CHAPTER TEN
EVALUATING

Evaluating is itself a learning activity inseparable from the language arts, because feedback is part of communication. But it must be done very thoughtfully. Much waste and much harm occur when evaluation is run off routinely without considering its exact function and the possibility of negative side effects.

Collecting periodic data on reading ability, say, as most schools and school districts do today, merely for "knowing" what is happening is a monstrous waste of time and money and often merely misleads the public.[1]

DIFFERENT FUNCTIONS

Language arts evaluation usually serves about five functions. It should indicate

- to the individual student, how fully he is developing in discourse.
- to the parent, how much the student is learning in school.
- to the teacher, the needs of the student, for coaching and advising.
- to the administrator, how good a job the teacher is doing.
- to the school board and community, how effectively the curriculum and materials reach their goals.

Too often educators expect a few standardized test scores to fulfill all five functions at once, and yet it's obvious that one such narrow type of evaluation cannot serve such different purposes. Students need to know if they're making themselves clear to others and understanding what others are saying or writing. Standardized tests will not tell them this or give them perceptions that they can use in the future. Of course students sometimes care about test scores or grades but mostly just because adults do. Otherwise, they don't need the blanket judgments of themselves, comparing one with others, except to the extent that the environment whips up a competitive atmosphere. Comparison satisfies a need many parents feel to know where their child stands in relation to his peers.

[1] Richard L. Venezky, *Testing in Reading* (Urbana, IL: National Council of Teachers of English, 1974), 5–6.

Grades, too, exist mainly for this comparison, since they provide only a blanket judgment based on some idea of a norm. Some tests, like those administered by the National Assessment of Educational Progress for each subject every four years, compare not individuals but whole state or district school systems. Such testing purports to aid curriculum development, but since comparison with other school districts is misleading and irrelevant, it really serves only political purposes and public relations.

If the teacher is assessed by the very same test scores that assess the student, this puts the teacher in an impossible conflict of interest. No one wants to turn in a bad report on himself; if job or salary depends directly on these test scores, teachers may teach to the test and precious little else, or simply cheat, or try to drive out of their class those students who score low, or compete to teach in schools with high test scores. This sort of evaluation explains some teacher hostility to disadvantaged children, some of the proliferation of separate classes for the so-called handicapped, undue pressure on counselors to classify some children as learning disabled, the commitment to tracking by ability, and large numbers of "force-outs" from school. Far from guaranteeing students good teaching, tying teachers' jobs to their students' test scores creates a me-or-them atmosphere hardly conducive to learning.

Standardized test scores do reassure (or unsettle) the administration and the community about the efficacy of various programs in the curriculum, and this is really what they are for. But to coach and guide individuals, teachers must have far more specific and extensive information about them than standardized tests can provide. These tests merely show where a student ranks against past and present peers on a few skills, not what he needs to do next or where he's coming from or the pattern of his whole verbal life. They are for people outside the classroom, not for those inside, who will have to find out what they need to know in daily ways as they go about their real business.

How is it possible to do justice to all five of the functions outlined above, indeed, without letting evaluation take over the classroom? And how may different parties be furnished the right sort of evaluation for their purpose without interfering with the sort the others need? The best answer may be to follow strictly two cardinal principles.

The first principle is: *Each party should do its own evaluating.* No one should be asked to evaluate himself *for another person.* Some teachers feel, for example, that letting students grade themselves is fair and "liberal," but actually it shifts unfairly the teacher's burden to the student and merely bypasses an important issue of internal versus external assessment. These teachers exclaim that students usually rate themselves too low rather than take advantage of the self-grading, but that's not a good sign, and the student should no more be put in the position of evaluating himself for another's purpose than the teacher should. If good reason exists for others to have an evaluation of students, then others should do it, availing themselves of the students' self-evaluation as part of their own. A dual inside-outside view makes for excellent assessment. But a student should evaluate his work only for his own reasons and by his own means.

Likewise, the administrator must size up a teacher's ability and effectiveness by more means than such evidence as the teacher collects and passes up to him. To prevent self-incrimination teachers can make this evidence self-serving when job security is at stake. Observations and talks with the teacher may be considered

along with the pattern of the teacher's activities and attitudes, with broad, consensual assessments of the teacher by parents and students, with student performance as measured by outsiders, and with student progress and performance as measured by the teacher.

Nor should the school or the system be judged by its officially promulgated results, which could cover up, intentionally or not, a low quality of education. Parents judge schools, in fact, not just by grades and scores but by what learning they can see for themselves when they observe and talk with their children. This is right and healthy.

The second cardinal principle is: *Evaluation should not dictate, distort, or displace what it measures.* It is difficult but essential to follow this rule, for to the extent that the institution breaks it, it defeats itself. Learning is the mission. The only goal of evaluation is to further learning. If evaluation ends by determining what is taught and how it is taught, by grossly or subtly turning learning from one thing into another not originally intended, it is bad evaluation. If it appropriates to itself the time and energy that could be used for more learning, it is bad evaluation. Most traditional assessments break this rule by shrinking the curriculum to fit their own instruments. Teaching to the test causes it to act backward and determine what it is only supposed to measure. The evaluation tail wags the curriculum dog.

Because standardized tests are less frequent, they hurt perhaps less than the daily and weekly tasks that are assigned only or mostly to get a grade off of students or a glimpse of what's going on in their heads—the quizzes, oral questioning, "reports," and so on. In this way, school writing has been too much just a testing instrument of the reading. And the "marking" of papers in the name of evaluation has made generations of students hate to write even so much as a personal letter later in life and probably accounts, more than any factors of intrinsic difficulty, for the poor writing ability of most high school graduates. Likewise, constant testing for reading comprehension by oral or written questions makes students feel punished for reading. So long as educators give reading such negative associations, it is pointless to rail about the abominations of television and to blame other extracurricular factors. Until schools divest reading and writing of the stunningly negative effects caused by breaking this second cardinal rule of evaluation, no one will ever know how well they can compete with other media. Incessant testing can virtually kill off the very two R's everyone is most worried about. Then as scores decline, schools frantically increase testing! The more score-keeping, the lower the score.

So we seem to face this dilemma: a lot of evaluation is needed, because a number of different parties and purposes must be served, and yet a lot of evaluation destroys the very learning it's supposed to facilitate. A narrowly programmed curriculum that teaches small things in small steps seems to solve the problem because of the claim that all items are taught and tested at virtually the same time, but the kinds of items that can be so taught and tested do not rise to a high enough level of mental organization to constitute significant education.

EVALUATING WITHOUT ACTIVITIES THAT ONLY TEST

A solution does exist to this dilemma. Evaluation can be done by means of valid learning activities themselves without making students do additional activities

only for the purpose of evaluation. The most efficient curriculum allows students to spend all their time learning without winding up in the position of accounting for something that is of little account. The dilemma is unreal. An environment that fosters authentic communication makes learning and evaluating compatible. The same passivity, paucity, and poverty of classroom dynamics that can make learning to read and write seem harder than they really are make evaluation seem like an inevitable parasite. The brute fact is that *ordinarily students don't produce enough to provide the evaluator something to see.* But if students are constantly producing and receiving discourse in great volume and variety, and if the teacher is freed from emceeing to circulate and observe, then good evaluation becomes possible without resorting to special activities that detract from learning and make students hate reading and writing.

