BASIC WRITING

SPRING / SUMMER 1978

4. EVALUATION

JOURNAL OF BASIC WRITING

Unsolicited manuscripts are welcome. Please enclose two copies with a stamped self-addressed envelope. Authors should note that each issue of the *Journal* is devoted to one topic. The next two issues will be titled APPLICATIONS: THEORY TO PRACTICE and PROGRAMS. Inquiries should be directed to:

The Editors
Journal of Basic Writing
Instructional Resource Center
535 E. 80th Street
New York, New York 10021

$2.00 per copy
$3.50 per individual subscription
$5.00 per institutional subscription

Cover design by Jamie Ross

EVALUATION

Rexford Brown

WHAT WE KNOW NOW AND HOW WE COULD KNOW MORE ABOUT WRITING ABILITY IN AMERICA

In the area of measurement of growth and proficiency in writing, one of the major difficulties has always been that word "measurement." Few of us in the English teaching profession feel comfortable with the associations of precision and icy objectivity that accompany the word "measurement," and most of us were brought up thinking you can't "measure" writing. All of us have been "grading" essays for years—by which I mean doing a range of things from simply saying "uh huh" to students as we hand back their virginal papers, to actually granting two or three letter grades and obliterating their text with such strange glyphs as "awk", "punc", "frag", "dang", and "rewrite by Friday."

A major advantage to the word "grading" seems to be that it supports the widespread feeling among too many of us that standards for evaluation of writing are somewhat personal. We are all very careful to respect each other's right to a private grading system, even if it is arbitrary, wrong-headed, nasty, or capricious. Criticizing a colleague's values in this area is academically equivalent to crossing a picket line. In such an atomistic climate there has been little room for the idea of measurement because we have assured ourselves that there are no shared units of quality, there is no bureau of standards. Proficiency in writing in this climate is expressed as a letter grade, and growth can only be expressed as an improvement in grade. The fact that a writer can improve his writing between his freshman and sophomore years, but receive a lower grade because his second teacher holds different views from the first, bothers no one but the poor student. Nor do we seem particularly concerned about the fact, easily borne out in a number of studies, that our own grades are subject to many kinds of bias, and fluctuate randomly in ways that few of us can control.

Rexford Brown is Director of Publications with the National Assessment of Educational Progress.

1

It is interesting to note that the first major advance toward large-scale measurement of writing samples was successful largely because it did not seriously threaten the picket-line principle or challenge in any way the professional conspiracy of silence about quality. Educational Testing Service has for many years called together radically different people, trained them to recognize certain papers as 3s or 6es (don't ask why, just internalize the training papers), lavishly paid them to read hundreds of papers and respond knee-jerkily to each with appropriate scale numbers—and never ask any of them to lose face by revealing that they might have harbored perverse or insane notions about what constitutes quality writing. Holistic scorers need never explain what they are doing; and thus did holistic scoring achieve a certain amount of respect in our profession. Measurement got a foot in the door by pretending it was not measurement.

We've learned that large numbers of essays can be reliably scored with the holistic method and that these scores are accurate predictors of college success. And we've learned that teachers can be trained to agree on something. But what do holistic scores mean? All anyone knows after a holistic scoring is that paper A is higher on the scale than paper B. But, since no one discussed quality criteria, no one knows why. Furthermore, it is possible that all of the papers at the top of the score are horribly written. They may be better than the rest, but still may be unacceptable to most teachers of composition.

Not only is this traditional holistic scoring incapable of establishing proficiency in any concrete sense, it is a very unsatisfactory system for the evaluation of growth. If a student's first paper is rated 5 at a September scoring session and her 20th paper is rated 6 in a May session, we know nothing, because experience has shown that holistic scorings cannot be replicated reliably. We know more about growth if both papers are included in the same scoring session and the second paper comes out higher on the scale; but we still don't know why it is better or how good it is in an absolute sense.

No matter how reliable holistic scoring is as a way of rank-ordering papers it is inadequate as a measuring tool in itself because it is entirely relativist and value-free. It is not tied to any absolute definition of quality. The most promising *modified* holistic scoring approach I know of is the National Assessment of Educational Progress (NAEP) "Primary Traits" system. Developed to counter a glaring fault in traditional holistic scoring—that you cannot report results in useful or even meaningful ways—the system rests upon elementary rhetorical

2

theory. It assumes that a carefully defined writing task is a statement of certain rhetorical imperatives; that successful completion of the task entails understanding of and responsiveness to those imperatives; and that degrees of success are definable in concrete terms. We have found the tasks hard, but not impossible to define, the scoring guides complicated but teachable and the actual scoring reliable. Most importantly, we have found the results reportable in terms that have curricular implications.

For many teachers, holistic scoring has been a luxury only the rich could afford anyway. Still reluctant to define quality, but nevertheless in need of evaluation systems, they have used objective, multiple-choice tests of writing ability. Such tests are cheaper and easier to score; best of all, they enable any user to say "Well, *I* sure don't define writing the way those test developers do, but I'll accept their claim that the results correlate with writing ability; and after all, these are the only tools available."

But machine-scorable tests *also* suffer from some glaring weaknesses. Their primary function is, again, to rank order people on a scale. This leaves us again with no absolute knowledge about writing ability and a slight sense of embarrassment when we tell people we'll test their writing ability by not requiring them to write a single word. *Of course* these tests correlate with writing ability and predict academic success; but the number of cars or television sets or bathrooms in one's family also correlate with his writing ability, and parental education is one of the best predictors there is. All existing objective tests of "writing" are very similar to I.Q. tests; even the very best of them test only reading, proof-reading, editing, logic and guessing skills. They cannot distinguish between proofreading errors and process errors, reading problems and scribal stutter, failure to consider audience or lack of interest in materials manufactured by someone else. Like holistic essay scoring, multiple-choice testing of writing is seldom diagnostic in any useful way. And since capacity to recognize problems in other people's writing does not insure capacity to avoid them in one's own writing—especially first draft writing—we can never be sure what the final scores on such tests mean, let alone the subscores.

There are even more insidious aspects to multiple-choice writing tests. They require a passive, reactive mental state when actual writing requires and fosters a sense of human agency, an active state. And they are necessarily incomplete, leading the student and perhaps even the teacher to believe that those aspects of writing most easily tested—sentence

3

structure, word meaning, spelling, punctuation and outlining—are the most important to teach and learn. Finally, since the approach of many such tests is to emphasize differences between standard and nonstandard usages, writing courses all too often become, unintentionally, cultural programming laboratories.

No, an objective test all by itself is not a very good measuring device either; it tells us something, but not enough that is concrete. But the proliferation of such tests over the years has softened the profession up just a bit more toward the idea of measurement and the possibility that there are some shared units of quality upon which to build more accurate and useful systems of evaluation.

We're ready now to work toward the creation of many such systems. The pressure is on from the public, the deans, and the students themselves to improve writing. In order to do it, we're going to have to know more about the process of composition than we do now, and we're going to have to know more about what is wrong—in concrete, absolute terms—with student writing. Even our agelong system of medieval fiefdoms—separating the Miltonians from the linguists from the English educators from the modernists from the rhetoricians from the Marxists from the graduate-student assistants who teach freshmen composition— even that is crumbling under the economic and social pressures so familiar to us all; and this crumbling makes possible a movement toward professional discussion of quality in writing. The picket-line principle is doomed.

We have learned a great deal in the last fifteen years about the strength and limitations of the various holistic scoring systems developed at ETS, National Assessment and elsewhere; we know what is useful and valid in such good objective tests as the Houghton Mifflin College English Placement Test and the ETS STEP test; our knowledge of syntactic maturity levels has been advanced by the work of people like Walter Loban, Kellogg Hunt, Lou LaBrant, Roy O'Donnell and others; the contributions John Mellon and Frank O'Hare to our knowledge of the relationship between sentence combining activities and syntactic maturity levels have opened new and exciting evaluation opportunities; the rebirth of rhetoric, and the particular contributions of Francis Christensen, Ross Winterowd and Edward Corbett have given us new frames of reference for definitions of quality that facilitate concrete evaluation.

We can create from this fund of knowledge and this special climate a number of evaluation systems that define proficiency in concrete terms, are sensitive to degrees of growth toward that proficiency, require people

4

both to write essays and test their editing skills, are valid and reliable, are cheap and—most importantly—*are coordinated with the long-range research effort we need to more fully understand and develop strategies for improving the process of composition.*

Here are some suggestions about how to develop an ideal instrument:

1. Make students write—but there's no need for more than 400 words on test essays.
2. Base essay evaluations on papers reflecting several models of discourse, because quality differs for each one and people are not equally proficient in all of them.
3. Teach testers how to write directions for essay examinations. If you want to evaluate an essay for certain characteristics, then you must be sure that you have requested them in the assignment. This is not a trivial matter: it is extremely difficult to write assignments that define precisely the rhetorical imperatives that will either be met or missed by the students. If you want to know whether they can elaborate upon a role expressively while maintaining control of point of view and tense then you have to set the task up in such a way that they must do so, and define acceptable levels of achievement that are concrete and realistic.
4. Use computers. Have people mark off T-units in the essays so you can gather information about number of words per T-unit, number of clauses per T-unit, number of words per clause, number of adjective clauses, number of noun clauses, and so on— information about embedding, in short, which ties you directly to indices of syntactic maturity.
5. When you have these counts, tie them to holistic scores. If the scorers cannot or will not tell you why some papers are better than others, the computer will at least give you an idea of what was influencing them.
6. Tie the counts to various criterion-scoring systems. The six factors that seem to affect judgment most are ideas, mechanics, organization, vocabulary, what Paul Diederich calls flavor, and handwriting. Each can be evaluated independently.
7. Define coherence in specific syntactical or transformational terms, have graduate students code papers accordingly and establish a concrete coherence scale.

8. Include in any instrument questions about writing attitudes, prewriting activities and rewriting activites and then look at results in the light of that information.
9. Require basic sentence combining exercises and tie results on such exercises to actual writing performance.
10. Include a battery of objective items that will at least remind students that they should edit.
11. Use the resources we already have. National Assessment's huge corpus of essays remains largely untouched by researchers. Ross Winterowd at the University of Southern California has received seed money from the NCTE and Carnegie Foundations to keypunch representative samples of NAEP essays for research into the syntactic features of coherence and other vital matters. But the research undertaking itself has not yet been funded. Various graduate students here and there have used bits and pieces of the corpus for various projects, but they have only scratched the surface. Much that could enrich our understanding of the composing process and those aspects of it that cause most confusion for students of writing remains undone. The situation will probably continue until more is known about the availability of national and state assessment score materials.

The next national assessment, currently under development, will include most of these features. In addition, it will include materials from 1969, enabling us to examine trends spanning a decade. But however good it is, it will not be sufficient to gather all the information we need at the necessary level of detail. For that we need a coordinated effort involving writers, teachers, linguists, anthropologists, rhetoricians, philosophers, data gatherers, and educational psychologists. Professional conferences, which bring together such people, must serve as the model for the inter-disciplinary approach which alone can promise sufficiently sophisticated understanding of our situation. Perhaps, after more such meetings I will be able to provide more concrete information about achievement in writing and more exciting and practical specifications for its assessment.

Joseph Williams

RE-EVALUATING EVALUATING

So many questions we ask about writing, about teaching it, about research in how to teach it turn on the problem of evaluation that we ought not be surprised at the energy we expend on devising reliable ways to measure our competence at putting one word after another. Who goes to which college, once there who remains, whether the quality of our national prose is sinking toward illiteracy—all such judgments depend on whether we can (1) identify what in a text is most salient to determining good and bad writing and (2) measure it consistently enough to make the measure more than a reflection of its inventor's good taste.

Nor ought we be surprised at those who wonder why the profession hasn't settled the question long before this. Not many other fields have devoted more effort at establishing clear-cut standards of evaluation with fewer results. The NCTE has published a whole collection of measures, none of which are unassailably reliable. The National Assessment regularly assures us that our intuitions about a decline in the writing ability of our students is not the product of irritable old age, but it continues to search for better criteria to evaluate student writing. The Educational Testing Service invites·only those it is reasonably sure can grade essays consistently to read the College Placement exams, but still devotes large amounts of time to regulating the grading for consistency and reliability.

The search for reliable criteria has gone in two general directions. One is toward objectively quantifiable features of a text that might correlate with different levels of maturation. These include clause/T-unit and word/clause ratios, counts of errors in grammar usage, number of words written in time, and so on. The other is toward systems that would make more accurate, valid, and consistent the wholistic judgments of paper graders. This has taken the form of training graders to be consistent in

Joseph Williams is Professor of English and Linguistics at The University of Chicago.

7

looking for and evaluating whole essays or particular features of essays, of finding ways to sum differential responses into valid wholistic scores, and more recently, of weighting by various mathematical operations the responses of different graders so that the score of those who tend to grade too stringently and too leniently will be respectively raised and lowered to fit a median response.

Our inability to find simple and reliable measures stems partly from the complex nature of written texts and from our equally complex responses to them. Different components of a written text elicit our judgments and responses from a variety of conscious and non-conscious levels. At any moment, any one of those components might touch most saliently on any one of our responses and thereby dominate the final wholistic judgment. More personally, we may not be able to agree on explicit criteria partly because so few of us are qualified to make reliable judgments in the first place. I suspect that most teachers of composition themselves write fewer words per week than their students, and the vast majority among us never have to write for keeps, never have to produce anything as consequential as a production report or a planning memo.

