Research on a system of developing and monitoring progress on individualized education program (IEP) goals has been conducted over a 5-year period at Minnesota's Institute for Research on Learning Disabilities. This research and development project is summarized by outlining the goal and rationale of the research, presenting the overall research plan, and providing a chronological description of the research and development process and results, and the current status of the research. The research has proceeded in three stages: (1) determination of what measures of student performance are appropriate in a formative evaluation system; (2) examination of the practicality and efficiency of the monitoring system for teachers; and (3) determination of the effects of formative evaluation on student achievement. Results have indicated that generic measures useful in indexing proficiency in reading, spelling, and written expression are available and that teachers who have used the measures and the monitoring system are pleased with it and do not feel unduly burdened by the time commitment. Preliminary evidence concerned with the effect of the monitoring system on student academic progress is positive. (Author/DB)
Director: James E. Ysseldyke
Associate Director: Phyllis K. Mirkin

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The material presented herein was prepared under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of the Office of Special Education.
Monograph No. 18

RESEARCH ON DEVELOPING AND MONITORING PROGRESS ON IEP GOALS:
CURRENT FINDINGS AND IMPLICATIONS FOR PRACTICE

Caren Wesson, Stanley Deno, and Phyllis Mirkin
Institute for Research on Learning Disabilities
University of Minnesota

April, 1982
Abstract

Research on developing and monitoring progress on IEP goals has been conducted over a five-year period at Minnesota's Institute for Research on Learning Disabilities. This research and development project is summarized by outlining the goal and rationale of the research, presenting the overall research plan, and providing a chronological description of the research and development process and results, and the current status of the research. Suggestions for classroom applications of the findings also are presented.
Research on Developing and Monitoring Progress on IEP Goals: Current Findings and Implications for Practice

Educators always have had to make decisions relating to instruction for students. However, in recent years greater demands have been placed on educators, especially special educators, to be accountable for the quality of these decisions and the ways in which they are made. A number of criteria to be followed in assessment and decision-making procedures have been outlined in PL 94-142, and while schools have attempted to follow the guidelines that accompany this law, the technical knowledge to fully comply with the intent of the law as well as the procedures of the law are, for the most part, not available.

During the past five years, the Institute for Research on Learning Disabilities at the University of Minnesota, under federal contract, has conducted a number of studies that focus on developing and monitoring progress on IEP goals, as is intended in PL 94-142.

The purpose of this paper is to describe a five year research and development project focusing on formative evaluation. Rather than describing the studies in detail, this paper will outline the goal and rationale of the research, present the overall research plan, and provide a chronological description of the research and development project including the process and results, and the current status of the research. In addition, suggestions for up-to-date classroom applications of the research findings will be provided throughout.
**Goal and Rationale**

The goal of the research and development program was to determine empirically the effects of using formative evaluation techniques on student achievement in reading, spelling, and written expression. The focus was on the IEP adjustment decision in the special education process. The formative evaluation system was to be an assessment device for monitoring the effectiveness of the IEP. (See Figure 1.) The hypothesis was that if an adequate system of formative evaluation was developed, teachers could use this system to monitor student progress and the effectiveness of their instruction. If student progress is not adequate, then teachers judge their instruction to be ineffective and make a modification in their instruction in an attempt to improve the student's progress.

Insert Figure 1 about here

The rationale underlying this hypothesis is as follows. The success of special education is defined by the extent to which students' academic and social behaviors are improved. For the individual mildly or moderately handicapped student, it is impossible to reliably identify special educational alternatives that will be more effective than the regular classroom program. Therefore, the IEP is a guess about what might be helpful to the student rather than a plan that is guaranteed to help. Given that the IEP is only a guess, there is no alternative but to continuously evaluate the effectiveness of the IEP and to modify it when it is not working. Teachers can
increase the success of special education by systematically measuring student progress toward the achievement of program goals and then adjusting student programs to improve that progress. Student performance data are the most useful "vital signs" of whether the program is working or should be changed. A good formative evaluation system allows teachers to test their best hunches about how to help a student.

The formative evaluation system must be useful for monitoring any type of instruction. For example, whether the teacher chooses DISTAR, Orton-Gillingham, or any other reading approach to teach reading, the monitoring system should accurately measure the student's progress in reading. The monitoring system must be unbiased in regards to various theoretical approaches to teaching.

Research Plan

In order to accomplish the goal of the research and development program, a three-stage plan was designed (Deno, 1979). Stage One included: (a) the identification of technically adequate measures of student performance and the development of these measurement procedures; (b) the investigation of a variety of frequencies and durations of measurement; and (c) an examination of the effects of systematic techniques for using the data generated by these measures to make decisions about the effectiveness of instruction. The studies in Stage One were intended to lay a foundation for subsequent studies. Identifying useful measures of student performance was critical since the results of later investigations would be meaningless if student performance data were not technically adequate.
Stage Two was an examination of the logistical feasibility of the formative evaluation system, as measured by teacher efficiency and satisfaction. No system of formative evaluation would be useful if teachers found it to be too time consuming or were dissatisfied with other aspects of the system. Without efficiency and teacher acceptance, the formative evaluation system probably would not be used regardless of how good the system was for measuring student performance.