To understand this point well, consider the difference between assessing receptive as opposed to productive activities. If a student says or writes or performs something, an observer can see or hear it and make a judgment about it. There is overt behavior or a tangible product. No need to make the student do something further to yield wherewithal for evaluating. When a student listens or reads or witnesses, however, there is ordinarily nothing to show for it. In order to turn the student's head inside out to look at his comprehension, the evaluator has to make him do an additional, unnecessary activity that produces something—traditionally, answering oral or written questions. The more the curriculum is oriented toward the receptive language arts, the more serious looms this problem of evaluation.

Now, listening, reading, and witnessing may be followed up by productive activities that while secondarily permitting the evaluator to see and hear a student's comprehension are, foremost, valid learning activities that students might do anyway for their own sake. Performing, discussing, and extending texts through writing externalize reading comprehension. A teacher or other evaluator witnessing performances or sitting in on rehearsals and other small-group discussions of common reading may not only note well the points of incomprehension but have a chance to hear incomprehension itself discussed in some detail. *Informal Classroom Drama, Performing Texts* and *Reading*, contain numerous productive activities that follow up reading and are valuable for their own sake. Translating texts into other media, such as illustrations or film, demonstrates comprehension also.

Among other virtues, oral and written directions are by nature meant to be carried out and hence naturally lend themselves to translating comprehension into visible action. Enacting words, in fact, is the chief way that truly scientific researchers—psycholinguists—employ to ascertain comprehension. It's of no small interest to us here that they do not measure comprehension the way schools do, by pencil-and-paper tests that translate words into other words; they go from verbal to nonverbal and thereby rule out the ambiguities involved in matching language to thought. They have subjects point to pictures or move game tokens according to verbal directions, and they watch what they do.[2]

---

[2] For example, consider a game called "Talk and Take," which Henry F. Olds, Jr. developed for a Harvard Ph.D. dissertation in psycholinguistics. Each card that players draw directs the player to move a board piece, and each direction represents a type of sentence of a certain difficulty: "Move a circle to any orange space, but do not capture a piece unless it is a square." A player who moves correctly according to this direction has to understand both the logic and the language in it.

Simply including directions as a category of discourse in the curriculum—how to do and make things—enlarges greatly the means of evaluation by increasing occasions for translating words into deeds. Using activity cards and game directions helps considerably also, because every time a student attempts to carry out their directions, he provides an observer an opportunity to assess his comprehension. In sum, evaluate comprehension mainly by seeing how students translate what they understand into either action, other media, or other words.

Implementing individualization, interaction, and integration in the classroom change utterly the picture of evaluation. Not only do they free you to observe constantly but they ensure the volume and variety that make daily observation the ideal means of evaluating. Even productive activities can be hard to assess if quantity is too limited. You may be uncertain how to judge learning from what a student says or writes or performs on only a few occasions but not from numerous samples. Similarly, if discourse is restricted to only a few types, you have no way of knowing what a student might do with other types and have insufficient data on which to base general evaluations about language strengths and weaknesses. Interweaving all of the language arts naturally alternates receptive and productive activities and allows one valid learning task to display what was learned in another. Consider, for example, a working party reading, discussing, acting out, and writing fables.

Finally, the pattern of decisions a student makes shows a lot about him. True individualization lets a student sift himself into those methods, media, and modalities he needs or prefers. You may decide, for your part, to intervene in this pattern, and that's part of counseling, but the point here is that by picking up on students' spontaneous patterns, you can assess tendencies you would not be able to see if they were all following a single, prescribed course. In a standardized curriculum evaluated by standardized tests, students all look alike except for some spread of scores. Individualization brings out their real differences, their full profiles. Much so-called diagnostic testing will tell you far less than what you can readily observe as individuals initiate and carry out their own programs of activity. Teachers who have worked with the curriculum described in this book say that they know their students as they never knew them under a traditional curriculum. This knowledge forms the basis of the most realistic and useful evaluation possible in the classroom, alongside of which standardized tests and quizzing seem slipshod and superficial.

EVALUATION FOR THOSE INSIDE THE CLASSROOM

Let's go back now to the different functions evaluation must serve for different parties.

■ SELF-EVALUATION

How does the student evaluate himself? The very essence of the action-feedback model of learning is self-evaluation. A person talks or writes or performs for a reason and for a known audience that responds to his production. Partner work, small-group discussion and improvisation, the writing workshop, rehearsal and performance, coaching from the teacher—all these reflect back to the learner the effects

of his language actions. If that action is receptive, it is linked to further action that is productive and hence can be evaluated by feedback. There is no other way, in fact, that people can ever know how their understanding of something compares with that of others than to take further action with it and see the results.

The curriculum presented in this book is so thoroughly committed to learning by doing, trial and error, interactive processes, and a responsive environment that student self-evaluation is a foregone conclusion and takes care of itself without need of any more setting up than the activities and materials already built in. We are saying, in fact, that only by continual self-evaluation can practice make perfect, and that language arts methods consist mostly of human feedback activities. The effects of action should be reflected back by as many different people as possible, by peers, teachers, aides, and whenever possible by outsiders and students of other ages. The practice of having students grade each other's work, by the way, will confuse and undermine this authentic feedback, because grades are an administrative matter and aim essentially at parties outside the classroom.

JOURNALS

From time to time students may find it valuable to reflect on all of the responses they have received, consider how they themselves feel about what they have done, and otherwise take stock of their own progress. A journal entry can be the focus of this. For beginners, suggest such prompts as:

I seem to be making progress in. ...

One thing I need to work on is. ...

One new interest I've discovered over the past few days (or weeks) is. ...

Veteran journal-keepers will find their own ways of expressing self-assessment. By looking at or talking over these entries you can get a picture of what is important to the student—information you can use in conferencing with him and setting up individual goals. See "Conferences" on page 250. See also page 158 on reading journals, which a student can draw on in pulling together his thoughts about his work. He can also look back over his portfolio of writing and over other productions such as tapes.

■ TEACHER EVALUATION OF STUDENT WORK

As the teacher, you have to assess student work not only for yourself, in order to coach and counsel, but to some extent also for students, as part of their feedback, and for parents, and for administrators. That is, what you perceive about student's work will naturally be of great value to all the other parties, because of your special, close-observer position, even though they should also assess independently of what you perceive. But you do not need to do separate evaluations for each outside party. All you have to do is transmit the same perceptions to each in different forms.

DAILY OBSERVATION

Follow the principle that evaluation is an organic part of your everyday role, not a separate function done on special occasions. Detailed, composite pictures build up

before your eyes of what each student can and cannot do, needs and doesn't need. The beauty of what you see when free to circulate and observe in your own classroom is that it gives you a slice-of-life view of the truth, because students are not thinking about being tested.