What follows is not especially a critique of any of the specific methods we now use to evaluate student writing, much less a new one. It is intended rather to raise some questions that I don't think we have attended to as carefully as we might have. I wish I could say that I think the questions will help simplify this matter of evaluation, but in fact their answers, such as they are, seem to complicate it.

Let us suppose that we finally devise a system of training an English teacher to respond consistently within his own grading and with the grading of others, and that we can reliably count objective data such as T-units, errors of grammar and usage, and so on. When we have done this, we would have a means to rationalize and defend admissions procedures, grading, the adoption of better teaching methods, and judgments about any national decline in the writing ability of our college population.

But it is not at all clear that such a system would be more than a self-justifying instrument that had taken its values and hence its measures from those who have not demonstrated any special competence in distinguishing competent writing in any world except their—our—own. That is a harsh charge to make against a whole profession and by no means includes every member in it. But I think it is essentially true.

I want to begin indirectly. Consider for a moment, the American Heritage Dictionary panel on usage and its findings. The criticism heaped on them and their judgments by those familiar with the realities of

modern usage is deserved. It is not merely that they did not represent educated, literate writers. (They averaged a year past retirement age and were by and large, Eastern educated or Eastern employed or both, and for the most part no more technically qualified to pass judgment on good and bad usage than those who edit them.) More seriously, that their judgments were solicited and quantified virtually assured the most Neandralithic sort of majority opinion. No one who has spent a life-time tangling with editors, themselves steeped in 19th c. rules of usage, will easily contradict a body of knowledge it took them years of abuse to acquire. Asked point blank whether the verb *contact* meaning "to get in touch with" is appropriate in formal usage, what could a 66-year-old writer educated at an Eastern university and writing for an Eastern seaboard publication edited largely by others of the same sort be expected to answer, particularly when he knew his opinion would be recorded and printed? The very fact that a writer had achieved an editorial eminence sufficient to call his name to the attention of the AHD staff suggests that he had accepted the values his position implies. And the very acceptance of the solicitation to join the panel constituted the final step in guaranteeing that the panel would be a bastion of linguistic conservatism.

But even if the members of the panel did fairly represent those in the world of letters, their judgments, no matter how close to a consensus they might come, ignore two questions which all such overtly compiled evaluations fail to address. First, even if the proscribed items do not appear in edited, publically printed prose (and it is not the case that they do not), we do not know how often they may appear in that considerably more voluminous quantity of unedited and unpublished prose generated by educated writers in government, industry, commerce, and the professions for their purely internal and private institutional consumption.

Now on the one hand, our professional response is to assert that the standards of usage in studiously re-written, edited published prose should constitute the standard of usage for *all* prose. It is, after all, the sort of prose that is written and presented with the greatest care. But the concept of "care" here is a misleading one. There is no analogy to being careful in, say, medical practice or engineering, where carelessness can have immediately self-evident, objective consequences. Patients die and bridges fall. In writing for publications, the concept of "careful" in regard to a rule of usage has good or bad consequences only to the degree that a reader responds to a violation of that rule.

But if in private prose any rule that holds for public prose is broken

9

and not responded to as a violation, then to justify the rule we would have to assert that such readers "should" respond negatively, that if they don't then their education failed them. Now this is a curious argument. It requires us to accept the idea that we must arbitrarily generate consequences where none before existed. (The argument that by observing some set of rules we prevent the language from degenerating is, of course, empty.) The only non-arbitrary non-socially based argument for honoring a rule would be if the rule contributed to ultimate clarity. But we know that the overwhelming majority of the usual rules of usage we find in the manuals have nothing to do with clarity or economy, but represent only a set of items whose capriciousness guarantees their imperviousness to mere logic.

In truth, we have publicized a variety of linguistic items as distinguishing literate from illiterate speech, but we have accepted these rules without determining whether educated writing that is *not* edited by people especially trained to identify violations of rules displays those items. We do not know the degree to which these items of usage have been circularly perpetuated as a standard for educated writing because of our assumption that public, printed writing, self-consciously edited by those paid to perpetuate those items of usage, should constitute the standard for *all* educated writing.

Unfortunately, we cannot answer any of these questions by asking. We are all thoroughly familiar with the way almost any educated but linguistically naive person who is put on the spot about correct grammar begins to speak quickly and nervously about grammar being his worst subject, and so on. To directly ask educated but linguistically naive informants would invite only those answers that they could dredge up out of their most insecure memories of junior high school, particularly when it appeared that they were being interrogated by the types that trained those who terrified them in the first place. Nor can we ask them to correct papers in which we have inserted a variety of usage problems, for that would induce even greater uncertainty since such readers would not only have to worry about the correct answers but the correct questions, as well. And even if we examined the writing of this group and found few or none of the items of usage we were looking for, we could conclude nothing, because their absence says nothing about the possible responses of readers if those items were present.

Theoretically, the best way to determine what counts as an error in the minds of non-academic, non-print-oriented writers would be to have them read reports, memos, and so on that each reader had to approve

and send on to his own superior, memos and reports into which had been inserted one or two items of debatable usage, and to repeat this process with many such readers and several items until we found those items for which they would not risk their own prestige. Any more direct method is certain to call up the most regressive sort of response.

Three cases to illustrate what I mean:

(1) I am in the process of drawing up a program to evaluate the quality of writing in the investigative office of a Department of the Federal Government. This particular division has been having increasing problems with the reports prepared in the offices around the country for the rest of the divisions in the Department. In the last two years, according to the director of the office, some of the reports have been delayed for up to six months while their prose was being revised and re-revised into a modest degree of intelligibility. During those two years, the division set up tutorial writing programs staffed by English teachers from the areas around the regional offices. In discussing with the officials the sort of program this division might find useful, I asked to see the comments those teachers had made on the reports they reviewed. They were about what we would expect to find on a carefully marked freshman essay. I asked whether one of the corrections, faulty parallelism, was a serious problem among the report writers. First response: Hesitation; second response: "If he says so."

Now this is an interesting response. A problem exists if the English teacher says it does, even though it may not be felt on the nerves of those who read the reports. None of the administrators would need an English teacher to tell them which reports were disorganized or illogical or pointless or lacking in supporting evidence. Nor would they need English teachers to tell them which sentences might be manifestly nonstandard: *Don't nobody know what goin' on in them offices.* The English teachers were called in to address a perceived problem that seemed to fall between areas which are not the peculiar domain of English teachers. The problem is for us to understand what that domain peculiar to our profession properly includes.

It certainly includes style, particularly in those sentences so confused and prolix that they fail to express what the writer meant. And it ought to include all the rules of usage, both those that are observed by the best publications and those that are observed in literate non-published, non-edited private writing. The crucial problem is not to define literate by the rules germane to print.

But the response, "If he says so," suggests that some believe that there

are other problems which either impinge on our understanding of a text but are beyond the conscious articulation of naive readers or that there are problems which in some metaphysical way violate platonically defined rules of usage. Parallelism, at least in some of its manifestations, may be such a rule. Are the readers of those reports faulted for violating strict grammatical parallelism conscious of that violation? Always? Never? Only certain kinds?

(2) Professor Rosemary Hake of Chicago State University and I have been conducting some research into the ways English teachers respond to different kinds of styles, particularly what we have been calling nominal and verbal. Here are two contrasting examples.

> Nominal: There is a need on the part of this office for a determination in regard to the resolution of these matters.
> Verbal: This office needs to determine how it is going to resolve these matters.

Given these two sentences point blank, no English teacher reading this would recommend to his students the first as a prose model. And yet when pairs of essays differing only in these two styles were at different times given to English teachers from a variety of institutional backgrounds, most tended to grade the essays written in a nominal style higher than the essays written in a verbal style. What many of us claim we reject we seem tacitly to prefer. The connection between (1) and (2) seems fairly clear: Not only do we not know how readers outside our profession regard different features of language; we cannot even say that we are entirely confident that we know how we respond to them ourselves.

(3) We have replicated this research under a number of different conditions. We have given the papers to graders to take home and mark at their own convenience. We have brought them to the University of Chicago on two Saturday mornings to provide responses they knew we would examine. We slipped papers into a state-wide examination required of all graduates of public colleges. These three contexts set increasingly stringent demands on the graders. The first was entirely non-threatening. The graders had done exactly the task given to them with other papers on other occasions for other purposes. No one was watching them and so far as they knew, they had no reason to be insecure in their responses. The second situation was the campus of a prestigious university where a different set of graders were providing data for research they knew would reflect on them (though they did not know the

12

specific nature of the research). In the third situation, the graders knew they were constantly being reviewed by grading proctors supervising the entire state-wide examination, proctors who would recommend which of them would be invited back to grade again, a decision that would have both financial and professional consequences, for an invitation to read the essays is regarded as a significant mark of professional recognition.

As the pressure of explicit review increased, the overall average of paper grades declined. And it declined most markedly among those on the lower end of the totem pole: among high school and junior college teachers.

The conclusion that suggests itself would certainly seem to be that the more explicit and personally consequential the task, the more conservative and disapproving become the responses. In light of points (1) and (2), we must become even less certain of what we know. Most of our evaluation is done under self-conscious circumstances. Our own performance is subject to review, if not by our peers, at least by our students, who would like justification for whatever grades we give them. We are only too happy to find criteria to defend strict judgments, judgments which testify to our strict standards. But when we read as unself-conscious readers, we seem to respond rather differently from what we might predict. In what you have read so far, for example, there are a number of errors in usage.

One of the tasks in the preliminary evaluation of the government writing project is to answer as many of these questions as we can. We will circulate among a variety of officials reports into which we have inserted particular errors. We will ask them to read the reports for their content, and only incidentally, to suggest any changes in the texts they think appropriate. The primary task will be to read for overall quality. Of course, even if no one identifies any of the items we insert as errors, we cannot conclude that those items are entirely irrelevant to how readers actually respond, for it may be that they respond to them at some non-conscious level. For this reason, we will recirculate essentially the same documents with the "errors" corrected to determine whether the "corrections" raise their evaluation.

When we turn to the less objectively quantifiable and more subjective questions of style, the problems of evaluation become no less tangled. I have already mentioned the results of Professor Hake's and my research on responses to nominal and verbal styles. Despite the fact that we might all claim that we prefer a clear, concise, direct style with lots of strong

verbs and few abstract nouns, a very large number among us, if our findings are accurate, grade an essay in a nominal style higher than exactly the same content in a verbal style.

But we are faced with essentially the same problem here that we faced with problems of usage; We do no know what counts as good style in places not familiar to those of us in English departments. This is one of the problems our preliminary evaluation of the government writing project will also have to speak to. We are familiar with the turgid bureaucratese that all of us hoot at. Indeed, this is one of the problems of the division we are investigating: Its administrators refer to it as a lack of clarity, as confused sentences and so on. But much of the problem seems to derive from the most common feature of bureaucratese, indirect nominalizations. The deeper problem is why report writers may value this heavy, indirect style more highly than a simple direct style. (At least we tentatively assume that it is valued more highly, since that is the style they use.)

It may be that two systems of values are competing here. On the one hand, the administrators want something that they can read quickly and easily, but the report writers are unwilling—perhaps unable—to be simple and direct. It may be a consequence of bad writing habits, but it may also be the consequence of the first rule of a bureaucracy, not to make oneself responsible for anything. Findings and recommendations couched in governmentalese at least partly cover the writer's ass from recrimination.

Under these circumstances, there is no simple answer to what counts as a good style. In our scholarly innocence, we might value the simple and direct as transcendentally good, much as Thomas Spratt did in the 17th century when writing about the ideal style for scientific prose. But in the real world of government bureaucracies, GS 10's and 12's are—or may be—looking over their shoulders to see who might be watching. And considering the state of a good deal of academic bureaucratic prose, we might have a hard time deciding who among us should cast the first stone.

Questions such as these, of course, also touch on attempts to quantify syntactic maturity. If we can define bureaucratic prose as that hyper-mature writing with more than one nominalization every five or six words, then most recent pedagogical efforts seem to be directed more toward increasing the syntactic maturity of a writer in the direction of bureaucratic abstraction than toward the pellucid prose of an E.B. White. Despite Hunt's disclaimers that increased syntactial maturity is

not to be equated with increased quality, the sense of accomplishment in most recent research papers reporting such gains suggests that raising the syntactic maturity of a ninth grader to the level of a twelfth grader is an unqualified good. The fact that graders also reported an overall improvement in the papers only underscores the value they attribute to syntactic growth. And in the absence of any evidence or arguments to the contrary, there is no reason to disagree.

What follows is not that evidence nor even the argument as much as some questions about syntactic maturity and its unqualified use as a means of evaluation.

As a writer matures, syntax is not the only feature of prose that becomes more complex: organization, a sense of audience, clear intentions, close logic, and so on also mature. One important question is the order in which these mature. We know that projecting ourselves into the role of audience is something most of us never completely master. Nor are logical arguments as natural a level of achievement as, say, puberty or 11.5 words per clause. Thus we ought not accept quantitative measures of syntactic development as good indications of—what shall we call it—rhetorical maturity, regardless of the attractive objectivity that the quantitative measure seems to provide. In fact, syntactic maturity may be a misleading measure, at that.