Stage Three was designed to be a test of the original goal, that of determining the effects of using formative evaluation on student performance. Other effects of formative evaluation also are being considered in Stage Three. In particular, current studies are examining the degree of structure in a student's educational plan as a function of formative evaluation.

A Chronological Description of the Research

The research conducted in each of the three stages will be outlined before addressing some procedures for applying the results of these studies to the classroom.

Stage One: How to Measure Student Performance

The procedure for determining student performance measures. The first step in Stage One was to determine what measures of student performance would be ideal for use in a formative evaluation system. The search for these measures began by generating a list of desired characteristics.

To be considered for inclusion in a formative evaluation system, the developed measures had to fulfill the following criteria:
(1) They must be valid with respect to widely used measures of achievement in reading, spelling, or written expression.

(2) They must be immediately sensitive to the effects of relatively small adjustments made in (a) instructional methods and materials, (b) motivational techniques, and (c) administrative arrangements (e.g., adjustments in grouping, setting for instruction, teacher/tutor, time of instruction, etc.).

(3) They must be easy to administer by teachers, parents, and students.

(4) They must include many parallel forms that are frequently administrable (daily, if necessary) to the same student.

(5) They must be time efficient.

(6) They must be inexpensive to produce.

(7) They must be unobtrusive with respect to routine instruction.

(8) They must be simple to teach to teachers, parents, and children.

The basic strategy used to identify measures with these desired characteristics involved a process of elimination. Initially, a pool of commonly measured behaviors was generated through a review of the available literature in reading, spelling, and written expression. The next step was to develop simple standardized measurement procedures. The third step was to determine the criterion validity of the measurement procedures by correlating the scores obtained from them with scores on commercially available standardized measures, with program placement, and with grade level. The measures that were not reliable or valid, or those that were deemed lower with respect to any of the other desired characteristics, were eliminated from the pool.
For example, several measures that were relatively insensitive to growth were dropped from consideration.

**Identifying the reading measure.** Some of the studies conducted for the reading measures will be described in order to illustrate the process for determining student performance measures. Five reading behaviors were generated from a review of the literature and placed in the original pool for consideration:

1. **Reading isolated word lists** consisted of three alternative forms of 60 words each that were randomly selected by grade level from the core list of 5,167 words listed in *Basic Elementary Reading Vocabulary* (Harris & Jacobson, 1972).

2. **Reading isolated words in context** measure consisted of three passages of approximately 600 words selected from the beginning, the middle, and the latter parts of the sixth grade book for three different basal reading series: Allyn-Bacon, Ginn 720, and Houghton-Mifflin. These were typed with every fifth word in each passage underlined.

3. **Reading aloud from text** included three additional passages of 300 words each. These were selected from the same basal readers as the words in context measure and typed on a sheet of paper. Each passage consisted of the first part of the story.

4. **Identifying deleted words in text (Cloze)** was developed from three additional passages of 300 words each that were selected from the same basal readers. The 5th word was deleted from all the other sentences in the passage.

5. **Giving word meanings** involved the use of three passages consisting of 300 words each that were selected from the same basal readers. Every fifth word of the passage was underlined, unless it was a function word, and students were required to say their meanings.

Given these measures, a series of criterion validity studies was undertaken. For example, one study correlated performance on these five reading measures with performance on the Stanford Diagnostic
Reading Test and the Woodcock Reading Mastery Test for 33 learning disabled and non-learning disabled students (Deno, Mirkin, Chiang, & Lowry, 1980). The results indicated that all correlations were statistically significant, ranging from .60 to .91. Additional concurrent validity studies using other samples and other standardized measures followed this investigation. Finally, scores on these measures were used to discriminate between students in LD programs and regular class programs, and among students at different grade levels. The conclusion was that reading aloud from a basal reader, reading aloud from lists of isolated words, and guessing at what words had been deleted from a reading passage (i.e., cloze comprehension) all related closely to performance on standardized tests and discriminated between program and grade placement. The word meaning measure was eliminated from consideration due to its relatively lower correlations with standardized test scores. The cloze measure and the isolated words measure also were dropped from consideration because of the time required to use the procedures in the classroom; teachers preparing to use these measures would have to delete or underline words from passages, a time consuming procedure. In contrast, minimal preparation would be required of teachers if isolated word lists or reading passages were used for formative evaluation.