You stand near a group discussing a story they've read together and hear a student defending an interpretation that shows the same literal-mindedness you have noticed in the way he responds to others' figures of speech in conversation. You sit in on a group helping each other to revise some essays they've written individually and note what they are able to help each other with and what they are not, who shows confidence in his writing, who has trouble taking constructive criticism, and what aspects of essaying this particular writing activity is helping different students in the group to understand and create. You coach a student as he reads to you alone a selection he wants to work up to perform for others; while using you for rehearsal he's letting you assess his word recognition and comprehension. You join a group in playing an educational card game and can easily tell from the way each member plays how much he knows about the content, how well he can classify items, or how strong is his memory or understanding of directions or social cooperation. Watching a group perform a rehearsed reading of a poem, you note when the interpretation shows insight or incomprehension. You stop for a moment to watch a trio improvising and see how well they listen to and pick up on each other's words and body English, how inventively they exploit the situation, the range of language and role they take on that is not ordinarily their own. As you pass by the bulletin board, you note some new fables for old morals that a group has written, illustrated, and posted. And so on.

There is hardly anything you do to facilitate the learning itself that won't help you evaluate, for in order to coach, counsel, and consult, you must observe constantly anyway. The same information you act on daily you can selectively communicate to student, parent, or administrator when you need to.

Active, involved students produce so much to judge that it's not hard to remember your judgments, and less bookkeeping is needed. Immersed daily in this richness, you have stronger, deeper judgments that you won't easily forget. You probably should carry around a little notebook, however, in which you can jot down specific observations. Gradually you'll find out how many such notes you need actually to write down. If you have aides, involve them in evaluation. Ask them what they notice about different students and use their commentary to corroborate or complement your own observations.

Oral work particularly requires this ambulatory observation. Although many improvisations, discussions, rehearsed readings may get taped so that you can see and hear them out of class, many of them pass forever, and since you can catch only a certain fraction of what's going on at any one time, you need to overhear or sit in a lot. A major reason oral work usually gets so little emphasis in the curriculum is that it leaves no record for evaluation. Encourage students to tape often so that they can evaluate themselves and so that you may listen later if you were not present. It's critical not to slight, or let students slight, the many valuable speech activities simply because they do not leave marks on paper. Let all parties know that all activities are assessed all the time, but don't ever give the impression that the assessment is intended for anything but help and encouragement.

For many years, researchers working with group process inside and outside education have been developing various ways of doing "interaction analysis" of groups according to their emphasis—on the content of the task, the dynamics

among members, the emergence of leaders, the roles that various individuals take, or the differences that changes in size or purpose or organization make. But this is an excellent occasion for teachers to act as their own researchers, because you understand your students, and as teachers you have special things you want to know about group process.

We suggest you practice analyzing group process with a small group of other teachers. Listen together to an audiotape or watch together a videotape of some student exchanges, discuss afterward what you perceived, and develop criteria for assessing the processes of improvising, small-group discussion, writing workshop—what students are doing for each and what each is getting out of it. Work out for yourself and with other teachers some ways of analyzing what happens in groups so that you can assess the worth of the exchanges. In doing so, you can evaluate both individuals and the group process. If each teacher supplies some material from his class, and if you critique these together periodically, you will generate for yourselves about the best kind of staff development possible. Pool insights, troubleshoot together, and share the burden of formulating what to look for and how to evaluate these difficult but vital processes.

The chapters in Part Two, "Basic Processes," contain or imply things to look for. For some criteria of dramatic interaction, for example, see "The Value of Informal Classroom Drama" on pages 91–94. For help with evaluation of reading, see pages 152–155; of spelling, page 227; of talk, page 82; of performing, page 178; of the writing workshop, page 206. The specialized chapters in Part Three provide many indications of what to look for in the writing and reading of various kinds of discourse. For help with many specific aspects of development in language and thought see *Detecting Growth in Language*.

STUDENT FOLDERS AND PORTFOLIOS

You can best judge a student's work by generalizing from as many instances of it as you can have access to. Ask each student to keep a student folder into which all writing eventually goes after it has been posted, printed, performed, or whatever. Besides compositions (both early drafts and final versions), this folder could contain drawings, some sorts of journals, and any other productions or records that establish what a student has been doing. Actually, a box will replace a folder very soon if students are very active at all.

Review a student folder or box periodically, before a conference with student or parent or when you have to make reports. Some of the compositions you will know already from having seen them performed, from reading them posted, from hearing them discussed in a writing workshop, or from simply having read them alone for a conference. Other of the compositions you will encounter for the first time during your review. Students need to write more than you can process, and papers should not be simply gathered and "marked" or "corrected." They should be used first, as intended, then accumulated in the folder, sometimes after copying.

As research has demonstrated, most commentary written on papers is wasted. It's better to confer periodically with individuals about their writing, at which time you can talk about both particular papers and general tendencies. We recommend a mixed approach. Give some of your feedback during writing workshops, some during conferences, and some via written comments on papers. Don't insist that you should personally respond to all writing. When you do respond, take the role of a real reader who is also a helpful coach.

Ideally, written work is passed on from one year to another, or at least some selection from the student folder, which becomes a portfolio of items the student and perhaps also the teacher consider representative of the best the student can produce. The process of selection of a student's best efforts can in itself be a very valuable experience. Who selects? The student can bring his folder to his writing workshop, or to the teacher in a conference, and discuss criteria for selection—whether this or that piece is better and for what reasons, etc. Teachers who have a portfolio of work from a student from the previous year can counsel on a more informed basis. Also, students can look back over their work and sometimes use an old composition as a starter for a new one. Looking backward also helps students feel their progress. Passing on portfolios can replace so-called diagnostic testing.

Schools and teachers may do many different things with portfolios, depending on purpose and the coordination of them with other activities for learning and evaluating. Portfolios may, for example, be very selective and contain only one kind of work or only certain samples of the work. Other writing and material might go into another folder. You and your students can work out understandings about different sorts of folders according to who sees the material and for what purpose. Some journals and certain other writing, for example, might be accessible only to the author and the teacher, whereas if a school or district or state wants to sample student writing in various genres, a student's general portfolio might be available to copy compositions from if the author is consulted. A sound principle might be to save all or most material in the student folder or other container until the end of a year, making selections from it at any time for any purpose, including a selection to comprise the portfolio to be passed on to the next year.

The process of keeping student folders and selecting from them for assembling portfolios is a valuable educational activity, whether or not these portfolios are ever used for evaluation by those outside the classroom. However, portfolio assessment has now become an alternative way to conduct large-scale assessments at the school or district level, and state boards of education, state university systems, and the Educational Testing Service are discussing ways to use portfolios and to implement statewide portfolio assessments of not only writing but reading and math as well.