The figures most often cited are these:

grade:	7	8	12	superior adults
words/T-unit	9.99	11.34	14.4	20.3
clause/T-unit	1.30	1.42	1.68	1.74
words/clause	7.7	8.1	8.6	11.5

Hunt and O'Donnell have suggested that of the two, words-per-clause most sensitively indicates growth. The main problem with this measure is that we have no idea what affective consequences these figures entail. Do we affectively discriminate between texts whose word/T-unit ratios differ by one word? two words? three words? Physiological maturity is ordinarily accompanied by a change in the ratio of cartilege to bone, but under most circumstances, the results of those changes have no appreciable consequences on how clothed adults relate to one another. Growth is a fact of maturation, but it makes no social difference. Word/clause ratios increase as a writer matures, but where is the threshold for perceived differences? There must be some difference at some point, but we have no idea where, and if we have no idea where, then we have no

valid way of making the evaluation relevant to our rhetorical concerns.

Furthermore, though Kellogg has figures that reflect the prose of what he calls superior adults, those whose self-consciously written, revised, and edited prose appears in the Atlantic Monthly and Harpers, we have no extensive figures for workaday world prose, the private prose I described earlier. A case in point: Professor Hake obtained a number of memos and reports from a large manufacturing concern in the Chicago area. We asked those administrators who had to act on the documents to rate the perceived quality of the prose on a scale of 1 to 10, according to whatever criteria seemed appropriate. We selected several from the extreme ends of the scale and analyzed their clause/T-unit ratio. Those rated low on the evaluation had a clause/T-unit ratio of about 1.5, roughly equivalent to the prose of a ninth grader. The documents rated high, on the other hand, had a clause/T-unit ratio of 1.3, about equivalent to the prose of a seventh grader.

Now at first glance, this would seem to contradict the figures that Hunt and O'Donnell gathered, but in fact, it tends to confirm them, unfortunately. A lower clause/T-unit ratio means a higher word/clause ratio, the figure they identified as most salient to maturity. When we recall how our evaluators responded to nominal and verbal styles, the pieces fit together. The memos with fewer clauses had more nominalizations, a construction which reduces the clause/T-unit ratio and increases the word/clause ratio. And a text written in a style with more rather than fewer nominalizations tends to be evaluated more highly than one written in a verbal style.

But doesn't this present us with a pretty problem? We English teachers —and virtually anyone else we might ask point-blank—would almost certainly prefer a verbal style for reasons none of us would find difficult to articulate: clarity, economy, directness, honesty, and so on. And yet when our preferences are probed indirectly, quite another set of values and responses seems to emerge, at least for a large number of us. If this is the case among writers of private prose as well, as the evidence slightly suggests, ought we English teachers adopt such criteria not merely as a measure of syntactic maturity but as explicit objectives in the teaching of style? Just as we have sentence-combining exercises we might have nominalization exercises that would by increasing the frequency of nominalizations lower the clause/T-unit ratio and raise the word/clause ratio.

An argument could conceivably be made that such an objective would not be entirely dishonest. As we have mentioned before, a heavily

nominal style sounds authoritative and judicious, but simultaneously allows a writer to avoid directly stating unpalatable or expensive truths. Every bureaucrat has learned to avoid taking responsibility not only for decisions but for the facts on which such decisions are based. Uncertainty leads to caution, abstraction and indirectness. If these two conditions are facts of bureaucratic life, of whatever industrial, commercial or governmental origin, then to persuade writers to write in clear, concise, and direct language is to ask them not just to change their habits of writing but, at least in their minds, perhaps, to risk their professional position.

Furthermore, we could find ourselves in exactly the situation I urged earlier: Just as we are perhaps wrong to insist that faulty parallelism and so on are mistakes if they do not elicit unfavorable responses in casual readers, so would it be a mistake to argue that a bureaucratic style is wrong, simply because it offends our sensibilities. But there is a difference: One of the problems with a bureaucratic style is that it resists easy reading. Often, it even resists strenuous reading. In virtually all matters of usage, the principle of clarity is rarely if ever invoked. *Data* as a singular, *irregardless* as a connector, *less* modifying count nouns—not one of them is obscure or ambiguous. But a sentence like this is virtually impenetrable:

> There is now no effective mechanism for introducing into the initiation and development stages of reporting requirements information on existing reporting and guidance on how to minimize burden associations with new requirements.

But one more inversion: From the bureaucrat's point of view, an opaque style is good, difficulty in understanding is good, confused meaning is good. Or is it? Is it really a bad habit that once corrected will give way to the concise style of an E.B. White?

We hope that at the end of our project with the government agency, we will know. What we know now is that we know very little; what we do know raises more problems than it resolves. One of the problems these considerations raise is that our understanding of good and bad, right and wrong, effective and ineffective may not be as straightforward as most rhetoric texts make them out to be.

17

Edward M. White

MASS TESTING OF INDIVIDUAL WRITING: THE CALIFORNIA MODEL

The most useful materials to come out of the institutional testing experience so far are the annual reports describing the California State University and Colleges (CSUC) English Equivalency Examination.[1] Although they detail the administration of a program for granting credit for freshman composition to students who have not taken the course, the procedures for evaluation can be applied to placement or proficiency testing at any level of competence. Because we think the reports can serve as valuable working papers for teachers and administrators charged with the responsibility of developing programs for evaluating writing, we are reprinting parts of them here. As space does not permit us to reprint all the material we consider valuable, we must refer our readers to the reports themselves to gain a full appreciation of the care taken in constructing this program.

With a rich source to select from, we have chosen to present material which can serve as a guide to the art of essay test design.[2] What follow are remarks on essay testing made by Edward M. White at the 1976 MLA

Edward M. White is Director of the English Equivalency Examination and Coordinator of English Testing Programs, California State University and Colleges.

1. Edited by Edward M. White, the reports are titled *Comparison and Contrast* and dated 1973, 1974, 1975, 1976, and 1977. They are available for purchase from the Office of the Chancellor, The California State University and Colleges, 400 Golden Shore Blvd., Long Beach, California, at $1.50 per copy.

2. There are sections that do not appear here but which we think might be of particular interest. In "Four Issues for Faculty in Equivalency Testing" (1975, pp. 101-111), Edward M. White responds to faculty concern about the testing program. In "Data Processing Procedures" (1975, pp. 37-48), Robert Bradley details the CSUC methods for handling the practical and logistical problems of scoring over 3,000 examinations. "The Working of a Controlled Essay Reading" (1976, pp. 68-75), assembled by Jan Green and Gae Goodrich, serves as a guide for those colleges and departments interested in using controlled essay readings for staff final examinations, placement or proficiency testing, self-study or research. Finally, each year's section on developing essay test norm samples, read sequentially, can give a sense of how a carefully planned large scale testing program can persuade teachers an administrators to look thoughtfully at instructional goals.

18

convention (A); an account assembled by Journal of Basic Writing *editors of the procedures followed to develop essay questions (B); the 1974 essay scales and accompanying writing samples (C); and the questions and scale used in the 1975 test (D). Although many of our readers are no doubt familiar with several essay ranking scales, we think they will be interested in reading the CSUC scale and writing samples. It is the only scale we know of which describes levels of writing samples composed in response to a single, carefully designed, college-level task.*

A. PRINCIPLES OF TESTING

The CSUC English Equivalency Examination program is designed to offer entering college students the opportunity to gain college English credit by examination. It is one of the very few such programs in the country directly controlled by English faculty; and we gained control and retain control of the program by developing (with, I should add, the continuing good faith and assistance of Educational Testing Service (ETS)) a considerable amount of expertness in essay testing.

To ensure careful testing we develop questions with great care, screen them through faculty committees and pre-test them with students. Pre-testing is essential for a good essay test. The basic principle here is that a test needs to be tested itself, and needs to demonstrate it is ranking students according to its declared criteria, to ensure what test specialists call validity. We need to demonstrate that a writing test in fact discriminates among students according to writing ability, and we need to define the particular kind of writing ability a particular question is designed to measure.

We also need to be aware of the desireability of giving assignments that are clear and whose scoring criteria are relatively apparent. Pre-testing will usually reveal problems in clarity, but careful question writing calls for consideration for the student writing the test. Far too often we compose questions that are vague and confused, in the hope or expectation that we are freeing students to write as they wish.

Another very common practice that damages the validity of essay testing is to give students a choice of topics. Most of us feel that we are helping students if we allow them to choose either topic A or topic B. We are wrong. There is no evidence to show that, when given such a choice, students will choose the topic on which they will do best; and giving the

19

choice we lower the possibility of fair grading. After all, one of the topics is sure to be easier than the other, and we will be assigning grades in part on the luck of the choice rather than on writing ability, or whatever else we think we are testing.

In addition to pre-testing, clarity in the question, and a single question for all students, I'd like to add the need to give more than one essay, if we are to achieve a ranking of students that will reflect their ability.

The ETS Advanced Placement Program essay readings are the basic source of this approach to scoring essays; controlled scoring sessions contradict the notion that it is impossible to reach agreement on the value of a piece of student writing. If we want to give fair, reliable tests, we need to be scrupulous in developing questions, and in conducting controlled essay readings. Only if we do so, will we be able to claim that we can test writing as effectively as, or more effectively than, the multiple-choice tests.

B. FORMULATING THE TEST [3]

For the CSUC testing program, students were required to write two forty-five minute essays, the first informal and personal, the second a comparison and contrast response to two literary passages, and to take the Analysis and Interpretation of Literature *examination which was developed by ETS and is scored by computer.*

Procedures for developing the essay question have followed the same pattern over the years. English department members on different CSUC campuses are invited to serve on one of two question committees. One committee develops a question to test the student's ability to move from description to abstraction. The other committee develops a question which asks for a comparison and contrast of two short passages in order to examine the student's ability to respond incisively to others' ideas. Committee members bring sample questions to the first meeting. The questions are discussed and the committee agrees on one or two to be pre-tested. The professors then assign the sample questions to freshman composition classes and examine the essays to see if the question has elicited a range of responses that indicates different levels of student abilities. (The step of pre-testing has proven invaluable for separating

---

3. This account was assembled by *Journal of Basic Writing* editors using as a guide the reports of question committee leaders which are included in each issue of *Comparison and Contrast.*

those questions which may serve as valuable teaching devices from those which serve as discriminating testing devices.) At the next committee meeting, members rephrase the question if necessary and develop directions for scoring the essays. When the examination has been given, the members of the question committees take on the responsibility of training and supervising those who score the essays.

C. CSUC ESSAY SCALE, 1974 [4]

Question 1. The following question was distributed to all students:

Think of a personal experience that has in some way changed your life, either for better or worse: a particular event, a person, a place you have visited, a book you have read. Describe the experience in detail and explain fully why it was an important one for you.

The following directions for scoring were distributed to all readers engaged in the grading of question 1:

The student is asked to write about a personal experience that has in some way changed his life, either for better or worse: a particular event, a person, a place, a book. He is specifically asked to 1) describe the experience in detail, and 2) explain fully why it was an important one. The student should be rewarded for what he does well in his response to the assignment. Papers should be scored for their *overall* quality.

An extremely well-written response may be scored a point higher than it would on the basis of content alone.

A poorly written response may be scored a point lower.

Errors in spelling and punctuation which occur in writing a draft under examination conditions should not ordinarily be counted against the score.

NOTE: Since the student is asked to write about a *personal* experience and its importance to him, a wide range of individual choices and attitudes must be allowed for. Answers should therefore not be penalized simply because the writer may regard even his most important experience as relatively insignificant, because he seeks to provide a philosophical perspective, or because he views the experience in humorous or satirical fashion. Imaginative responses should be recognized and rewarded, as distinct from 'cop-outs.'

---

4. Question 1 and directions for scoring can be found on pp. 16-18. Writing samples for question 1 can be found on pp. 21-27. Question 2 and directions for scoring can be found on pp. 27-30. Writing samples for question 2 can be found on pp. 31-39, *Comparison and Contrast,* 1974.

Possible Scores:

6 A superior response will be a well-organized essay that does the two things asked for in the assignment. It will *describe* an experience in sufficient *detail* to make it distinctive, and it will explain the *importance* of the experience. An essay getting a score of six will show a high degree of competence generally, though it may have minor imperfections.

5-4 These scores apply to responses that deal with the two tasks specified in the assignment less thoroughly than the essays scoring 6. The description may be somewhat general or abstract, and the explanation more implicit than explicit. However, essays in this group should have an effective, logical order and be reasonably free from errors in the conventions of writing.

3-2 Papers in this category respond only partially to the assignment. They may:
   —give adequate attention to one of the specified tasks but little to the other;
   —treat both tasks rather superficially;
   —be lacking in supporting detail;
   —drift away from the topic or display considerable irrelevancy;
   —have serious faults in writing.

1 This score should be given to any response that is on the topic but suggests incompetence.

* Non-response papers and papers that are completely off the topic should be given to the table leader.

The following student responses to Question #1 were sample papers used during the reading to illustrate the grades on the 6-point scale:

SCORE OF ONE

I was becoming rather pessimistic in my view of life in general; Because of the injusticies, corruption, lies, and hypocracy I saw in almost everyone; in school etc. Until I started to think why many of these things were, what circumstances brought them about, and I realized that life is pretty much what you make it. From pessimism I came to believe that everyone wanted to be good but they weren't sure how to do it.

This change in my way of thinking didn't come all at once. It come by gradual perceptions of human behavior; such as why a person should become nervous in a certain situation, or why some people seemed to understand better than others. This experience was brought about mainly by a combination of several events that helped me to think more clearly; and I think another major factor was an atmostphere at home and school of calmness.

The importance of this revelation or understand is manifold. It has shown me a new way of learning to live in this society. I have noticed several times that it has saved me from doing foolish things. I have become more at ease with myself. Frustration doesn't bother me mentally, make me upset; it may make me mad but I understand how to deal with it. In general it has made me a better person, enlightened my life, given me an ambition to live my life the best I can and to be proud of it.

