
-

- educational problem and what can be done about it. *Elementary School Journal*, 85, 105-126.
- Hills, J.R. (1991). Apathy concerning grading and testing. *Phi Delta Kappan*, March, 540-545.
- Hoge, R.D. (1988). Issues in the definition and measurement of the giftedness construct. *Educational Researcher*, October, 12-16.
- Johnson, D.W., & Johnson, R.T. (1985). Motivational processes in cooperative, competitive, and individualistic learning situations. In D. Ames & R. Ames (Eds.). *Research on motivation in education. Vol 2. The classroom milieu*. New York: Academic Press.
- Johnson, D.W., Johnson, R.T., & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and a meta-analysis of the research. *Review of Educational Research*, 53, 5-54.
- Johnson, D.W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta analysis. *Psychological Bulletin*, 89, 47-62.
- Johnson, T. (1990). Alternative Assessment Workshop, Las Lomitas School, Redwood City, California, October 20.
- Jongsma, K.S. (1989). Portfolio assessment. [Response from Jay Simmons] *The Reading Teacher*, 43(3), December, 264-265.
-

- Krechevsky, M. (1991). Project Spectrum: An innovative assessment alternative. *Educational Leadership*, February, 43-48.
- LeMahieu, P.G., & Wallace, R.C., Jr. (1986). Up against the wall: Psychometrics meets praxis. *Educational measurement: Issues and Practice*, 2(1), 12-16.
- Lichota, A. (1981). Gains expected in skills tests. *New York Times*, June 28, 13.
- Linn, R.L. (1988). State-by-state comparisons of achievement: suggestions for enhancing validity [Perspective]. *Educational Researcher*, April, 6-9.
- Lumsden, J. (1976). Test theory. *Annual Review of Psychology*, 27, 251-280.
- Maehr, M.L., & Stallings, W.M. (1972). Freedom from external evaluation. *Child Development*, 43, 177-185.
- Maier, L. (1991). In an alien land. *Phi Delta Kappan*, March, 555-556.
- Marton, F., & Säljö, R. (1976a). On qualitative differences in learning:1. Outcome and process. *British Journal of Educational Psychology*, 46, 4-11.
- Marton, F., & Säljö, R. (1976b). On qualitative differences in learning:2. Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology*. 46, 115-127.
- Marzano, R.J. & Costa, A.L. (1988). Question: Do standardized
-

- tests measure general cognitive skills? Answer: No.
Educational Leadership, 45(8), May, 66-71.
- Matthews, R.M., Whang, P.L., & Faucett, F.B. (1980). Behavioral assessment of occupational skills of learning disabled adolescents (Research Rep. No. 5), Lawrence: University of Kansas Institute for Research in Learning Disabilities.
- McCall, R.B. (1977). Childhood I.Q.'s as predictors of adult educational and occupational status. *Science*, July, 482-483.
- Miller, G.E. (1978). "Teaching and learning in medical school" revisited. *Medical Education* 12, Supplement, 120-125.
- NAEYC Position statement on standardized testing of young children 3 through 8 years of age [adopted November 1987] (1988). *Young Children*, March, 42-47.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22(2), 155 -175.
- O'Neil, J. (1991). Drive for national standards picking up steam. *Educational Leadership*, February, 4-8.
- Popham, W.H. (1983). Measurement as an instructional catalyst. In R.B. Ekstrom (Ed.), *New directions for testing and measurement: Measurement, technology, and individuality in education*, No. 17 (pp. 19-30). San Francisco: Jossey-Bass.
- Samway, K. (1987). Formal evaluation of children's writing: an incomplete story. *Language Arts*, 64(3), March, 289-298.
-