Not surprisingly, because the idea of portfolio assessment is so popular, the testing divisions of commercial book publishers are responding to this innovation by developing assessment tools they market as necessary parts of a portfolio program, and many of these are not the authentic assessment and teaching tool that a portfolio can be but rather a new way to sell simplistic reading, writing, and grammar tests. Teachers must ever be alert to the ways that good ideas get turned into commercial marketeering slogans and into commodities. The only authentic "portfolio product" is the portfolio itself. To look at either a student folder or a portfolio together, all a student and a teacher need is their combined experience in reading and writing themselves. They don't need commercial checksheets or lists of criteria.

CHARTING EXPERIENCE

On page 25 we described the charting and counseling system necessary for individualizing. The student needs to keep some kind of record of what he's doing, and the teacher needs to translate this into coverage of general language arts goals.

STUDENT TRACKING CHARTS. The form of these has to balance simplicity against utility. Don't overburden students with bookkeeping. All one can expect from primary students is probably a checking or coloring or circling or dating of kinds of materials worked with. Whenever possible, it is very valuable to know the titles of activity cards, reading selections, games, recordings, and so on, but less mature students may do well just to check or date an activity category, such as "Making Up Stories" or "Reading Books." During conferences you can elicit more detail from the student, such as specific titles.

THE TEACHER'S EXPERIENCE CHART. You'll probably want to keep an individual experience chart for each student. Some teachers find that all they need do is keep a photocopy of the student's record in their own file with a note as to the general overall quality of each piece of work for the end-of-term assigning of grades.[3] Others prefer a record that shows how often a student has worked toward each of the language arts goals in each of the kinds of discourse (crossing Part Two of this book with Part Three). Basically this consists of some layout of the ten kinds of discourse (stated as objectives on page 18 and treated in Part Three), permitting you to log under each kind the amount of experience a student has so far accumulated. Since these kinds of discourse are to be practiced orally and in writing, as sender and receiver, you could log not only the experience accumulated toward each goal but also indicate by which language art—listening, speaking, reading, writing, or performing—the experience was gained. An experience chart helps you in counseling. You can look at a student's chart before advising him about which areas of discourse and kind of activity to stress next. Teachers' records of cumulative experience can be passed on to other teachers at the end of the year or term in a portfolio or in a computer data base when teachers can agree on a common format.

Such a chart registers "how much," not "how well." That is, it doesn't attempt to measure quality of achievement. But because experience consists of direct practice of the target language activities, the charting of work with recommended materials and activities should in large measure indicate higher achievement in each goal area. Comparison with other measures such as direct observation may show that experience is high and achievement low in some areas. This may yield valuable knowledge about a student's learning efficiency. Other measures may indicate that a student is already so proficient in a certain goal area that he does not need any more experience in it, even though your chart shows he hasn't yet spent much time in that area.

CONVERTING FROM STUDENT CHART TO TEACHER CHART. Translating an individual's particular selection of materials and activities into general learning terms is hard, but because the very feasibility of individualizing depends on it, the difficulty must be faced. Many activities and materials may teach toward many discourse and literacy goals at once. You're not trying to determine if each student did certain required specific activities, because what is required are not *certain*

---

[3] Nancie Atwell carries around a status-of-the-class chart attached to a clipboard. See Chapter 5 of *In the Middle: Writing, Reading, and Learning With Adolescents* (Portsmouth, NH: Boynton/Cook, 1987).

activities but *some* toward each general objective—as many as each student can do, given where he started from. If the curriculum array comprises the rich variety that we advocate and that effective teaching requires, and if students themselves are constantly making and bringing in other materials, any correlated listing of goals to materials and activities would not only constitute a staggering compilation job but would be mind-boggling and time-consuming for you to consult if you had to keep looking up items in order to connect them to goals.

Such a job fits a computer perfectly, of course, if furnished with some kind of data-base software. Computers can, in fact, do both the students' and the teachers' record-keeping, including converting an individual's work into terms of language arts goals. If computerization could be limited just to logging student activities and correlating them to goals, without altering the activities or interfering with the student-teacher relation, it could be a strong help in solving a difficult problem.

Lacking such mechanical aid, the most practical way to convert a record of individual experience into goal coverage seems to be an informal means worked out differently by different teachers but relying a lot on conferences, where various records can be pulled together and you can find out from the student what each activity accomplished.

CONFERENCES

Meet the student with his portfolios, journals, experience record, and your own chart of his cumulative experience. Other useful materials are notes you may have made observing the student at work, doing miscue analysis of his oral reading to you, and examining his products. Incorporate into your perceptions those of aides who may have worked with him. Go over all this with him and note patterns, such as emphasis on one language art or type of activity, sameness or variety in both the materials and the people he chooses to work with, and any other traits or trends that will be helpful.

Try to think of this session as an opportunity to find out further how the student feels about his work and to help him set goals, rather than an evaluation or grading meeting. The more you can get him to level with you as he elaborates, the more you can help, and the more he will learn, but he must feel that frankness won't be used against him. This is how grades make counseling difficult. To encourage him to break ground in new activities, assure him you'll protect his grade by allowing for temporarily reduced success as he grapples with new challenges. Be sure to give reasons for recommending new directions, shifts of emphasis, and particular activities and materials. This is part of the student's education, and the more he understands the kinds of discourse that exist to become acquainted with and the ways you are trying to open for him toward all language use, the more he can successfully take over this decision-making.

Say what you see in his writing and other products. Try to describe rather than rate. Mention which *kinds* of spelling and punctuation errors he makes, what you notice when he reads aloud to you, the roles he takes in groups, the sort of reading and writing he gravitates toward, and so on. Generally focus on traits and trends. Although these often imply a value judgment, they emphasize fact and act as a useful reflection of the student to stimulate and guide growth. This makes you an ally instead of a judge.

Because of parental and societal pressures, students' feelings of self-worth and competence are in large part determined by their perceptions of their academic

progress. Many of them need to see evidence of this progress often (though this not need be in the form of a "grade"). One goal of a conference should be to enhance the student's feeling of mastery and progress. Try to do this in such a way that you wean him away from dependence on your judgment, however. You might start a conference with a question like "How do you feel your work is going?" or "What were you trying to do in this paper?" Concentrate on keeping the ownership of the work and the goal-setting with the student and not with you.

EVALUATION FOR THOSE OUTSIDE THE CLASSROOM

The best way for outsiders to evaluate results is to see and hear what students have done. Again, the typical slice-of-life most convinces, because it doesn't depend on rare, special occasions, as do standardized tests.

Parents have a right to know how their children are progressing. The same variety and volume that provide you with plenty to judge also display for parents the evidence of learning. They should see many of the papers from the folders, hear audiotapes, and whenever possible see routine performances, live or videotaped. Compositions and transcriptions will show handwriting, spelling, and punctuation and the creative abilities to think and imagine. An audiotape can catch an improvisation, discussion, or rehearsed reading. If a curriculum enables students to reproduce for themselves the kinds of materials they find in the classroom—books, recordings, learning games, photos, films, and so on—then you'll have no problem rounding up many things for outsiders to examine.