SCORE OF TWO

A persistantly used topic in novels and films is that of the 'art student' in a garret in Europe. Humble, naive, and left out of the mainstream of culture and society, he spends his days mooning through plaza and cathedral. Typically he subsists on a meager income gleaned from selling a small painting or two. Such a romantic existance was the antithesis of my sojourn in Firenze, Italy.

Caught helplessly in a rush from private school to villa, and back, I was a captive of a widowed teacher bent on spending a small grant for the luxury of touring Italy in a new Mercedes-Benz. Culture was deprived from my visual perusal by the constant onslaught of theatre engagements and expensive restaraunts. No, I could not boast of an increased understanding of Verrochio, but only of the finest wines and meats.

Such a fictional account must be the meat of many an essay dealing with milestones in life. Yet I regret to say that I am not able to paint such a lurid tapestry, if only because most of my life is yet ahead of me. I find that when faced with the challenge of recognizing a major catalyst in my existance, I am unable to do so because of several important considerations.

For some event to be meaningful in the necessary contest, surely its aftermath must be multi-decodous in length. At 18, and aware of only the last half of my duration to date, I lack the required insight to appreciate such a remarkable, if not violent, motive force. Oh, I could speculate to the hearts content, but this method falls short of reality.

To guess at the probably longterm outcome of anything short of death or grevious injury is grossly unwise. Understand that any predictions of the future are always clouded by optimism, or perhaps pesimism, but rarely the correct confluence at the hands of the adolescent writer.

In short, careful retrospect and insightful analysis can not be taken from a medium that has not, as yet, had ample opportunity to mellow with experience. One can not stand at the mouth of many tunnels and know what dragons lurk within. Only with the eventual outcome of the drama can I afford to rest and then comprehend the reasons for the structure of the plot. Such a report may indeed be forthcoming in future years.

As a note to the preceding piece, it is not meant to be acid but rather the

only reply imaginable to me, in light of the nature of the question. Perhaps I interpret the meaning too gravely.

SCORE OF THREE

There has been one person in my life that has changed me very significantly. This person, who's name is Leslie, makes me realize just what kind of an individual I am. Before I met her, I had a variety of problems. One very drastic problem that I had was the lack of self-confidence. There was no self-motivation behind me what so ever, and it showed very much. I also had another very serious problem which was the constant thought of death. The idea that we will not *be* any more or more specifically that *I* will not be any more was running through my mind at all times. Still another problem that troubled me was an inferior attitude towards myself. This attitude was not only mental, but in my physical features and abilities also. I was constantly believing that I was very ugly inside and outside, and this feeling developed into deep stages of depression. Depression was so much a part of me that others did not want to be around me. Then I met this person and started talking out each problem with her. She explained to me the seriousness of my problems and allowed me to solve them for myself, just by talking. Leslie said that if *I* had no confidence in myself, *who would* have confidence in me? She also showed me how fun and exciting life could be to the point where I no longer had to think about death. By this time, depression seemed one of the farthest possibilities for me.

Because of Leslie, I am now a changed person. She made life worth living for me, and most of all, she allowed me to understand it all. Now I realize that I am a changed individual. An individual different and unique from all other persons, with attributes that are unique also. This realization has been very important to me. Everything was against me before Leslie helped me understand all of this, but now I have a lot to learn and experience. Now I am even starting to understand and help other people with their problems. This also is very important to me. It raises my self esteem to know that I can be of help or service to another individual in need. To some people the importance of life is not realized unless they find out they are going to die. By this I mean someone who finds out they have a terminal disease. Then, the whole world changes before their very eyes. Each day is lived to it's fullest; like it was the last day of their lives. Everything becomes beautiful and simple. Leslie made me realize that I don't have to think of life as a terminal illness. By understanding myself better, I can live each day with enthusiasm just like it was the last day of my life.

This feeling alone that have described explains the importance of Leslie being a very significant change in my life.

SCORE OF FOUR

When I was approximately ten years old, I joined a synchronized swimming team called the San Francisco Merionettes. I heard about the team from a friend of mine who's sister had been on the team for about five years. For the first few years we trained on Tuesdays and Saturdays from 5 p.m. until 7 p.m. As the years went by and I advanced within the team I began to train more often during the week—three days, four days, five, six, and finally seven days a week. The club was divided into smaller teams ranging from the "A" team (the best) down to the "G" or "H" team. It took me seven years to reach the "A" team but it was well worth all of the time and effort spent to reach this goal. As a member of the Amateur Athletic Union I competed in many meets here in San Francisco and also in other cities throughout the Bay Area. My coach, Marion Kane; was known as one of the best in the business so I frequently placed in the top three and received medals for my accomplishments. I also made many lasting friendships not only with the girls on my team who I trained so often with but also with girls from other teams who I competed against. Though the competition was tough there was always a friendly atmosphere at the swim meets and it was a good chance to make new friends from other cities. For some meets we would travel to other states such as Ohio, Pennsylvania, and Texas to compete. This was usually a National Meet where girls from all over the United States met to compete in the sport of synchronized swimming. National meets occur twice a year; usually in April and in July. Synchronized swimming is different from speed swimming because you swim to music in a team of eight, a duet of two, or a solo routine. The routines take from three to five minutes each and are prepared and practiced for months ahead. When your routine is completed you are judged by seven judges on a scale from one to ten and when the scores are calculated, you are ranked from first on down. The competition was sometimes very close which made it exciting and suspenseful when the medals were being awarded. The travel was always exciting and fun and it was a good chance to see new and different places. In the summer of 1973 a team of ten girls from our team toured Europe for three weeks to help the European teams and also to do demonstrations. I was included in this team and had a very fun and interesting trip. Now, our coach has retired so I no longer swim, but I have the friends and the memories from the seven-year experience. Swimming was a very big part of my life during those years. I enjoyed the daily exercise and the idea of getting out and working with a group headed towards a major goal rather than eating and watching television every day after school. I feel the discipline did me alot of good, also, physically as well as mentally. I feel I am a much more rounded person from all of the travel and the meeting of people from different areas. All in all, my membership in the Merionettes

had a lot to do with the person that I am today and I am very glad that my young friend got me interested in the team. I got alot out of it by putting alot into it and I feel that if I had it to do over, I certainly would.

SCORE OF FIVE

"Oh - Amerika!" I remember her face scrubbed clean in the sub-zero air like an old apple someone had picked up and polished. Her eyes sparkled against the snowy forest as she looked up at me, bent now but still so alive.

We were walking through the snow drifts to the chapel; I was nineteen, she was ninety-three; and we were both just a little suspicious of why the other was there, the Russian border only ten miles away.

Of course, she had more reason to be there than I. She was Russian. It showed in the way she tilted her hat to the right, as all Cossacks did before the revolution struck; it showed in her thick, gnarled hands that helped in the fields for ninety years before she decided to slow down and just work in the barn; and it showed in her eyes as she looked at me then.

"Amerika."

What could I tell her? How could I tell her why I had come? Was Russia as different as we were told it was? Russia, now the USSR and 'Amerika,' now the U.S.—were they that different? Was there some kind of mutation in the human race that made our ideologies so diametrically opposed?

I looked at the wizened, ancient woman as if she could give me an answer, but I spoke only Finnish and my companions teeth had either fallen out or served her so badly that I could barely understand her as she spoke. She smelled of the barn. Of green hay and warm milk and geraniums-in-the-window, in an old patched coat that seemed to bury her —but not her eyes. She was looking past the forest at the sun as it began to rise and bathe the sky in velvet;

"It's beautiful," I said as my eyes followed hers.

"Herosheni," She said.

"What?" She smiled at my question.

"I am too old, and my Finnish is very bad, but it makes no difference. Everytime we see the sun rise, I will say 'Herosheni,' and you will say whatever it is you say in Amerika, and it will make no difference. We will be as one." She smiled as she walked ahead of me.

"Herosheni."

"I have never found a dictionary that could define that word."

SCORE OF SIX

Sometimes, people are not able to mature properly; others get a head start early in life. A trip I went on to Canada with my father at age eleven

changed my whole outlook on life, from that of a frolicking boy to a serious young man.

We had gone mountainclimbing for several years, and our experience was extensive. My father and I both belonged to the Sierra Club, and had participated in many of its' events together. But we had grown tired of climbing in the High Sierras. We wanted to climb in another country, on an expedition. At one particular club meeting, we picked up a brochure put out by Mountain Travel, an expedition organizing corporation. We read through it, and found a trip to Canada to our liking. The first requirement, we learned, was to send resumes of yourself and your experience. Needless to say, the expedition leader was astonished at my amount of experience in the mountains, on all types of terrain. My age cast some doubt on my eligibility, but the amount of experience I had compiled more than made up for it. After several months of planning, buying, assembling, and packing, we were ready to go. The first stage was to fly to Vancouver. Everything from then on was left to the organization and its' leaders.

We left Vancouver in an Amphibious aircraft headed for Mimpo Lake. This was to be our base, from where we were to attempt to explore the Monarch Icecap region of British Columbia, and conquer several of its glorious peaks. Every other day it rained, so we did not get much done for the first week or so. The leader, Gary, was not much to my father's liking. He was a very immature man of about thirty. We attempted two or three minor peaks, with him leading, and he would just walk off and leave the group to catch up or get lost. His wife was on the expedition also, and once, when we stopped at the base of a large glacier to put on crampons, he just left his wife behind, still struggling with her crampon straps. My father was not pleased with our leader's conduct; to say the least, he was furious. There we were, 200 miles from the nearest city, fifty miles from a farm or cow pasture, and we were stuck for two more weeks with a man who might just walk off and leave his own wife to die. My father did not like the situation, and the friction between him and Gary increased. The other members of the expedition were also aware of Gary's immaturity, but what could we do?

One night, in our tent, my father and I decided we were going to leave on our own. We were not having any fun, and we were being herded about like cattle by the leader. We packed up everything we would need, and, at about midnight, we set of down the glacier towards Bella Coola, where we could get a plane home. Walking in the dark, with only starlight to guide us, we worked our way through the maze of glaciers that could swallow a man before he could shout. The deep crevasses all around, some 300 feet or more to the bottom, loomed toward us like hugh abyssas. In the light of early dawn, we were almost down off the glacier when my father slipped and broke his ankles. I was horror stricken! What could we do, out in the

27

middle of a glacier, my father who could not walk and myself, an eleven year old boy. We worked to erect a sort of shelter, my father directing and me lifting or tugging. When he was safe and warm, I set off alone, back toward camp to get the others to come and help us.

It took me a day and one-half to get back and get help, winding my way around gigantic crevasses, over huge snowbridges that threatened to give way beneath me, over cliffs so slimy my boots would not stay put. It was a miracle I got back alive. I enlisted the help of the other climbers, and we used the camp radio to call a helicopter to get my father and I back to civilization.

I think this event in my life was an important one because it changed my whole outlook towards people, especially adults. It made me realize that there are dumb, incompetent people at every age level, and that people like my father were very extradinary indeed. I had previously thought that all adults were like my father; calm, mature, collected, not like little children. That talk with my father the night we left camp was very enlightening. He showed me how Gary had been a very mean, immature man, self-centered and unreliable. I came to realize that many people never really grow up, but die as immature as young children.

I also learned to take on the responsibility of an adult. I saw what had to be done, and I faced the crisis head on, instead of crying or turning away from it. I feel that all the temporary grief this incident caused was nothing compared to the changes in my character that were brought about by this calamity.

Question 2. The following question was distributed to all students:

A. 'If a society is to strive with any hope of success toward peace and prosperity in a commonwealth, the authority governing that society must not only be able to pass laws and to reassess those laws constantly as circumstances change. . ., it must also be enabled to enforce those laws and to exact penalties for their violation.'

B. 'Under a government that imprisons any unjustly, the true place for a just man is also in prison.'

Assignment: Write an essay on the two passages above in which you answer the following:
In what ways are these statements alike and in what ways do they differ?
What strong or weak points does each position have?
To what extent might a person accept both positions?

The following directions for scoring were distributed to all readers engaged in the grading of question 2:

"The student is asked to write an essay in which he explains 1) in what ways the two statements are alike and in what ways they differ; 2) what strong or weak points each position has; and 3) to what extent a person might accept both positions. He should be rewarded for what he does well in his response to the assignment. Papers should be scored for their *overall* quality.

An extremely well-written response may be scored a point higher than it would be scored on the basis of content alone.

A poorly written response may be scored a point lower.

Spelling errors should not ordinarily be counted against the score.

Possible Scores:

6 A superior response will be a well-organized essay that does the three things asked for in the assignment. It will compare and contrast the meanings of the two statements. It may explain the meanings by *means* of comparison and contrast, or it may explain the meanings *and* compare and contrast them. The best essays will note that while the quotations both say something about government and laws, the first asserts the need for law and order and takes the point of view of the state, while the second affirms the principle of justice as superior to the laws of the state when those laws are unjust, and it is written from the perspective of the individual. The best essays will show conscious-ness of the possible dangers inherent to the first quotation (that is, that it could mean that even unjust laws should be enforced, that it says nothing about individual rights, that it emphasizes punishment and authority rather than freedom); and the most perceptive may perceive dangers in an uncompromising position on the second passage. The best papers may show an awareness that the two positions, properly qualified, can both be accepted. An essay getting a score of six will show a high degree of competence generally, though it may have minor imperfections.