Related studies conducted during this time focused on determining the optimal duration of measurement and the type of data to record. Results from testing one, two, and three-minute durations indicated that reading proficiency could be indexed validly within one minute. Also, correct performance was a more valid measure of reading
proficiency than error performance; correct performance alone was found to discriminate among reading proficiencies as well as a combination of correct and incorrect performance.

Studies designed to investigate sensitivity of the measures to change over time also were conducted during Stage One (Marston, Lowry, Deno, & Mirkin, 1981). In order to be useful in a formative evaluation system, the measures must be able to detect relatively small changes in performance. Therefore, a study was designed to assess the sensitivity of two reading measures, reading isolated word lists and reading aloud from a basal reader. Sensitivity was assessed in two ways: growth across grade levels, and growth within each grade level from fall to spring. Both reading measures were found to be sensitive to changes. However, reading aloud from a basal reader was chosen as the optimal generic measure in reading because it has a broader range of scores than isolated words, relates somewhat more closely to comprehension, and requires little teacher preparation. The teacher simply picks a random passage and directs the child to read aloud.

**Identifying measures in spelling and written expression.** Similar research plans were followed in the areas of spelling and written expression. This research led to the conclusion that the optimal measure in spelling is the number of words or letter sequences written from dictated lists that have been repeatedly sampled from a constant list such as the same basal reader (Deno, Mirkin, Lowry, & Kuehnle, 1980), and that the optimal measure in written expression is the number of words or letter sequences written in response to a verbal or
pictorial stimulus (Deno, Mirkin, & Marston, 1980). Thus, in the academic areas of reading, spelling, and written expression, there are measures that can be used repeatedly and frequently to monitor student progress. These measures have the desired characteristics originally identified as necessary for use in a formative evaluation system. Further, although the research will not be discussed here, the measures also are useful for making other decisions in special education, such as in referral, identification, IEP selection, and IEP certification (see Figure 1).

Writing objectives. Given the identified measures, the next focus of Stage One in the research program was to investigate two procedures for writing objectives. Ten resource room teachers in a rural cooperative special education district participated in a year-long project during which they used the generic measures for developing IEP long-term goals and short-term objectives as well as for monitoring student progress.

In writing long-term goals, teachers used generic measures from various levels of material in order to find the level at which the student met entry level criteria. The goal set for this level was written using a basic formula. Figure 2 depicts the format for goals in the academic areas.

--------------------------
Insert Figure 2 about here
--------------------------

Short-term objectives (STOs) are based on the long-range goals. Two types of STOs are possible. Briefly, STOs can be written so that
measurement is on a standard task (e.g., reading aloud at a specific level of a reading series) or measurement can be based on a standard criterion applied to sequential tasks (e.g., mastery of units in a basal reader). A year-end survey revealed that although teachers preferred measurement on a standard task in most cases, for reading many preferred measuring progress through sequential tasks (Fuchs, Wesson, Tindal, Mirkin, & Deno, 1982).

Using the student performance data. The next step for these teachers was learning to use the student performance data. Prior to learning two data utilization strategies, teachers first were taught to specify each student's instructional plan, including activities, time, arrangement, materials, and motivational strategies used in each IEP area. Once the instructional plan was elaborated, teachers began to use a data utilization strategy. Each strategy was put into practice for approximately eight weeks. Therapeutic teaching involved using an aimline, which marked the line of progress required in order for the student to accomplish the IEP goal. The rule for using the data was that if the student's performance fell below the aimline on three consecutive days, then a change in the instructional plan was necessary. The other strategy, experimental teaching, required use of standard analyses of time series data. Specifically, teachers used the slope, step up and down, variability, overlap, and medians to determine the effectiveness of instruction (c.f. Jones, Vaught, & Weinrott, 1977).

Analyses of student performance data indicated that students showed more academic growth when a data utilization strategy was in
effect than when teachers did not use the data systematically. Questionnaires designed to evaluate teacher satisfaction with these strategies revealed that teachers preferred to use a combination of the two strategies over using either strategy alone. This finding contributed to the design of the data utilization strategy employed in Stage Three studies. The new strategy combined use of the aimline and calculation of the actual slope of the student's performance data. Performance lines less steep than the aimline have become the cue for changing the instructional plan.

After identifying the optimal generic measures, developing procedures for writing IEP goals and objectives, and determining a viable strategy for using the data, the next move was to examine the logistical feasibility of this system; this was the focus of Stage Two research.

**Stage Two: Is This Monitoring System Practical for Teachers to Use?**

**Efficiency of the monitoring system.** How much time does it take to measure student academic performance using these procedures? Teachers who were trained to time their measurement activities required 3.5 minutes per academic task at the beginning of the school year. By the end of the year, this time was reduced to approximately 2 minutes per task. These teachers measured each student in their caseload three times per week; they did not feel that measurement was too time consuming. During the school year, teachers systematically attempted to reduce their measurement time and continuously were prompted to improve their efficiency. They were, indeed, more efficient by the end of the year (Fuchs, Wesson, Tindal, Mirkin, &
Deno, 1981). Interestingly, other teachers who were not prompted to improve their efficiency or to systematically alter their measurement techniques became less efficient over the school year (Wesson, Mirkin, & Deno, 1982). The success of the prompted teachers was encouraging evidence that the system provided adequate training and support.