The problem is how to bring parents and products together. Open houses and parent conferences are fine to the extent that you can succeed in getting parents to come to the school. Alternating the two is a good idea. Many parents will not come to an open house but will come when appointments are made systematically. For conferences, arrange time so that they may look over their child's portfolio of work, sample tapes they may have made, and examine other of his creations. Then you can talk together.

You can describe traits of their child's work as it appears to you. Global value judgments mean less than specific comments about which kinds of talking, reading, writing, and performing their youngster tends to choose on his own, or has most and least experience in, which skills come easiest or need more work, habits and patterns, areas of recent progress. If you think you see why a learner is having trouble in spelling or in comprehending literature, explain this and say what you'll be recommending that he emphasize next. If you think, for example, both the spelling and the literary comprehension will progress better from reading while listening with recordings, then you could explain the connection to parents who might not understand how this practice can teach these two skills.

■ PARENT INVOLVEMENT

Try to involve parents as much as possible in helping as aides. The more they work with your students, the more they'll understand how the curriculum should function and will be able to assess results the same way you can, by observing while facilitating. These aides can help other parents to understand how much their children are learning.

Opening the classroom to the community not only shows confidence in yourself; it actually can head off unduly negative criticism of you. It often happens, for example, that parents concerned about "basic skills" will not understand why literacy is not isolated into drills and will fail to see how the skills are being taught by ways other than those they are expecting.[4] Such parents may believe that their child is just having fun and not learning anything. The more contact you have, one way or another, with parents, the more chance you have of correcting this misunderstanding by explaining to them where in their child's work these skills are being practiced and improved.

■ GRADES

Grades may be made up independently of any particular curriculum and may be done in myriad ways. We recommend that teachers not give a grade to individual activities but only to the totality of a student's work if they need to give grades at all. Make comments on work mostly descriptive and functional; make value judgments only to the extent that it serves a really good psychological purpose. The assessment of writing would be added in to evaluations you make of oral work and reading. Then, for a grade, make a blanket judgment on the whole of a student's work for the marking period. This is easy to do when you look it all over at once and confer with the student about it. Bookkeeping for grades alone is minimal this way.

Students who are used to receiving a grade on every piece of work may experience some frustration for a while, but they will come to appreciate the intrinsic rewards of authentic productivity. Once a school has operated this way beyond the memory of its current student body, most students won't even require adjustment. Practical feedback implies value judgments anyway, but if value judgment is minimized in favor of relating a learner's intents to his effects, the learner stays focused on the inherent learning issues instead of on grades. You can distill value judgments to satisfy parents and the institution from the evaluation you do anyway for purposes of coaching and counseling. Reporting that permits descriptive statements about a student's strengths, weaknesses, needs, natural tendencies, and so on (*qualitative* evaluation) will do more good than a letter or number on a report card, since it informs parents better and better facilitates administrative decisions within and between schools, including decisions about college admission.

Teachers should work toward the elimination of grades. Both students and parents must and do evaluate for themselves anyway. Grades maintain a competitive atmosphere that militates against learning. Students who receive low grades develop a low self-concept that often makes them perform worse than they would if no one had labeled them. Students who get high grades often think they know more than they do, especially if the grades are based on just a couple of things like reading-comprehension scores and grammar tests.

---

[4] A useful pamphlet to help parents understand the approach to writing recommended in this book is *How to Help Your Child Become a Better Writer*, available from the National Council of Teachers of English, 1111 Kenyon Road, Urbana, IL 61801.

So long as grades must be turned in on students, collaboration tends to be viewed as cheating and is therefore discouraged, because individual marks become harder to make up. Thus a powerful learning force is stymied. Further, growing up in an atmosphere that favors competition over collaboration produces adults unable to cope with either personal or public problems of today, which require collaboration. Surely the argument that school competition prepares for life must stick in the throat today, when all evidence indicates that far from helping people in life, competition is itself one of the major causes of both personal and interpersonal difficulties, not to mention international.

Grades distract students from the actual goals of effective communication. While competing and comparing themselves, they are also aiming to please adults, which should not be a school goal. Youngsters allowed to keep a pure learning focus will naturally please adults by becoming powerful learners. Methods and materials that cannot engage students without grades, extraneous rewards, coercion, or other irrelevant and artificial motivation do not belong in schools. It's not at all idealistic to assume that communication has its own rewards. If this has not appeared so in schooling, that's because purposeless exercises have too often reigned in place of real discourse.

Social reinforcement naturally plays a part in communicating and hence will always play a part in learning the language arts. Precisely because it is built into authentic communication, however, it does not need to be reintroduced by a reward-punishment system. Sender-receiver relations are broad and various and must never be simply boiled down to commands from a superior. Students should not practice discourse because big people make them, but that's the message implied by grades. A major goal of education is, precisely, self-evaluation. Grades constantly orient a learner toward what an outside observer thinks of his performance and encourage him not to judge for himself the effects of what he's doing. One judges communication, it's true, by its effects on others, but the "other" must be an authentic receiver, not a wielder of power over the sender.

Grades determine advancement to the next station, including eventually higher education, jobs, and careers. To the younger child, grades indicate acceptance or rejection of him as a person. To the older student they represent control of his destiny. Both feelings about grades play havoc with the learning process. The time must come when society removes entirely from schools this misplaced function of certifying. It is not the business of schools to certify people for jobs or for college. Again, each party should do his own evaluating. The mission of schools is learning, and that mission is impaired so long as schools continue to act as screening agencies for other businesses and other organizations, most in the private sector.

■ STANDARDIZED TESTS

One conventional way for those outside the classroom to evaluate what happens inside is to institute periodic testing with instruments sanitized supposedly against any bias of the teacher or the school. For this purpose many commercial tests are put out both independently and as part of curriculum packages. They are not of course sanitized, because teachers whose promotion is linked to their students' test scores may teach so closely to the test as to bias results heavily toward favorable scores. The only advantage of such evaluation is that student performances

may be compared with those of other times and places, which is not an advantage to those in the classroom and of dubious value to those outside.

For most well-known standardized tests, this comparison is either norm-referenced or criterion-referenced.

NORM-REFERENCING

In a norm-referenced test the score of anyone taking the test now is compared to the scores of some original "normal" student population. So a student obtaining today a reading score indicating "third grade, fourth month" is simply being scaled by the norms of that first population. Second, a student today may be compared with his contemporaries throughout a school district or, in most cases, the nation. The great weakness of norm-referencing, of course, lies buried back in the original "normal" population. How was its normality determined? Is it normal for today also or only for when the test was designed? But most of all, do norms established by performance of *any* population in public schools as we have known them do anything but set low standards? In an era of school reform this question becomes especially relevant. Norm-referencing reposes on what some students *did* do, under all the usual handicaps of conventional language teaching, not on what students *could* do under improved learning conditions. Teachers, parents, and administrators happy with "third grade, fourth month" may be accepting a meaningless standard, possibly a very low standard that holds back many youngsters.