5-4 These scores apply to responses that concentrate more on one quotation than the other, or that deal with both subjects somewhat less thoroughly than the essays scoring 6. Essays in this group may have minor errors in writing.

3-2 Papers in this category deal with both quotations but may:
—be lacking in supporting details, or treat both quotations superficially
—give adequate attention to one but too little to the other;
—fail to see that both are concerned with laws and the state but that there are important differences between them;
—misunderstand or misinterpret the meaning of either or both;
—be *primarily* critical or argumentative;

—have serious faults in writing;

—drafit away from the topics or display considerable irrelevancy.

1 This score should be given to any response that is on the toppic but suggests imcompetence.

* Non-response papers and papers that are completely off the topic should be given to the table leader.

The following student responses to Question #2 were sample papers used during the reading to illustrate the grades on the 6-point scale:

SCORE OF ONE

These statements have little in common except that they both talk of justice and penalties within society. "A" speaks of keeping up with the times while "B" speaks of justice for those accused of violating laws.

Both have something important to say, and deliver it with a certain amount of impact. I say "B" has much more impact than "A", because it's statement is made with one short (down to the bone) sentence, stripping it to the raw unclutter *point!* Which "A" trips out on

If society is to strive—blabber—," "A" gives an introduction to its statement which I feel isn't necessary, leaving me with the feeling that its more story than statement.

I would accept both positions to any extent. B is a little extream in its message but thats what gives it its impact. While "A"s position is one of a lot of peoples, I'm sure. Its a safe general statement of fitting penalties to the present day society. They are both reasonable, and complement eachother nicely.

SCORE OF TWO

The two passages are quite different from each other. Although they are both of the opinionated form, the second is much more poetic than the first. The first one states a warning or a set of instructions on which one might form a constitution. The second, on the other hand, gives a form of philosophy.

The first one is quite explicit in that it sets the goals and what must be done to meet them. I says that peace and prosperity are what you're striving for and the only way is through flexability in government.

The second is harder to understand. It says that if you live under a government that imprisons unjustly, a just man should be in prison. I find this hard to agree with. I feel if a just man lives under such a government he should strive to make it just. Another thing which is hard to take is that if all the just men were in prison only the unjust would be left to govern.

A person might accept both positions if he understood the the second is pretty dangerous.

SCORE OF THREE

The two questions are similar in that they deal with the just way to strive for justice in government. Laws are provided that hopefully nobody who is innocent of a crime can be punished. This creates a situation that enables many criminals to escape justice. By the second questions standards it is better than few guilty people are protected so that innocent people are protected also. There must be sufficient evidence to support guilt, leaving no doubt in the judge or jurie's minds of guilt or innocence.

The question arises of what is just or unjust. Who is allowed to set standards for society. Some argue that the majority rules in all cases leaving no allowance for any other possibilities. In many cases however the majority will be the same people and the minority will never be heard from, thus getting the shaft.

In other cases it is the ellect officials which we the public elect into office who create justice, and all that laws are followed. These people however are squeezed into tight limitations because of our Constitution, leaving no possibility for personal involvement in any case.

The Constitution creates another loophole in that it creates different powers, and leaves Congress open to decide what is meant by parts of the Constitution. They can interpret it a number of ways, changing it for individual cases.

A major weakness in the first statement is that it does not set limitations on law enforcement. There is a limit to how much power any one particular organization should have and ones own individual rights as written in the "Bill of Rights." Where does law enforcement end and 1984 begin. Do we want a police state, with no regard for personal freedom. If this were to occur the second question could likely be draft, with both just and unjust persons being the victims.

At the same time total anarchy with no rules or regulations would create total chaos, with everybody attempting to beat out his competitor. In creating laws you try to establish what will be the best good for the most amount of people, without leaving any individual out.

The system we live in creates a sense of competition, in which money is the eventual end goal. In many cases people are placed into roles of superior inferior, with the inferior having to prove himself to rise to the higher plateau. There are often obstacles which obstruct and impede this persons progress, which results in extreme measures by that individual to survive in society. He is left little option but to committ a crime under governmental laws. Is it fair that this person was put into the situation

31

where there was *no* alternative. Does 'fairness,' even enter into the picture as a possible motivating factor. That is a question which is often dealt with, but with no satisfactory answer for everyone. The question arises of everybody having equal opportunity in our society, but is that always the case or is that a non-reality.

The first question makes the statement, 'reassess those laws constantly as circumstances change,' which outwardly seems fine. Everything no matter what it is should be open to change, but is that change occuring fast enough. In many cases the statement is made that we are changing, but we can't do everything overnight. Is this an exaggeration by these people or are they justified in this comment. In some cases they do change, but in others they don't, but not everybody wants these changes to occur, so on the whole it would seen successful, but what about the time lag between a proposed change and the actual writing of it into law. In many cases it becomes obsolete, and has a negative reaction by all.

It is hard to please everybody, but the major thing which should be strived for it justice, even in one form or another, making a strong attempt to please everybody.

SCORE OF FOUR

Statement one, taken for itself, has many strong points but it is not entirely without fault. In any orderly society, there must be laws, and they must be enforced, so as to insure greater peace and protection for all. However, law must be not so terribly strict as to imprison a man unjustly. The justice of these laws must be considered in their reassessing, but, even then, a law should not be totally rigid.

Statement two, dealing with unjust imprisonment, also has strong and weak points. If a man is imprisoned unjustly, it should not be taken as an indictment against the whole system. It is true, however, that a law should be able to be considered differently in different situations. When just men see others imprisoned unjustly, their place should not be 'in prison' with the first, but out trying to do something about unjustness.

In many ways, the statements' basic messages can be both accepted by a person. However, qualifications must be made and neither statement should be accepted as it is. Laws are necessary in society, if it is to flourish, and they must not be ignored. However, in their enforcement, the justness or unjustness to the individual must be equally considered. The key to the reconciling of these two viewpoints is found in this sentence from statement one: authority must 'reassess laws constantly as circumstances change.' If the law is reassessed according to different and changing situations and times, then it also must be considered differently in situations involving different individuals. In this way, it will be insured

that laws, while being enforced, are not unjust in their imprisonment of persons.

SCORE OF FIVE

Statements A and B have both strong and weak points. They are similar in some ways, but different in others. It is actually possible for a person to accept both positions.

Statement A is basically sound government policy. It is true that a society should be able to pass laws and reassess those laws. As times change, the attitudes and needs of the citizens change, and the duty of the legislature is to meet those needs with progressive legislation. A society should also be able to enforce its laws and punish violators. With no executive branch to support the legislature, a society quickly becomes anarchy. Punishment must be administered to violators to rehabilitate them, deter other possible criminals, and protect society from dangerous individuals. These are not all strong points of Statement A. However, statement A does not mention any guarantee of personal rights to the citizens. To insure a democratic society, a constitution outlining these basic rights is a necessity. Without this basic framework, an oppressive government could result.

Statement B, on the other hand, says that in a government which imprisons any unjustly, the true place for a just man is also in prison. The strength of this statement lies in the principle of justice, where no innocent man can be punished for something he did not do. This statement is weak by not offering any solution or system whereby a government could operate efficiently and not risk persecuting the innocent.

Statement A appears to be written from the viewpoint of the head of society, while statement B seems to express the average citizens views. In this way the statements differ. They are alike in that they are both opinions on how a society should operate.

It is not difficult to accept both statements. Obviously, the men who set up our government took both points of view into consideration. Our legislature operates under a basic constitution and continually makes new laws to keep up with our changing society. Our judiciary uses a system whereby a fair trial is guaranteed and those convicted of crimes are punished. We also have policement to enforce the law and protect society from criminals. It is impossible to have a system where all criminals are punished. We also have policemen to enforce the law and protect society not persecute any just men, and still efficiently punish criminals. A good society should have a golden nean, as ours has, where the rights of the individual are protected and criminals are still punished.

We have seen that the two statements offer opposing views of a perfect

society, and that both have their strong points and weak points. By adopting the strong points of each, it is possible to accept both statements to a large extent, much as our society has.

SCORE OF SIX

Statements A and B differ greatly, primarily in their respective outlook upon Society in general. The first Statement is undoubtedly that of a political realist, dealing only in the black-and-white of the extent of governmental authority. The second is obviously the profession of a political moralist, to whom authority is useless if misdirected. In the *4th* Century B.C. a Chinese philosopher named K'ung Fu'tzu, better known as Confucius, stated that government exists for the benefit of the governed, and not visa-versa. Hence, while a government may possess power, it must also dispense justice fairly.

Examine each statement carefully for while both are well-founded, both contain damaging, perhaps damning contradictions of thought. The author of the first can be thought of as being quilty only of political realism and skepticism. All he has done was to put bluntly what every nation's political philosophy has stated indirectly since time immemorial. An ordered state is desirable, therefore laws must be made. If laws are to be made their violators must be punished. This is all simple political philosophy, with all conclusions resting on the basic premise of national survival. Had the author of this statement rested his personal argument upon 'natoinal survival', there would be little to dispute. However, his supposed aim was 'peace and prosperity in a commonwealth.' Under these circumstances, his statement is found to be inadequate. While bills of attainder are suitable for rational survival, genuine peace and prosperity requires a judical check upon legislative authority, a means by which justice can be dispensed in the commonwealth. Note that such a reference is non-existant. The exacting of penalties is left to the governing authority. This authority reaches omnipotency in that it exists and operates without the interaction with any independent power (as a check).

In the second statement a verbal profession of the ideas of Gandhi is seen. Indeed, it is nothing but a restatement of the noble theory that led many Indians to perform acts of civil disobedience (resulting in imprisonment) in order to call attention to widespread injustice. Yet, if a government is so lacking in justice, civil disobedience or non-violent publicity-getting is not adequate. If we are to believe the Confucian concept of government to benefit the governed, then we can conclude, as did the master's student, Mencius, that the people have the right to change their form of government, by whatever means are endemic to that nation's beliefs. Therefore, we see that the just man's confinement (by his own

design) to prison, defeats the just man's purpose in an unjustly governed society.

Although the two statements seem to be different, a man can, with clear conscience, subscribe to both. An omnipotent governmental authority can indeed bring 'peace and prosperity', as long as harsh laws are tempered with even justice, so that a just man need not feel his true place to be in prison.

D. ESSAY QUESTIONS 1975 [5]

Question 1. The following question was distributed to all students:

We are all made up of many selves. Describe some of your various selves— for example, food checker at a supermarket, big brother to a foster child, sole wage-earner in a large family, etc. How different are those selves? What do they have in common?

The following directions for scoring were distributed to all readers engaged in the grading of question 1.

The student should be rewarded for what he does well in response to the question. Here the student is set a three-fold task: to *describe* some of his selves, to *show* how those selves are different, and to *comment* on what they have in common. He is told to think about the question and plan his response.

Note that the question asks for a comparison-contrast commentary beyond mere description, simple autobiography, or generalizations about personality. Responses that do not go beyond such description, autobiography, or generalization should not ordinarily receive scores above 3.

An extremely well-written response may be scored a point higher than it would be scored on the basis of content alone. A poorly written response may be scored a point lower.

---

5. Essay questions and directions for scoring can be found on pp. 21-22 for question 1 and on pp. 31-32 for question 2 in *Comparison and Contrast*, 1975.

6 A superior response will describe two or more selves and state differences and similarities clearly. Though it may have occasional faults, it will be well-organized, well-detailed, and generally well-written.

5-4 These scores will be useful for a well-handled paper which is weak in one or two characteristics of the superior response, i.e., description of the selves, or in demonstration of the differences or the similarities, but is otherwise competently written.

3-2 These scores will be useful for the following kinds of papers:
—those in which only two parts of the three-part question are treated;
—those which treat the subject in superficial or overly generalized fashion;
—those which treat the selves only as moods, opinions, etc., instead of as roles.
—those in which the writing exhibits serious weaknesses in structure, syntax, or diction.

1 This score is to be used for papers which show very little understanding of the question or suggest incompetence in structure, syntax, and diction.

* Non-response papers or those which argue with or avoid the question should be given to the table leader.

Question 2. The following question was distributed to all students.

"My father was killed on Iwo Jima," he said.
"I'm sorry," I said.
"I guess there were good people killed on both sides," he said.
"I think that's true," I said.
"You think there'll be another one?" he said.
"Another what?" I said.
"Another war," he said.
"Yes," I said.
"Me too," he said. "Isn't that hell?"
"You chose the right word," I said.
"Each person does a little something," I said, "and there you are."
He sighed heavily. "It all adds up," he said.

Kurt Vonnegut, Jr., *Mother Night*

To be sure, whoever realizes the senselessness, the hopelessness of this world might well despair, but this despair is not a result of this world.

Rather it is an answer given by an individual to this world. Another answer would be not to despair, would be an individual's decision to endure this world in which we live like Gulliver among the giants.

<div align="center">Friedrich Dürrenmatt, "Problems of The Theatre"</div>

Write an essay in which you explain the views implied in each of these passages. How are these views different and how are they alike?

The following directions for scoring were distributed to all readers engaged in the grading of question 2.

The student is asked to write an essay in which he explains *what* the two statements mean, and *how* they are alike and *how* they differ. He should be rewarded for what he does well in his response to the assignment. Papers should be scored for their *overall* quality.

An extremely well-written response may be scored a point higher than it would be scored on the basis of content alone. A poorly written response may be scored a point lower.

Spelling errors should not ordinarily be counted against the score.