- Schell, L.M. (1988). Dilemmas in assessing reading comprehension. *The Reading Teacher*, October, 12-16.
- Schunk, D. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist*, 19, 48-58.
- Schunk, D. (1985). Self-efficacy and classroom learning. *Psychology in the Schools*, 22, 208-223.
- Searle, D. & Stevenson, M. (1987). An alternative assessment program in language arts. *Language Arts*, 64(3), March, 278-284.
- Shapiro, W. (1991). Tough choice. *Time*, September 16, 54-60.
- Sternberg, R.J. (1990). Thinking styles: Keys to understanding student performance. *Phi Delta Kappan*, January, 366-371.
- Stiggins, R.J. (1991). Assessment literacy. *Phi Delta Kappan*, March, 534-539.
- Stiggins, R.J., Griswold, M., Green, K.R., & associates (1988, April). Measuring thinking skills through classroom assessment. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.
- Teale, W.H., Hiebert, E.H., & Chittenden, E.A. (1987). Assessing young children's literacy development. *The Reading Teacher*, April, 772-777.
- Valencia, S. (Speaker). (1991, January 19). *Literacy: Assessment for the 90s*. Santa Clara, CA: Performance Based Assessment: A Bridge to Change [conference for teachers and administrators].
- Wiggins, Grant (Speaker). (1991, January 19). *Teaching to the*
-

(authentic) test: Assessment that measures and evokes quality student performance. Santa Clara, CA: Performance Based Assessment: A Bridge to Change [conference for teachers and administrators].

Wiggins, G. (1993). *Assessing student performance.* San Francisco: Jossey-Bass.

Wixson, K.K. & Peters, C.W. (1987). Comprehension assessment: Implementing an interactive view of reading. *Educational Psychologist*, 22(3 & 4), 333-356.

Appendix A

Ages of Children in Class from Oldest to Youngest

(Rounded to Nearest Month)

F preceding a number designates a female student; M
preceding a number designates a male.

<u>Rank</u>	<u>Student</u>	<u>Birthdate</u>	<u>School Year</u>	<u>Last 3 Months</u> (Mar. 14 - June 14)
1	F-15	6/12/84	6-3 to 7 yrs.	6-9 to 7 yrs.
2	M-11	8/4/84	6-1 to 6-10	6-7 to 6-10
3	F-18	11/7/84	5-10 to 6-7	6-4 to 6-7
4	F-20	11/18/84	5-10 to 6-7	6-4 to 6-7
5	M-13	12/2/84	5-9 to 6-6	6-4 to 6-6
6	M-10	12/4/84	5-9 to 6-6	6-3 to 6-6
7	F-3	12/7/84	5-9 to 6-6	6-3 to 6-6
8	F-23	12/19/84	5-9 to 6-6	6-3 to 6-6
9	F-7	12/31/84	5-8 to 6-5	6-2 to 6-5
10	M-5	2/21/85	5-6 to 6-4	6-1 to 6-4
11	F-16	2/24/85	5-6 to 6-4	6-1 to 6-4

Appendix A (continued)

Ages of Children in Class from Oldest to Youngest

(Rounded to Nearest Month)

F preceding a number designates a female student; M
preceding a number designates a male.

<u>Rank</u>	<u>Student</u>	<u>Birthdate</u>	<u>School Year</u>	<u>Last 3 Months</u> (Mar. 14 - June 14)
12	F-14	3/6/85	5-6 to 6-3	6-0 to 6-3
13	F-1	3/6/85	5-6 to 6-3	6-0 to 6-3
14	F-4	3/19/85	5-6 to 6-3	6-0 to 6-3
15	M-19	3/24/85	5-5 to 6-3	6-0 to 6-3
16	M-9	3/30/85	5-5 to 6-2	5-11 to 6-2
17	F-8	4/16/85	5-5 to 6-2	5-11 to 6-2
18	M-21	4/24/85	5-4 to 6-2	5-11 to 6-2
19	F-6	5/10/85	5-4 to 6-1	5-10 to 6-1
20	M-12	5/19/85	5-4 to 6-1	5-10 to 6-1
21	F-2	6/5/85	5-3 to 6-0	5-9 to 6-0
22	M-22	7/1/85	5-2 to 5-11	5-8 to 5-11
23	M-17	8/7/85	5-1 to 5-10	5-7 to 5-10

Appendix B

Ages of Children in the Class by Randomly Drawn Student Number

(Rounded to Nearest Month)

F preceding a number designates a female student; M preceding a number designates a male.