Furthermore, comparison itself should be challenged. It serves, of course, the immediate practical purpose of selecting out students for this group or that class or certifying some for admission and employment. In other words, the more schools operate by limiting membership or admission, by segregating and screening for their own or other's institutional purposes, the more comparative evaluation seems to make sense. Parents need comparison only to the extent they're using their child to keep up with the Joneses. The student doesn't need comparison to "know where he stands" because good learning processes always show him by feedback how well he's performing, so that his only reason for comparing his performance with others' would be to know where he stands in the eyes of adults manipulating his destiny. For your own coaching and counseling purposes, ranking students has no value.

CRITERION-REFERENCING

To ensure meeting "minimal standards," schools are choosing more and more another type of test called *criterion*-referenced, which measures students absolutely, against a fixed standard, rather than relatively, against each other. The idea of it is by no means new. A Civil Service test to screen applicants for a certain job simply tests for those skills necessary to do the job, with little regard for how many applicants are likely to fall above or below the passing point, so long as enough pass to fill the job vacancies.

If you give a test to your students and grade afterward "on the curve," that is, by setting the passing point only after seeing how students do, you're setting up a kind of norm-referencing, because although you have no prior set of scores to go by, still you're scoring each individual according to norms that the class as a body

provides. If, on the other hand, you decided in advance how many errors constituted a passing grade, how many an A, and so on, without knowing how well students would do, then you would be criterion-referencing the test, because you're setting standards according to a desired performance, not according to comparison of student scores.

Is there, then, any more use for criterion-referenced testing than for norm-referenced? Is its specific performance a virtue? Consider the main purpose of criterion-referencing. It isn't to distribute students against each other on a curve. It's to find out which students or how many students can do certain tasks. The tasks tested for tend to set a floor, because they are selected as indicators that student and teachers are achieving *at least* such and such. Any testing tasks have to gear themselves somehow to realistic expectations of what students may achieve. If this gearing is not built in by some kind of averaging of what students do in fact achieve—that is, by norming—then it has to be accommodated another way, because schools find it politically disastrous to administer tests which too many students fail.

Criterion-referenced tests ensure that too many do not fail by including mostly safe items. They focus on "minimal standards." They're a pass/fail kind of test and assume that the large majority of students will pass. But how can they assume this without a prior score group? Obviously there *is* a kind of score group—in the minds of the test-makers—only it is not a particular population actually run through a particular test but rather a general notion of what most students have done, and can do, based on common school experience. Most children learn to master the long-vowel spellings—at least long enough to pass such a test, even if they never really read.

The chief value of criterion-referenced tests is to cover schools against charges of negligence or malpractice. It came to the fore in an age of legalistic accountability. Since each teacher must cover himself for each individual, students are tested each year for virtually the same material and hence taught the material again each year, so that they tend not to rise far but rather to hover over a required floor. In short, criterion-referencing differs not so much from norm-referencing as might appear at first blush, because both set low standards based on lifting large masses a short way. In a democracy, schools must keep a low center of gravity so that students can be passed on up the line. Standardized testing, ironically enough, tries to implement the democratic ideal of equality for all. But it is individualization, not standardization, that realizes this ideal.

For true individualizing, the only relevant measure is the student against himself. If schools take each individual as far as he can go, charting experience year to year, they will accomplish manyfold what they attempt by standardizing. If student achievement is measured by student-to-student comparison and by minimal thresholds, school achievement will remain low, perpetuating low standards and further low achievement.

Direct observation and direct examination of student products are the best ways to measure individual student progress. Standardized tests don't measure nearly a broad enough range of language activities or over a broad enough range of difficulty to be useful in individualized learning, which requires the same breadth of possibility in evaluation that it does in curriculum array. Tests covering all learning by all individuals would be impossible and obviously contradict the whole idea of standardized testing.

Criterion-referenced tests can be used as only one, inadequate sort of evaluation—just to reassure everybody that most students are getting over a threshold. School districts with students who are performing poorly will be under pressure to buy easier tests and to give students textbooks that are simpler to read, so the district can look as if its students are performing "at grade level." Any major evaluation would have to go far beyond these tests. Standardized testing overfocuses on a few, easily testable skills and ignores what is hardest to teach and learn and ultimately most important. The alleged strength of criterion-referenced tests is the concrete specification of the behavior to be evinced by a student on that special occasion of the testing twice a year. To fit the tight time compass, the test catches only the most specific, not the larger, more complex behaviors that cannot be seen or heard on one occasion but can only be built up into a composite picture by continual observation.

All too often what we have is a closed circle of test to textbook to teaching to curriculum guide and back to test. The designers of standardized tests often assure the validity of their test by asserting in their manuals that they have determined which items to include by consulting experts in the field and curriculum guides in use in schools. If these experts disagree, they use items included on at least three other leading comparable tests. Thus the test is devised. Textbook editors in the publishing houses use the tests as guides for designing workbooks and other materials advertised to ensure mastery of skills needed for the tests. Teachers who don't ask what else they might be doing look at the tests and teach by the book. The state or district language arts curriculum committee may well look at what is being taught to decide what should go into the district curriculum guides. Then the testmakers look at the curriculum guides, and round and round it goes—unless someone jumps outside the closed circle to ask what students *should* be learning.

■ READING COMPREHENSION TESTS

A widespread practice, critiqued on page 142 of *READING*, confounds testing and teaching by having students read short passages and answer comprehension questions immediately afterward. In other words, a whole "instructional" program, misnamed "practice reading," is made out of the examination situation by which comprehension is measured on standardized tests. This epitomizes teaching to the test, which can be carried no further than this.

Scores often do rise in these programs partly because the activity itself is nothing but constant test-taking. In addition, the increased scores may

- show that a student's reading skills may actually be increasing, since any practice may help, even if inefficient.
- mean that a student is learning to take this sort of test. Some youngsters quit reading the passages and simply go straight to the questions, referring to a passage if they need to. "You get used to the sort of question," as one explained. Even if the student can't see the questions until after reading the passage, he knows as he reads that he'll be questioned afterward in a particular way. This creates a very different frame of mind from ordinary reading. Also unrealistic is the short length of text and the short time span between reading and testing. For most real reading, one has longer texts to remember for a longer time.

- show merely normal growth in thinking and knowledge acquisition that would have occurred anyway without the exercises. This classic bugaboo of testing looms most at precisely the age when reading comprehension is tested most, which is when youngsters' mental growth is bounding along.
- reflect other learning occurring elsewhere in the curriculum such as the growth of concepts and inference through social and environmental experience.

The only kind of control to evaluate the effect of such treatment would be to put the same youngster through the same period of his life twice—once with and once without the treatment. Though a powerful reason to minimize *any* test scores, this lack of experimental control hurts comprehension testing far more than testing of factual material, because people have to acquire facts, whereas they're born with faculties for comprehending that will grow anyway. Such constant comprehension quizzing can never show if youngsters will read if they don't have to, or will want to read. More likely, as many teachers learn the hard way, it will misrepresent reading and kill interest in it.