Possible Scores:

6 A superior response will be a well-organized essay that does the three things asked for in the assignment. It will explain briefly the meanings of both quotations and compare and contrast them; it may explain the meanings by *means* of comparison and contrast, or it may explain the meanings *and* compare and contrast them. Essays in this category will interpret the quotations intelligently and coherently, though there may be minor errors in interpretation. An essay receiving a score of six will display a high degree of competence generally but may have slight flaws in writing. An essay getting a six will support generalizations with appropriate details.

5 The essays in this group will concentrate somewhat more on one quotation than on the other or deal less thoroughly with both quotations than essays scoring 6, but they will clearly demonstrate competence.

4 This score will apply to responses that are generally well written but may lack the development of those essays in the 6 and 5 categories, fail to perceive the distinctions and similarities in the passages quite so accurately, or reveal somewhat less facility of expression.

3 Papers in this category may show signs of clear writing but contain misinterpretations of both passages; radically misinterpret one passage; deal superficially with both passages; display some evidence of serious deficiencies in writing; or deal almost entirely with one passage to the exclusion of the other.

2 Essays receiving a score of 2 may start with the assigned topics but drift away from them; be primarily critical or argumentative rather than expository; display considerable irrelevance; fail to see similarities in meaning between the two passages or make distinctions between them; have very serious faults in writing.

1 The response in the 1 paper shows almost no understanding of the question or the passages, *and* the writing strongly suggests incompetence in structure, usage, and idiom.

* Non-response papers and papers that are completely off the topic should be given to the table leader.

Rosemary Hake

WITH NO APOLOGY: TEACHING TO THE TEST

A question on the entrance diagnostic exam at Chicago State University read:

> Should students have to demonstrate certain skills before being allowed to graduate from high school?

Below, Essay Sample One, is the entering student's response.

Rosemary Hake is Director of Basic Writing and The Tutoring Center at Chicago State University.

that the student is ready for the
experience ahead.

Although the knowledge that the
students achieved help them a great
deal, all that can be learned in a
high school is all a person can expect
a student too know. Knowledge of
knowing just how to cope with the
many situations in life is great.
High school graduates have this little
knowledge. It might be the beginning
of the experiences they might find
themselves in, but at least this
is the start. Nothing can qualify
more better, than the experience
of knowing.

Knowledge and the ability to
cope with the many situations in

school at home, just the total environment be which the student is surrounded by. The student has the ability to do anything. Ability should be most emphasized in the high schools ceremonies. Because without achieve ability, which mean the effort, the student cannot achieve in any way. High school graduation should base their requirements around ability. The ability to learn, the ability to cope with problems and many situations that the student as a person have to cope with in life.

Another major requirement that a student should have is that is the ability to face reality with one self. Reality is an experience that is going to happen. Mental reactions to real situa-

tions can yyect the student. He has
to be able to except what lie ahead.
He has to govern his life around
realty. This student have to think
and be able to make decisions — Decisions
that will benefit him in his full
adventure ahead. Therefore, the
requirements for a high school graduation
should be the ability to learn, the
ability to except whats ahead, the
right use of knowledge that one
has, with all ability added together
that high school graduate can
achieve in society.

This is not an isolated example, nor a hypothetical paper. It is a fact of life at urban Chicago State University (CSU), as at many other colleges and universities in the country. The fact, evidenced by the above entering test paper, is that a high proportion of our students arrive unable to write competently. The question—again one confronting many English departments besides ours—is what can be done to insure that such students are not equally poor writers when it comes time for them to

leave the university. The purpose of this article is to describe the program of testing and instruction we have developed over the past several years at Chicago State to see that our students meet at least a minimal standard of competence in writing.

An example of that minimal competence is the following paper, written by the same student under the same circumstances after he had completed English II. To complete the two composition courses took this student one solid year. During that year, he was enrolled in a structured writing class for three terms—namely Composition I twice and Composition II once (See Appendix A for this student's failing and passing Composition I exams, Essay Samples Two and Three). The student also received individual tutoring for one term. A question on his Composition II exit proficiency was:

Should foreign students be required to pass an English proficiency exam?

Below, Essay Sample Four, is the student's response.

> Regrade response _____
>
> The Rights of Foreign Students
>
> America is the land of wealth and opportunity, and indeed, America's education system enables students from all countries to achieve career goals never dreamed of in other nations. Because foreign students must break the language barrier to receive the most benefit from the American educational system, an English proficiency exam would exhibit a clear indication, in most instances, of the foreign students' competency in the English language.
>
> In America today, speaking English is an essential part of every facet of life. Most jobs, with the possible exception of some types of manual

43

labor, require a knowledge of at least conversational English. Persons of foreign origin would have almost difficulty time even arriving at intended destinations if they could not speak some English. The implication of this can certainly be extended to include the foreign college student. Any person that is unable to speak English would be incapable of succeeding in any average American college or university.

The foreign student, if he does not have adequate knowledge of the English language, faces a major problem. However, the American college system, cannot be dramatically changed to fit these students' needs. The system lacks the material an human resources for this.

DP-2694-2/75

44

Therefore, the foreign student must make an effort to learn English.

Many colleges in all parts of the country offer courses in English for the foreign student. This provides an excellent opportunity for the student to learn the fundamentals of English before attempting to study in his intended major field. In this way, the student will be gaining the skills of the English language. He will also be learning about college life. He will then feel more comfortable when enrolling in general college courses.

An English proficiency exam will give insight to the college administrators, as well as to the foreign student. Both parties

the college and the student will benefit if the foreign student has adequate knowledge of the English language

In both instances the student was given a general content question and one hour to write an essay response. The results of all four of this student's exams are listed on a computer outprint (see Appendix B).

Two points about our approach should be mentioned at the outset. One has to do simply with the type of writing on which our efforts are focused. It should be understood that our concern is with that practical type of composition known as the expository essay—the standard form for the written communication of information, whether in a paper by a student, a report by an employee, or an article by a professional writer. The other point is that the key to our instruction in composition is the examination we have developed to measure the competence we aim to produce. Without any apologies, we do, as the saying goes, teach to the test. The reason seems obvious enough: we believe that the processes of teaching and evaluating should function as both cause and effect of one another; the components stressed in evaluating writing skills should influence the teaching of them, and what is stressed in teaching the skills should influence evaluating the student's performance.

In social terms, the function of an institution like Chicago State, and of its required program in basic English composition, is to serve as a means of entry to the middle class. Almost all of our students, about 65% of them black and the rest representative of the other ethnic communities of the south and west sides of the city, come from working class families and are the first generation in those families to enter college. While pursuing their studies almost all are employed at part-time or full-time jobs to support themselves and sometimes their families as well. By and large, the aim of these students in investing the time, effort, and money required to get a college degree is entirely practical: they want to qualify for more satisfying and higher paying jobs than those which they presently hold. Such students, perhaps more than others with less pressing needs and concrete ambitions, make one feel accountable as the instructor in a required writing course and the judge of an essential skill.

The very fact that, in spite of the trend of the past decade, our basic writing courses continue to be required of every undergraduate in the university has afforded us the stimulus and opportunity for scholarly and scientific work in the composition area. In our courses it has been possible to address the thinking and writing skills of all our students and, with the evidence of the essay, observe the result of our efforts to help develop those skills. Instead of having to seek data, we are flooded with them.

What is new in our composition program originated in an experiment conducted several years ago. To get some insight into department

grading practices, we had the staff as a whole read a set of ten essays written for a regular composition course examination. Not much to our surprise, the result was a very wide range—the same essays receiving grades all the way from A or B to D or even F. Such inconsistency among ourselves helped to explain why many of our instructors in second term composition sections had been finding that they had to spend most of their time teaching over again skills supposedly covered in the first term, rather than progressing to new levels of performance. Given the discrepancies in our evaluation of student writing, it also became possible to believe that students entering our required composition courses with serious writing problems might be passing on through them with their deficiencies intact. Clearly we had to seek a collective standard of judgment and try to develop some common strategies to help students meet that standard.

Recognizing that we evaluated essays in different ways, but believing that skill in writing is best demonstrated *in* writing, we set out to design an essay examination format that would account for our differences and provide some of the objectivity in observation and measurement usually lacking in such examinations. Over a two year period of research and experimentation we developed a procedure which provides, among other features, a means for computer readable scoring.

Our first step was to construct an observation framework (see Appendix C) based on the theoretical assumption that a whole generates its parts.[1] This framework, which has been re-worked many times, is designed to help the grader categorize the flaws he observes. Lest the reader consider our observation framework itself flawed because we address vices, not virtues, I should explain that there are both practical and theoretical reasons for constructing it as we have. First of all, it should be understood that what the test aims to discriminate is not fine writing but simply a level of basic competence. The practical point is that while each of the virtues in a piece of writing is virtuous in its own way, the vices or flaws are capable of being classified and counted. The theoretical point concerns the relation of whole and parts, a central concept on which our thinking is based. We assume that if there is a

---

1. Given the thesis that the whole generates its parts, I assume that until the whole has closure the parts lack focus; they are framented, unrelated, incoherent. Once the whole has closure, at a deep structural level, a directive formulating step causes the parts to go through a series of transformations to move to surface representations. I do not identify closure, at the deep structural level, in an absolute sense which indicates a polished finish, a pre-knowledge of all parts, but an "almost finish" where one is aware of enough parts to be able to operate with a directed search as one creates.

question about whether a whole exists, the judge who has some idea of what that whole should contain can clearly identify what it doesn't contain—that is, what parts are missing. In the case of the essay, we assume that a missing part causes a block in the communication. We call this block which the grader observes in the communication a flaw in the essay.

Just as our assumptions about language behavior suggest that in writing there must be a conception of the whole before functioning parts can be generated, so we also assume that teachers or graders have a conceptualization of the essay as a whole; therefore, when reading an essay, they expect an integrated whole with meaningful and logical connections in the essay's paragraphs, sentences, and syntactical and phonological structures. If the concept of the essay as a finished product suggests that the essay is made up of integrated parts, then the whole essay should have harmony among its parts. Because breaks in harmony cause blocks in communication, flaws are likely to be a conspicuous feature for the grader. Hence, the flaws are what we count, and the question becomes one of ordering them in a sound and workable manner. (Given this method, of course, the lower the score, the better the writing.)

In our system the grader reveals explicitly what he has observed. When we have these observations from the grader, we use a mathematical model to transform them into measurement units. These units are then translated into an evaluation.

The observation framework falls into four dimensions: (1) organizational coherence of the essay as a whole, (2) coherence within and between paragraphs and sentences, (3) mechanics and usage, and (4) punctuation. These dimensions (detailed in Appendix C) form a hierarchy. The first dimension provides the most global and formal reference to the essay, the fourth the most specific. There is also a qualitative difference between the first dimension and the other three. The first dimension requires the grader to rate the whole essay with respect to the flaws listed. Since this dimension reflects the essay's structural whole, the flaws are such that they can occur only once. The number of errors possible in dimension one is finite. Dimensions two, three, and four, on the other hand, reflect the essay's functioning parts. The flaws represented in these dimensions may occur repeatedly and are, theoretically, infinite.

The framework is a guideline for the grader; it does not insist that he must find the flaws it lists. He is only expected to note, in the appropriate

dimension, the flaws he observes. Printed along the right margin of the essay paper itself are four columns, each corresponding to one of the four dimensions. The grader records each observed flaw by blackening a space in the appropriate column on the same line as the flaw itself.

The grader is expected to read the paper twice: once for the flaws in dimension one and again for flaws in the other three dimensions. After he has recorded the flaws he has observed, he makes a summary recommendation—for placement if the examination is an entrance examination, for Pass or Fail if it is an exit examination. The advantage of listing specific flaws *and* making a holistic judgment is that by graphing the relationships between judgments and flaw counts we are able, each time the examination is given, to establish (i.e., compute) a minimum competence standard which is based not only on the idea individual graders have of what is proper but on what collectively they have in fact recorded while reading the examinations.[2]

The next step, aimed at strengthening this collective judgment and increasing the reliability of the final results, involves an adjustment of the recorded scores themselves. To insure maximum uniformity, we use the computer and a statistical formula to "calibrate" ourselves as graders. That is, a certain number of flaws, determined by the formula, is added to the scores recorded by reader A, who stands low on a group scale of severity/leniency and a certain number subtracted for reader B, who stands high on that scale. This process, built from the Rasch mathematical model, makes possible the transformation of our observations into measurement units and the translation of these units into an evaluation.

On the basis of the information arrived at through the above procedures a decision (whether in diagnostic or pass/fail terms) is computed. Should an inconsistency emerge, however, between any aspects of the evaluaton (for example, between the grader's summary

---

2. The procedure for determining this standard—or better, the boundary area between clearly passing and clearly failing scores—involves charting the flaw counts of papers recommended for Pass against those of papers recommended for Fail. For reasons already touched on, we, in fact, compute two boundary scores: one for dimension one and one for the functioning dimension; dimensions two, three and four. To illustrate: suppose that 94 papers have been recommended for Pass and 97 for Fail. The graphs which follow show this procedure. The first graph deals with dimension one. You may observe a set of dimension one flaws. The Pass/Fail or minimum competency boundary for dimension one is determined by the point at which the two lines intersect. The result, in this case, is a boundary between papers with two or fewer dimensions one errors, and papers with three or more. The second graph deals with dimensions two, three and four combined.

recommendation and the numerical scores registered, or between the score for dimension one and that for dimensions two, three, and four), the essay is rechecked before judgment is made. Two further protective measures are also built into our procedure when the examination is used as an exit examination. If the student's instructor disagrees with the grader's observations on an examination paper, he can appeal the grade to a review board. On the other hand, if the instructor does not disagree with the grade but feels that the essay was not up to the student's usual level of performance, he can request that the student be permitted to retake the examination.