<u>Student</u>	<u>Rank</u>	<u>Birthdate</u>	<u>School Year</u>	<u>Last 3 Months</u>
F-1	13	3/6/85	5-6 to 6-3	6-0 to 6-3
F-2	21	6/5/85	5-3 to 6-0	5-9 to 6-0
F-3	7	12/7/84	5-9 to 6-6	6-3 to 6-6
F-4	14	3/19/85	5-6 to 6-3	6-0 to 6-3
M-5	10	2/21/85	5-6 to 6-4	6-1 to 6-4
F-6	19	5/10/85	5-4 to 6-1	5-10 to 6-1
F-7	9	12/31/84	5-8 to 6-5	6-2 to 6-5
F-8	17	4/16/85	5-5 to 6-2	5-11 to 6-2
M-9	16	3/30/85	5-5 to 6-2	5-11 to 6-2
M-10	6	12/4/84	5-9 to 6-6	6-3 to 6-6
M-11	2	8/4/84	6-1 to 6-10	6-7 to 6-10
M-12	20	5/19/85	5-4 to 6-1	5-10 to 6-1
M-13	5	12/2/84	5-9 to 6-6	6-4 to 6-6
F-14	12	3/6/85	5-6 to 6-3	6-0 to 6-3

Appendix B (continued)

Ages of Children in the Class by Randomly Drawn Student Number

(Rounded to Nearest Month)

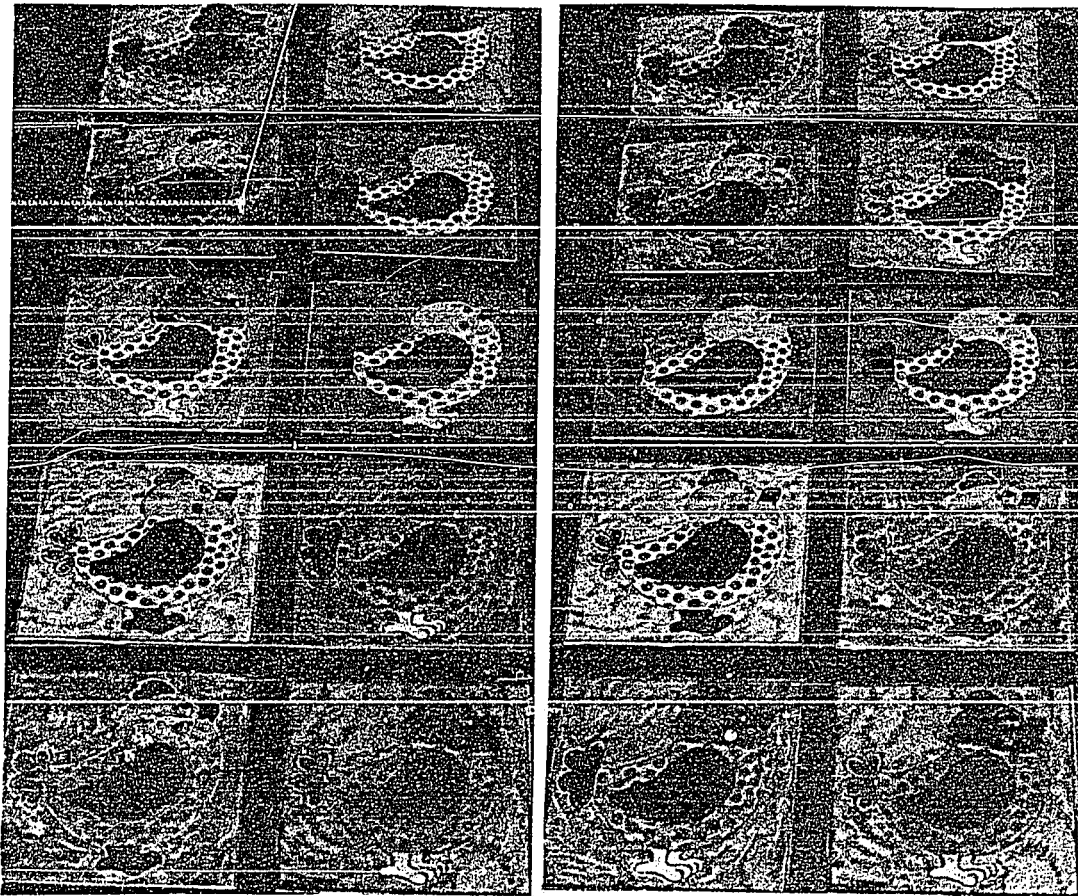
F preceding a number designates a female student; M preceding a number designates a male.