Finally, the right-or-wrong multiple choice answers unintentionally teach the pupil that only a certain predictable set of implications and conclusions can be drawn from a reading text. (Notoriously, on standardized comprehension tests, brighter students often make mistakes because they see inferences other than the conventional ones the test-maker had in mind.) This falsifies profoundly the nature of reading by making interpretation of texts appear absolute, in contradiction to some main thinking in both literary theory and research in reader response. Readers frequently come away from a text with different understandings and, at the least, different emphases, because they bring to it different knowledge and attitudes.

It may be argued that the comprehension questions in practice readings test only obvious, consensual points. But that itself establishes a shallow value system, as if what's worthiest about reading is the obvious and impersonal, not what is personally significant. But this value system is only part of a terrible mind-set such tests induce. Readers are probationers waiting to be judged by external authorities. You read to provide scores. You read what people put in front of you for *their* purposes. Texts are arbitrary. Even if you answer the questions correctly, the snippets of text from hither and yon are meaningless because they have no relation to the rest of your life. The tragedy of comprehension tests is not that so many students fail them but that so many who pass them recoil from reading for life. The scores do not show this.

It would be hard to find a student who doesn't resent the inevitable quizzing, by the teacher or the printed questions, on what he has just read even when he *has* chosen the text. He has enjoyed the story and now he must face the music, endure the commercial, pay the piper. Has anyone attempted to estimate the damaging effect of this on a youngster's will to read? In rat-and-pigeon psychology, this administering of a pain after a certain act would be called "negative reinforcement," when it's intended to discourage the act. How many adults would read if they had to face a battery of questions afterward? Indeed, how many adults don't read because they *did* have to?

Apparently even designers of state assessments recognize the limitations of their programs. After saying, "Test results are not good measures of what is taught in school, strange as it may seem," the assessors of the Michigan accountability

system continue, "Even if the tests were completely valid and reliable, it would not be possible to attribute achievement gains to the school or teacher."[5]

It would be very good if a parent committee undertook to study the standardized tests used in your school, that is, to look closely at the tests themselves, read what testing experts say about them, and hear what teachers think of their influence on curriculum. Generally, the more people know about these tests, the less they want to rely on them and the better they understand the destructive side effects. You may find that parents can become good allies for changing evaluation. In fact, only the community may succeed in pressuring school systems to seek alternative ways of evaluating for those outside the classroom, who are, after all, the community and the system. Parents and other taxpayers certainly don't want learning spoiled in their name, once they know this is happening.

■ HOLISTIC EVALUATION OF WRITING SAMPLES

Blanket judgment of realistic writing samples has now become well-established in school systems that want to assess the efficacy of their writing programs. By holistic scoring, raters assess the total effect of a piece of writing. In order to establish reliability, scorers are trained by reading and discussing large numbers of papers similar to the ones they'll be rating. Because raters so trained are able to come up with high reliability in scoring, holistic evaluation has become increasingly relied on as a way of standardizing the evaluation of student writing. Like grades, it gives only a rating that permits ranking. Inasmuch as the samples are whole discourses, often now differentiated by actual types as described in Part Three, this way of assessing writing comes closer to fitting the approach of this book than other comparative testing, but it stands only about halfway between other standardized testing and the external assessment we're recommending.

School systems evaluating writing programs this way, like the National Assessment in Educational Progress, don't report on individuals but on schools, districts, or states, testing only readily selected individuals. Such scoring of writing samples frankly serves only parties outside the classroom. Even if scores were reported for individuals, they would tell students nothing that would help them to improve their writing. It's critical not to rationalize assessment for those outside the classroom as useful to those inside, for whom ranking can only be a distraction.

Furthermore, the basic problems of the standardized testing *situation* still haunt even this improved form of external evaluation. Writing something in a single sitting and under stringent time constraints, with no personal motive and no authentic communication context, rules out many hard-won principles of learning to write that most schools are still struggling to establish—prewriting, the writing workshop, a "process approach," and writing for real audiences. Some very important kinds of investigative writing, moreover, which schools need badly, will be ignored even more because the examination circumstances don't permit the interviewing, site visits, and other researching of original material on which this writing depends. True, some sophisticated evaluators are trying to make

---

[5] Ernest R. House, Wendell Rivers, and Daniel L. Stufflebeam, "An Assessment of the Michigan Accountability System," *Phi Delta Kappan* 55 (June 1974): 668-669.

examination circumstances flexible enough to permit some prewriting and group work; but the more they succeed, the more these examination circumstances will approximate the classroom conditions in which writing samples should be created to begin with. So carry this reform effort to its logical conclusion.

In other words, why not forget the conventional standardized situation, which artificializes the samples and in which students have no reason to do well? Instead, draw writing samples from portfolios, which can supply a variety of kinds of writing at any time. If they can't, a vital assessment has already been made by that fact alone. Formalists may complain that this is too uncontrolled, but the special examination situation controls so much as to falsify results, whereas the whole point of sampling a writing program is to find out what is really going on in it and (rightly or wrongly) to influence it by feeding comparisons back to it. The timed sit-down exam also sets a bad model of how writing occurs and indeed has created thereby much of the difficulty encountered in getting across the new, realistic approach to learning to write.

If, then, schools insist on assessing the writing *program*, they should do so through the slice-of-life way, which will disrupt learning the least and will actually accomplish better the purpose of tapping off a realistic sample of a school or school system's writing flow.

■ ALTERNATIVES TO STANDARDIZED TESTING

Granted that continual observation assesses best, how can outsiders like parents and administrators avail themselves of this means? Don't they have to send standardized tests into the classroom as a kind of reporter on their behalf? A reporter who can detect only a few limited things and who eventually causes those he's observing to do only what he can see is no reporter at all. Parents and administrators can understand this if they match the real learning goals against the tests. Then those in and out of the classroom can consider alternatives together.

The needs of the community and the administration overlap but also differ. The parent is interested first in his or her child; the taxpayer, like the administrator, in the efficiency of the whole curriculum. What's my child getting, and how much does the system deliver for what it costs us? The parent can evaluate the child at home by both observation and examination of products. The administrator need not be interested in individuals as such but in their aggregate welfare—in how many children are faring well, not mainly in which ones (except as subgroups or types). He wants to know how well the teacher and the curriculum are functioning. Nothing can satisfy any of the parties more than a highly productive, thoroughgoing individualized program, because it creates the most information and evidence about each student and hence about the total student body.

SAMPLING STUDENT PRODUCTIONS

Except to the extent that parents and administrators can actually visit the classroom, they are indeed handicapped for live observation. You have to act as intermediary for them, but they can certainly examine the wealth of products, and machine recorders make it possible to let outsiders observe outside, since audiotapes and videotapes can be sampled at the convenience of the outsider. Random samplings can be periodically made of a class by principals, language arts coordi-

nators, department heads, other school officials, and by rotating parent committees. Parents and officials doing this sort of evaluation together might find the collaboration and contact useful to both.