The assumption reflected in these procedures is that although we vary among ourselves in our observations and judgments, we are all professionals and, within a predictable range, systematic and consistent within ourselves. The aim was to discover our different systems and, by accounting for them, to approach objectivity in grading essay examinations. Rather than being subjected to the unrealistic and oppressive requirement that each see and judge exactly the same things in exactly the same way as others, we are free to be ourselves as we grade and enabled to learn from our differences as we examine them in retrospect.

Having touched on various features of the examination format and grading procedures, I should mention a point having to do with the subject matter of the essay. My assumption here is that an individual can only appear competent if he is familiar—or at least thinks he is familiar—with the subject about which he is writing. To address this problem, CSU publishes, at the beginning of each term, five possible content areas from which we choose two for the topics on each exam we give. Because all of the students, those enrolled in our classes, those being tutored, and those transferring into the school, have a whole term to become familiar with the general content area, no student is forced to write on a topic of which he is totally ignorant. The five subject areas for this past fall term were:

College Degrees
Child Adoption
Medical Practice
Government vs Community Control of Schools
Retirement

The specific topics within these subject areas are formulated as questions. The topic for Composition I is a What or How question:

"What are some characteristics of people working for college degrees?" We assume that the question itself provides the basic outline for the essay: the student will provide a 250 to 300 word answer with a series of examples or illustrations to support his proposition. The Composition II question is in the form of an inverted subject and predicate stated in the subjunctive: "Should there be a mandatory retirement age?" Our expectation here is that the student, responding either affirmatively or negatively, will not only generate a thesis to be both illustrated and explained. He will have to provide the rational relationship between his illustrations and his thesis. This paper should be 450 to 500 words long. (We are still actively researching this area in hopes of greater refinement in formulating topics designed to elicit particular essay responses.)

With the examination format as our control, our program in basic composition is designed to serve every undergraduate in the university. There are three variants of the examination, all with the same format:

1. the *Entrance/Diagnostic Examination*—required of all beginning composition students and administered to:
   a. all entering freshmen,
   b. all entering transfer students with 0 hours of composition;
2. the *Exit/Composition I Examination*—required for entry into Composition II and administered to:
   a. all Composition I students, as the course examination which must be passed in order to get credit for Composition I,
   b. all entering transfer students with one course in composition, as the examination which must be passed in order to have prior composition credit recognized;
3. the *Exit/Proficiency Examination*— required for admission as a major into any university degree program and administered to:
   a. all Composition II students, as the course examination which must be passed in order to get credit for Composition II,
   b. all entering transfer students with two courses in composition, as the examination which must be passed in order to have prior composition credit recognized,
   c. all students in English 222, a group tutoring course designed principally for transfer students who have failed the Qualifying Examination at entry.

The strength of our examination design is twofold: the student knows the content he will write about and how he will be judged, and the faculty can develop teaching strategies to meet a defined goal. And since the examinations serve both as entry to and exit from our composition

program, they provide a comparative measure of a student's ability before he takes a composition course and after he has completed it.

As a diagnostic exam at entrance, the test enables us to place students in classes best suited to their particular requirements. We have designed three distinct versions of Composition I:

Category 3: those who indicate inability to conceptualize a whole, or create a whole essay, and who therefore cannot generate its parts or see relationships among them. These students will usually also need help with mechanics and usage but will not be ready to function with usage level drills. Prior to anything else, they need work in analysis and organization of total compositions, probably beginning with exercises aimed at recognition and analysis of wholes outside the medium of writing altogether and progressing to parallel processes of composition in the essay, starting with the simple narrative.

Typical Category 3 student score on Diagnostic Exam:

Dimension 1: = 6 errors   Dimension 2, 3, & 4 = 24 errors
(See Essay Sample One)

Category 2: those who can almost create a whole composition but don't relate all of the parts to the whole. These students have surface level problems and some problems with meaning, but they can deal with meaning relationships and hence can identify surface relationships and create new structural relationships. They can also investigate simple essay forms to create essay patterns or use simple essay patterns to organize their own essays and then compare the process of organizing the essay to the process of composing paragraphs and sentences. They can also identify differences in the surface representation of different essay forms or sentence parts.

Typical Category 2 student score on Diagnostic Exam:

Dimension 1: 3 errors   Dimensions 2, 3, & 4: 21 errors

Category 1: those who are competent or nearly competent in the simple essay form. These students may have some usage and mechanical problems and some organization problems, but they are what we would consider the typical entering freshman. They should be able to correct surface writing problems if given clear structural definitions or guidelines. They should investigate essays to see how ideas are developed by rhetorical patterns and then be able to expand the patterns. They should also identify simple sentence and paragraph structures and be led to compose more complex structures.

Typical Category 1 student score on Diagnostic Exam:

Dimension 1: 1 error   Dimensions 2, 3, & 4: 19 errors

There is also a class of students who are beyond simple competence. Probably using a thematic text, they should be able to explore ideas, work with the variety of ways (and the strategic reasons for them) of making similar statements, and/or embark on an analysis of styles aimed at developing style(s) of their own.

Obviously, one result of our Composition I "tracking" is that students in the more remedial classes, though exposed to some elements common to all sections, are less likely than the others to meet the requirements for passing the course the first time through. An Incomplete grade is entered for a student who has faithfully done the work in any basic composition course but failed the course examination. A student with an Incomplete may either be assigned to a tutor for assistance in preparing to re-take the examination or, more normally, be directed to re-enroll in the course, usually in a more advanced category.

A further result of this procedure, we hope, is an increased likelihood that both students and instructors will feel a sense of progress: returning, if necessary, to first principles in our teaching, we start somewhere and go somewhere. One evidence that such progress is occurring is the avowal by various Composition II instructors of their increasing ability to build on skills developed in Composition I rather than having simply to repeat the effort to develop them in the first place. The student, on the other hand, enabled to analyze and compare his own respective performances, can chart his progress not only by comparison with his peers or with a national norm which does not necessarily reflect him but by a comparison with himself. He is able to see, for example, that even though he has not yet met the Composition I passing score of 2-15, he is better with his current score of 0-24 than he was with his entering score of 6-26. And when he sees that he is making progress, he may be encouraged to feel that he can, with the help of his instructor, find some system in his problems that will help him to solve them systematically.

We have tried to define as precisely as possible for ourselves and for the students the skills we expect to be developed in basic composition. The general objective, wholly unoriginal, is that students who have completed both Composition I and Composition II should be able to write coherent, unified, and organized expository essays free from serious mechanical errors. Though some components of the course work have more ambitious designs, it is this sound and modest standard that a student must meet to pass his composition requirement. The essay model toward which we teach follows:

THE EXPOSITORY ESSAY HAS:
I.   AN INTRODUCTION WHICH
    A. has a stated or implied proposition; that is, a statement with which the reader may agree or disagree (what)
    B. includes one or both of the following
        1. the proposition placed in an overall context (why)
        2. an implication of how the proposition is to be developed (how)
II. A BODY WHICH
    A. is logically organized
    B. has statements which are
        1. relevant to the proposition
        2. relevant to one another
        3. developed with specific details
III. A CONCLUSION WHICH
    A. restates (not repeats) the generating proposition
    B. does not have information irrelevant or contradictory to the introduction or the body.

This model generates an essay of at least four well developed paragraphs: e.g. Paragraph one: Introduction, Paragraphs two and three: Body, and Paragraph four: Conclusion. It can, and usually does, accomodate more than the simplest four-paragraph design; it can also accommodate an expansion of this deductive model for the more sophisticated inductive model: An introduction which only implies the proposition; a body developed with an analogy; a conclusion which finally and definitely states (not restates) the proposition.

We have found these distinct advantages in teaching to our test:

1. It provides detailed feedback understandable to the student.
2. It provides the student with scores which demonstrate his progress.
3. It can diagnose specific problems so that we can identify both the remedial and non-remedial students and develop a system to place them in classes which will address their needs.
4. It generates data about our students which we can use for further research.
5. It does not impose an arbitrary outside set of standards on the grader; it makes the grader's particular observations and judgment central to the process of evaluation.
6. It allows the development of a departmental consensus on grading standards, a consensus based on the practical working judgment of the graders.

Despite the humanistic antipathy generated by competency based learning, we have found it productive to teach to our test. Even though a testing format can be limited by the skill and imagination of those who employ it, it provides us with a rigor too often lacking in composition programs. The approach is mainly a refined self-conscious application of what most good teachers have practiced willy-nilly. By systematizing the criteria for students and calibrating graders, we have systematized our common sense. The difference between willy-nilly common sense and a systematized approach lies in stating our expectations about what we are to teach and what students are to learn and then devising a means to measure the performance of those expectations.

As a response to humanistic antipathy we can only ask questions: If there are valuable writing performances which cannot be defined and therefore measured, should we not still insist upon identifying and measuring those that can be and finding better ways to teach them? As we isolate performances which resist precise statement and measurement, may we not, even so, find better ways to state, measure, and teach them? Attempting to answer these questions helps us do another thing universities are supposed to do—research.

General Note: Readers interested in the theories of language and mathematics involved in this procedure may consult Rosemary Hake and David Andrich: *The Ubiquitous Essay: A Discourse and Psychometric Model to Identify, Measure, Evaluate and Teach Essay Writing Ability,* 1975 (unpublished research monograph). Copies are available from Chicago State University or the University of Western Australia.
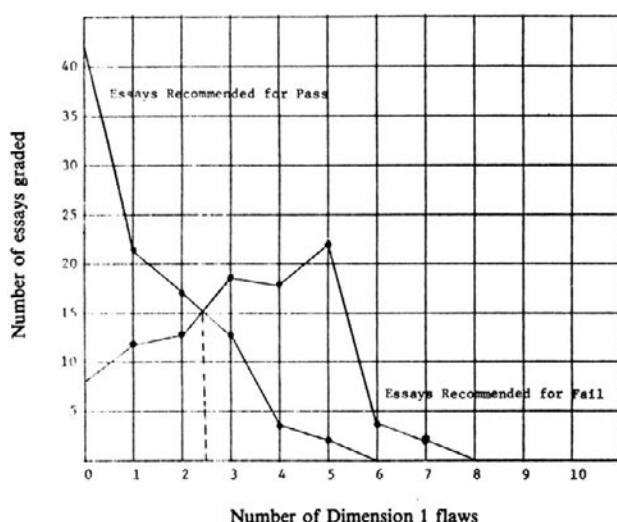
Figure 1: Graph of method for computing allowable number of Dim. 1 flaws

The same procedure is followed to determine the error limits in the functioning dimensions (two, three, and four counted together). If the intersection occurs, say, at 13 errors, the boundary with respect to dimensions two, three, and four will be between papers with 12 or fewer errors and papers with 14 or more.

Number of D2 + D3 + D4 flaws
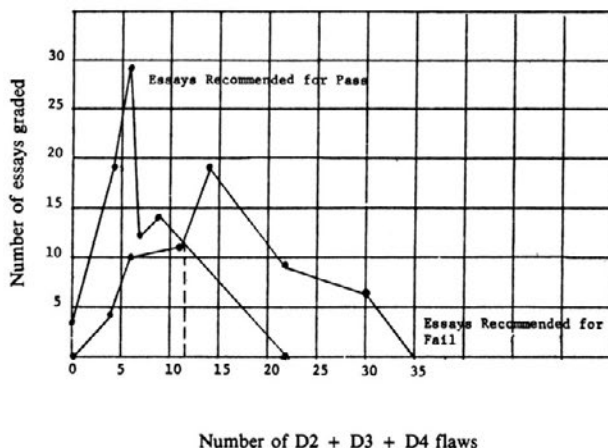
Figure 2: Graph of method for computing allowable number of functioning dimension flaws

EXIT: COMPOSITION I EXAM—ESSAY SAMPLE TWO

Pornography

There would be several effects if a pornography shop was in a neighborhood. First, if a pornography shop was in a neighborhood it would it would lower the creditials of the people in the neighborhood. Second, it would be harder for the parents to control their children. third, if a pornography shop was in a neighborhood some people might not live there.

If a pornography shop was in a nieghborhood it would lower the creditials of the people living in the neighborhood. This neighborhood could have been a very good neighborhood, and the people living in it might have been very respectable people, but in the presence of a pornography shop would change that right quick. Other neighborhoods would get a wrong impression about there people now, and wouldn't associate with them. The presence of a pornography shop in a neighborhood would effect their social standing, who would want them to elect anyone for office. They couldn't run for anything and get elected because the pornography shop would be the cause of it. If there people don't get together and set there pornography shop out of their neighborhoods, they would never get by in good social standings with other neighborhoods

The presence of a pornography shop in the neighborhood would make it hard for the parents to keep their children away from there. The children would go wild knowing theres a pornography shop just around the corner. They would be hanging around the shop all day tring to see what they can see, or waiting for someone old enough to buy a pornographic book so they can get him to buy one for them. Now what can the parents say to there children when they find them hanging around the pornography shop. They may be upset, they get angry, they may even but the children on punishment, but this won't stop most children especially if they think they can get away with it.

The presence of a pornography shop in the neighborhood would upset so many people that they may not want to live there. When people invest in a home, they usually invest in one they find they like or in a nice

neighborhood. The presence of a pornography shop would upset them so, but they're willing to fight it first. If all they tried failed, these people will move out. I mean who wants to live in a neighborhood with a pornography shop right on the same block. These people respect themselves enough not to be humiliated by a place like that.