<u>Student</u>	<u>Rank</u>	<u>Birthdate</u>	<u>School Year</u>	<u>Last 3 Months</u>
F-15	1	6/12/84	6-3 to 7 yrs.	6-9 to 7 yrs.
F-16	11	2/24/85	5-6 to 6-4	6-1 to 6-4
M-17	23	8/7/85	5-1 to 5-10	5-7 to 5-10
F-18	3	11/7/84	5-10 to 6-7	6-4 to 6-7
M-19	15	3/24/85	5-5 to 6-3	6-0 to 6-3
F-20	4	11/18/84	5-10 to 6-7	6-4 to 6-7
M-21	18	4/24/85	5-4 to 6-2	5-11 to 6-2
M-22	22	7/1/85	5-2 to 6-0	5-9 to 6-0
F-23	8	12/19/84	5-8 to 6-6	6-3 to 6-6

Appendix C

Chicken Pieces

Below are examples of teacher-made Chicken Pieces created to give children practice in sorting between People Pieces products and Toad and Rabbit products.

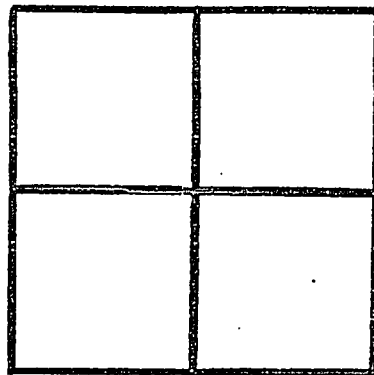
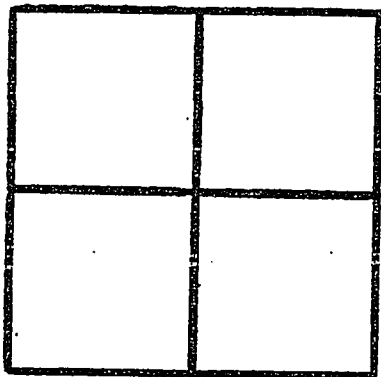
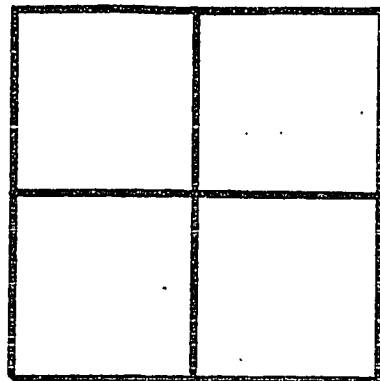
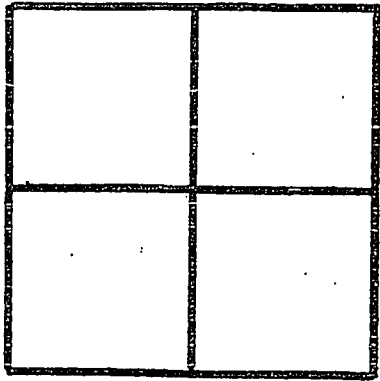


Source: Ditto material from
The Center for Early Learning
Laconia, NH 1974

Appendix D

Squares

Below is an example of squares used by children in the Squares project. On the sheet used by the children, twelve of these squares fit on an 8 1/2 by 11 inch piece of paper and six additional squares fit on the half sheet stapled to it.




June 26, 1994

To Whom It May Concern:

I give Joan Damm permission to use Very Important Person graphs created by me as part of her Master's thesis. If you have any questions regarding the creation of these graphs, you may reach me at 560 Sixth Street, Hollister, CA 95023.

Sincerely,



Jerry S. Damm