The products examined should not be especially selected by the teacher but should be pulled out by the outside evaluators in the classroom, though the teacher can cooperate by helping them find samples that show this or that sort of activity. Reading student reportage, for example, will allow them to connect composition with assignments that entail gathering raw material by visiting and interviewing. Hearing a taped improvisation while looking at the students' own transcription of it will show perhaps not only how improvisation may teach thinking and theater but also spelling and punctuation. Unlike standardized tests, this kind of evaluation permits relating cause to effect, in many cases at least. The more materials you and your students produce, the less you need worry about how any one of them may strike outside evaluators, and the more confidently can you trust the total impact of it all. If evaluators have an embarrassment of riches to choose from, that's their problem, and the quantity itself will surely count in favor of you and the curriculum. Many parents and administrators may need guidance, however, to know what to look for in such a rich setting.

When students learn by doing and by getting feedback on what they've done, the curriculum and the teacher can be evaluated by examining processes as well as products. *Classroom* assessment should be partly based on how well the individual's *self*-assessment systems are working through small-group processes and your coaching and counseling, samples of which outsiders can witness in person or on the same tapes that you and the students make for your own purposes. Assessing these has to involve subjective judgments, but all evaluation—make no mistake about it, "objective" tests or whatever—always comes back down anyway to someone's subjective judgments, however hidden. The more consensus, however, the more impartiality. A curriculum or a classroom operation can be very effectively evaluated by combining judgments of different human raters.

These can be combined with test scores, but to the extent that you and the curriculum are to be judged by the latter, then make sure parents and administrators understand that (1) they will either gain data about only a fraction of what you are trying to do, or (2) they may force you, in order to cover yourself, to teach to the tests and hence to teach only a fraction of what you should. In return for broader evaluation, you must willingly open your classroom to inspection any time. You can do so with confidence if you set in motion the practices recommended herein.

TEACHER RESEARCH

Teachers are in an excellent position to find out about which kinds of children learn from which kinds of activities. Indeed, many teachers now find it worthwhile to do their own kind of research about this as part of running a successful learning environment. Sometimes such a project may require special support, but often it's generated by the daily business of monitoring how well individuals and groups are working. Suppose you want to know if youngsters of a certain age are capable of a certain activity or whether, if you changed the directions a certain way for an activity that is showing problems, the results would be better. You experiment a bit, watch what happens, and make some notes, perhaps with another teacher. Talk it over with students and ask them to help you work it out. It's

their business too. In other words, only a fine line separates classroom research from what the inhabitants have to do anyway in order to troubleshoot and think more fully about what they're doing.

This research amounts to an excellent way of evaluating curriculum and methods for everyone's purposes, since notes and results can be shared with other school or district staff and become, indeed, an important part of an in-service program.[6]

CASE STUDIES

Other professions such as law, medicine, and business have found that getting down to cases helps practitioners understand better how well an enterprise is functioning and how it might be improved. Some colleges and private schools have done case studies of certain students to learn how they fared as they went through the institution. This can be a most valuable way to evaluate how well a program is working, at least for that type of student. A type might be based on a certain background or personality. How do children of nonreading parents respond to certain literacy activities; from authoritarian cultures, to individualized learning; from privileged homes to collaborative learning? Or simply, what happens to each of these, and others, as they go through a program combining these approaches?

Standardized tests give only quantitative information. You can sometimes parlay this data a bit farther by correlating some with other data and learn that working-class or minority children score lower on some tests, but you still have no *qualitative* information that can prompt improvement. Rating and ranking may tell you something is wrong for some children but not *what* is wrong or what to do about it. Case histories give another sort of slice-of-life that conveys the particularities of why some learning does or doesn't occur, the qualities of experience itself.

They can do this by bringing together much information already accumulated by and about an individual for daily learning purposes. Teachers don't have the time to be writing case studies *in addition* to their other work. But like other classroom research, it can arise out of evaluation that has to take place anyway to make individualization work. You're coaching and counseling and observing a student whose products are known to you and who is himself perhaps keeping reading and writing journals as well as writing about himself part of the time, revealing himself every day in choices he makes about reading and other activities. Together you and he are charting his past and future. No biographer has better data or opportunity.

Experienced teachers walk around with case histories in their heads. It doesn't take much more effort to get some of this down on paper or tape where you and others can consider it. Perhaps some money budgeted for evaluation and for in-service programs could buy released teacher time and secretarial services to compile booklets of case studies (that don't reveal the identities of the students). Reading and discussing these will help teachers a great deal to see how to proceed

---

[6] See Dixie Goswami and Peter Stillman, eds., *Reclaiming the Classroom: Teacher Research as an Agency of Change* (Portsmouth, NH: Boynton/Cook, 1987). See also Amanda Branscombe, Dixie Goswami, and Jerry Schwartz, eds., *Students Teaching, Teachers Learning* (Portsmouth, NH: Boynton/Cook, 1992).

in the future with other students.[7]

Schools and school systems will probably find case studies more valuable than test scores for monitoring the curriculum. After all, no one needs these scores to know if students are talking, reading, and writing well. That information for the outside can come from the internal evaluation going on all the time among teachers and students, who know anyway how everybody is doing. What external evaluation needs is precisely the qualitative description from the inside that teachers and students can supply together.

■ EVALUATION AS PART OF RESEARCH AND DEVELOPMENT

The resources put into standardized testing, in other words, would pay off better if combined with in-service resources to fund a kind of ongoing language arts research and development for both curriculum and staff at once. The evaluation that people in the classroom do every day could then be crystallized with not very much research effort into a major alternative to standardized testing while serving other purposes as well. This fits with the multiple uses of portfolios and tapes for student productivity and self-assessment, teacher guidance of students, curriculum development, and staff development.

The same audio recordings or videotapes that students make to preserve a performance or to play to younger children or to play back a group talk about a text, a panel discussion, or a writing workshop can be discussed by teachers for staff development and shown to parents and administrators for external evaluation. Similarly, if written work is frequently printed and posted and performed to afford students authentic discourse circumstances, it also becomes handily available for in-service use and examination by outsiders (this in addition to culling material from portfolios). In this way, investment in desktop publishing facilities and other means of dissemination can be justified for in-service and evaluation as well as instruction. Most good things serve several purposes at once. The best evaluation for all parties will come about as part of a total research and development program for curriculum, staff, and students.

SUMMARY

Tolerate standardized tests if required, but don't count on them much for evaluation. Only daily slice-of-life observation carried on without distracting students from honest language tasks will really tell you what you need to know and avoid negative side effects. Students taught by this curriculum should score well on standardized tests and do much more besides. And, similarly, tolerate grades if required but depend on student products and parent conferences to convey progress and problems. Encourage all outsiders to evaluate as you do, by observing processes and examining products. Assess the curriculum mainly by how well it enables students to assess themselves, that is, to get useful feedback about their efforts to comprehend and compose.

---

[7] See Glenda L. Bissex and Richard H. Bullock, eds., *Seeing for Ourselves: Case-Study Research By Teachers of Writing* (Portsmouth, NH: Heinemann, 1987).