**Teacher satisfaction with the monitoring system.** The second piece of evidence of a logistically feasible system is the feedback from the teachers. Are they satisfied with the system?

Independent evaluators questioned the teachers about this monitoring system. Special education staffs expressed that:

1. The system eliminated much of the jargon, ambiguity, and vague descriptions once found in IEPs.

2. They felt more confident that the system meets the real intent of the law.

3. They felt their own testing is now relevant to the instruction being provided in the classroom. ("We've changed our objectives from improving test scores to improving the student's performance in the classroom.")

4. They felt confident in the reliability of their test, making decisions easier and meetings shorter.

5. Their testing was more meaningful because a student is compared with peers from his/her own school and grade level.

6. They felt students were more aware of their own progress because of the frequent charting required by the data based system. This charting also increased the motivation of teachers and students toward reaching goals and objectives.

7. Their ability to measure the effectiveness of their teaching strategies with any particular student was improved. The system notifies a teacher when to change his/her current intervention.

8. The system made writing IEPs much easier.
They had a positive feeling of confidence because the system meant current information always was available on any special education student's progress.

These comments clearly suggest that this monitoring system not only is logistically feasible, but, in fact, has practical advantages.

**Current Research**

**Stage Three: Does Formative Evaluation Increase Student Achievement?**

Stage Three of this research and development plan brings the focus of research back to the original goal: to determine the effects of formative evaluation on student achievement. Currently, two types of Stage Three studies are in progress in rural, suburban, and urban sites.

One kind of study is a matched pair design in which two students who are similar in age, grade, and reading ability are assigned either to a control group or to a monitored group. The hypothesis is that students in the group in which progress is monitored will make greater progress than students in the unmonitored group. The second type of study, called the causal model, uses correlational techniques to determine the relationships among the degree of implementation of the formative evaluation system, the amount of structure in the student's instructional program, and the student's rate of academic progress. (See Figure 3.) The hypothesis here is that the extent to which teachers implement this monitoring system influences the degree to which their teaching is structured, which in turn influences the extent to which students share academic progress. All studies will be completed by the end of 1981-82 school year.
Summary

Four years of research in Stage One of the research plan served to lay the foundation for work in Stages Two and Three. Now, generic measures useful in indexing proficiency in reading, spelling, and written expression are available. These measures meet all the criteria necessary for inclusion in a formative evaluation system. These measures also are useful in developing IEP long-term goals and short-term objectives. In addition, a viable strategy for using the student performance data generated by these measures also is available. Furthermore, those teachers who have used these measures and this monitoring system are pleased with it and do not feel unduly burdened by the time commitment. Now, the crucial test is at hand. Can we teach teachers to use this monitoring system to increase the student's rate of academic progress? Preliminary evidence is encouraging.
References


Figure 1. Special Education Decision Making Processes
<table>
<thead>
<tr>
<th>CONDITION</th>
<th>BEHAVIOR</th>
<th>CRITERIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading: In (total # weeks until next annual review) when presented with stories from Level (Level # at which you expect student to be proficient by time of annual review) in (name of reading series) for one minute</td>
<td>(student) will read aloud</td>
<td>for grades 1 and 2: 50 correct words with 5 errors or fewer; for grades 3-6: 70 correct words with 7 errors or fewer.</td>
</tr>
</tbody>
</table>

| Spelling: In (total # weeks until next annual review) when dictated words for 2 minutes from Level (same as reading level) in (name of series) | (student) will spell words | for grades 1 and 2: 40 letter sequences correct or 5 words correct; grades 3-6: 60 letter sequences correct or 3 words correct. |

| Written Expression: In (total # weeks until next annual review) when presented with a story starter or topic sentence and three minutes in which to write a story | (student) will write | a total of __________ words or letter sequences correct. (see Table 7-1) |

Figure 2. Format for Writing Long Range Goals
Figure 3. Causal Model Research Design Hypothesis
The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.


Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.


Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.


Epps, S., McGue, M., & Ysseldyke, J. E. Inter-judge agreement in classifying students as learning disabled (Research Report No. 51). February, 1981.

Epps, S., Ysseldyke, J. E., & McGue, M. Differentiating LD and non-LD students: "I know one when I see one" (Research Report No. 52). March, 1981.


Graden, J., Thurlow, M., & Ysseldyke, J. Instructional ecology and academic responding time for students at three levels of teacher-perceived behavioral competence (Research Report No. 73). April, 1982.