Pornography is not dirty but there is a certain place they could be sold, and a neighborhood is not the place.

EXIT: COMPOSITION I EXAM—ESSAY SAMPLE THREE

CSU An Asset To Society

When I first arrived at CSU, I looked over the listing of prospective majors that they had to offer. In comparing it with that of other schools, I found that CS.U. was lacking some opportunities that others had to offer. Nowhere on CSU's listing did I see course offerings for doctors or dentists. It confused me, and I have wondered about it since then.

If CSU expanded its horizons by offering more majors to their students, then people would appreciate the school more. In fact, the reason why most high school graduates go away from home to attend school is because they can't find what they want here in the city.

Another reason why there should be more offerings is that everyone has to go elsewhere for their educations, then that shows very little for CSU as a college. Other high school graduates may feel that CSU. will eventually be a low rated school, causing them to go other places.

It is my opinion that CSU is a good school. However, if they want to improve their images in the surrounding community, I feel that they should expand their horizons by adding more major offerings to their listings. Chicagoans would appreicate it very much. People from out of town, state & even country would look at CSU another way too. CSU would then certainly become a better asset to society.

APPENDIX B

On each printout, the circled entry is the score of the student whose essays you have reviewed in this article.

Keep in mind this scheme as you see his progressive scores:

ENTRANCE DIAGNOSTIC EXAM	400 to 500 words
EXIST COMPOSITION ONE EXAM	200 to 300 words
EXIT COMPOSITION TWO/	
PROFICIENCY EXAM	400 to 500 words

Keep in mind also our assumption that the greater the number of errors, the less competently the essay was written.

Social Security	Name	Def Dim Wtd No	Fun Dim Adj No	Def and Fun Joint Decsn	Grader Codf Recom	Gder and Dim Joint Decsn	Final Decsm
Entrance: Diagnostic Exam—Essay Sample One							
OBS 8 28	9	6	24		37 3		1 2 3
Exit: Composition I Exam—Essay Sample Two							
	0	24		Anom Chek 23 Fail			
OBS 11 10 5							F
ADJ 13 4 5							
Exit: Composition I Exam—Essay Sample Three							
	0	10		Comp Pass 37 Pass			Pass
OBS 3 4 1							
ADJ 5 3 1							
Exit: Composition II/Proficiency Exam—Essay Sample Four							
	0	11		Comp Pass 37 Pass			Pass
OBS 2 4 4							
ADJ 3 3 5							

APPENDIX C

The four dimensional framework which follows is just that—a framework for measuring the rudimentary competence of essays. The framework reflects the basic essay model and assumes the use of standard American written English. It does not imply that this is the only possible model but the basic model. The model and its dimensions are used to provide a means for a relatively systematic and specific recording of what a judge perceives as flaws in the report. As you will note, the whole framework is built on the principle of inclusion and exclusion or omission. Namely, the essay, paragraph, sentence or even word does not have something it should have or has something it should not have.

DIMENSION ONE   FLAW IDENTIFICATION 1-12

The Essay as a Whole: Logic, Organization, Development

The essay is flawed because
1. it does not address the question and is ineligible for grading -10
2. it has no introduction -3

3. it has a faulty introduction which does not imply or state the proposition -2
4. it has a faulty introduction which does not place the proposition in an overall context or imply how the proposition is to be developed -1
5. it has no body -5
6. it has a faulty body which is illogically organized -2
7. it has a faulty body which has statements irrelevant to the proposition -1
8. it has a faulty body which has statements not related to one another -1
9. it has statements which are not developed with specific details -1
10. it has no conclusion -2
11. it has a faulty conclusion which does not restate the generating proposition -1
12. it has a faulty conclusion which includes information irrelevant or contradictory to the introduction or body -1

DIMENSION TWO   FLAW IDENTIFICATION 13-23

Meaning and Style

The essay's meaning or style is flawed because
  I. it has faulty paragraphing in the essay when
      13. a necessary paragraph is omitted
      14. an unclear, repetitious, irrelevant, misplaced, factually incorrect or illogical paragraph is included
      15. a paragraph should/should not commence
  II. it has faulty structuring in its paragraphs when
      16. a necessary sentence is omitted
      17. an unclear, repetitious, irrelevant, misplaced, factually incorrect or illogical sentence is included
      18. the paragraph is lacking necessary details
  III. it has faulty phrasing in its sentences when
      19. a necessary element (word or word grouping) is omitted
      20. an unclear, incorrect, inconsistent, irrelevant, redundant, misplaced, dangling or unparalleled element is included
      21. the sentence lacks necessary details
  IV. it has faulty sentence construction when
      22. there is a run-on
      23. there is a fragment

Usage

The essay's usage flaws include
  I. Verb usage
    24. improper subject/verb agreement
    25. verb phrase omitting a verb form
    26. incorrect verb ending or verb form
    27. an inconsistent tense, mood, or voice
  II. Pronoun usage
    28. no antecedent for a pronoun
    29. pronoun not agreeing with its antecedent
    30. pronoun in incorrect case form
  III. Noun usage
    31. incorrect plural form
    32. no plural form
    33. incorrect possessive form
    34. no possessive form
  IV. Adjective usage
    35. incorrect comparative or superlative form
    36. no comparative or superlative form
    37. adjective instead of adverb or vice versa

    Word usage
    38. misspelled
    39. misused

Punctuation

The error has been made by the omission or incorrect usage of the following:
40. Capital letters
41. Period (unless the period creates a sentence fragment marked in Dimension II)
42. Question mark
43. Exclamation point

44. Comma (unless the comma creates a comma splice marked in Dimension II)
45. Colon
46. Quotation marks
47. Dash
48. Underlining
49. Hyphen
50. Parentheses
51. Apostrophe

Roberta S. Matthews

THE EVOLUTION OF ONE COLLEGE'S ATTEMPT TO EVALUATE STUDENT WRITING

Inexperienced writers need a tremendous amount of support and encouragement from their teachers. Students with writing deficiencies must be assured that they can write and shown that they have ideas worthy of development. Thus, after spending a term praising student progress and shoring up student confidence, most instructors find it difficult to fail students who have not made enough progress, whose skills still need more polishing. Developmental writing courses designed to prepare students for college-level writing tasks need objective criteria to judge student writing. Yet the very nature of the course makes it difficult to reconcile the objective evaluation with the subjective and affective process.

At La Guardia, a combined concern for maintaining standards and respecting human relationships accounted for the creation of an exit exam from ENG 100, Fundamentals of Effective Writing. This developmental writing course prepares students to enter ENG 101, Basic Composition, the first of a sequence of freshman composition courses. Sarah Barber, the Director of Composition, and I felt the need for a simple procedure which would ensure some uniform standards, while helping instructors maintain the very important relationships of trust and mutual striving that teachers of writing try to create with their students.

Since instructors tend to be considerably more objective about the writing abilities of unknown students, we decided that all ENG 100 instructors would exchange students' final in-class papers with each other. Each class was given a code number so that graded exams could be returned easily to each instructor, but graders did not know either the student or instructor related to a particular paper. In order to allow for the value and necessity of judgments based on an intimate knowledge of

Roberta S. Matthews, whose special interest is modern British and Irish literature, chairs the English Department at La Guardia Community College.

63

student work and achievement, we established a Screening Committee to which instructors could appeal their own students' failing grades and, on the basis of other in-class writings, argue for a passing grade for an individual student.

In order to pass the exit exam, students had to be able to generate, in a 70-minute period, a reasonably developed, logical and coherent 300-word essay with a recognizable beginning, middle, and end. We have tried to develop exit exam topics which allow us to judge how well we were accomplishing our course objectives:

1. To generate in students an enthusiasm for and enjoyment of writing.
2. To introduce and reinforce basic grammatical rules.
3. To introduce students to mature sentence structure.
4. To develop the students' ability to locate and correct their own errors of writing.
5. To instruct the students in paragraph development.
6. To improve students' understanding of logical and stylistic strategies for effective writing.

The essay had to contain fewer than eight major grammatical or syntactical errors. As stated in our performance objectives for the course:

Such errors include the following:

1. run-ons	7. homonyms
2. fragments	8. word choice
3. capitalization and lower case	9. tense
4. apostrophe	10. punctuation
5. spelling	11. plurality
6. agreement	12. omitted words

We chose our particular criteria because experience suggested that students needed to meet them to have at least a fighting chance of passing ENG 101.

All of our full-time faculty and most of our part-time faculty teach the whole composition sequence. This is important for two reasons. First, in addition to objective criteria our instructors have another way of judging students: "Would this student be able to pass my Basic Composition course? Would I want to have this student in my Basic Composition course?" If the answer to either of these is "no," graders will not give a passing grade to an exit essay. Second, if we are not satisfied with the

quality of student writing in our college composition courses, we modify procedures to prepare our students for our courses. The system demands that we continually question ourselves and evaluate our methods of evaluation. However, the basic structure is essentially the same this quarter when 36 instructors will process 51 sections of the ENG 100 exit exam as it was several years ago when 17 instructors taught fewer than 30 sections of our Fundamentals of Writing course.

The following pages, excerpted from our *English Faculty Handbook,* outline the essential skeleton of the exit exam and appeal process. After presenting the skeleton, I will flesh it out by suggesting why and how certain procedures evolved and relating the kinds of revelations which continually convince us that, for all its problems, our exit exam is worth the effort.

The *Handbook* describes the mechanics of the exit exam as follows:

PROCEDURES FOR THE EXIT EXAM

Topics for the exam may be assigned by the Director of Composition. If this is the case, instructors will be informed when they can pick up the list of topics for their classes. All exams must be administered before noon on Wednesday of the tenth week of the quarter and turned in to the assigned tutor in the Writing Center. Then, you may pick up the batch of exams you are to mark. No instructor marks his or her own students' papers. Return all graded exams and pick up your own students' graded exams.

ADMINISTERING THE EXIT EXAM

1. Do not announce topics ahead of time.
2. Have students write their exams in "blue books" available in the English office.
3. Each instructor will be assigned a number for the exam. Have all students write this number clearly on the cover of the "blue book." Be sure that they do not write your name anywhere on the exam.
4. Check each exam to see that the student has written his or her name clearly and legibly on the cover.
5. Stress to students the need for legible manuscripts (their readers will be unfamiliar with any penmanship idiosyncrasies)—and the need to leave sufficient margins on the right and left.
6. Tell students that they shall not have time to "copy over" their rough drafts.

GRADING EXIT EXAMS

1. Use a red pen—or a pen of some other contrasting color—so all corrections stand out clearly.
2. Circle and mark each error with appropriate correction symbol.
3. All exams are graded either "Pass" or "Fail." In addition to judging the content, use the list of major errors in the performance objectives as your guide—eight or more constitute a failing exam. However, an error that is *exactly* repeated (e.g., a misspelling, the same wrong verb ending) should only be counted once. Note: of the eight errors, no more than two can be run-ons and/or fragments.
4. The matter of what constitutes a "major" error is, at times, judgmental. For instance, most instructors would not count a missing comma a major error unless its absence distorted the sense of the sentence. To make their criteria absolutely clear, many instructors follow this system: A) circle and identify with the appropriate symbol each and every error on the exam; B) go back over the exam, writing the symbol for each major error in the lefthand margin of the line on which it occurs; C) write the number of major errors on the cover of the booklet.
5. Read through the exams you are grading twice—the first time for a rough grading and the second for verification.
6. When done with an exam, write "Pass" or "Fail" on the cover and "Graded by #.....," filling in your assigned number. Grade as objectively as possible, without necessarily giving the student the benefit of the doubt. An instructor may bring a borderline case to the Screening Committee if he or she feels the student should pass.
7. Students see (but do not keep) the graded Exit Exam. Therefore, if you have a comment for the instructor (e.g., suggesting an appeal), write it on a slip of paper and put it in the booklet.
8. Graded Exit Exams from past terms are on file in the English office should you wish to consult them as a guide.

THE APPEAL PROCESS

A student who fails the Exit Exam can still pass ENG 100 if a member of the Screening Committee certifies that the student has, during the term, written at least two 300-word in-class essays that are clearly passing in terms of Exit Exam standards.

You should prepare for a possible appeal for all your students, since it will be too late to collect essays after the Exit Exams are returned and you know which, if any, students need to be appealed. Follow these steps:

1. Have all students write an "appeal theme" in the class session following the in-class Exit Exam.
2. *Grade the exam yourself*—strictly by Exit Exam standards. A passing appeal theme can be one of the two papers needed to appeal a failing Exit Exam.
3. Have students bring all their graded essays to class—either on Exit Exam day or appeal theme day. Select, for each student who has them, his or her two best passing essays.
4. Present your student's "case" to the committee.

Once an instructor and member of the Screening Committee agree that a student should receive an F for the course, that F may *not* be changed. A list of all failing ENG 100 students goes immediately to the Registrar.

SAMPLE FAILING EXIT EXAM

Below is a sample of a failing Exit Exam. It contains most of the major errors which students need to eliminate in order to pass ENG 100: insufficient development, repetition, run-ons, fragments, misspellings, plurality errors, subject-verb disagreement. Other samples of Exit Exams (passing and failing) are on file in the Writing Center.

Assignment: How has being in college changed your life? Answer any way you wish. You might want to consider any or all of the following: relationships with family, jobs, sense of time, life goals, values.

How College Has Changed My Sense of Time

Before I came to La Guardia I used to have alot of time. But now I find myself at home studying, reading and thinking more. I have read more books in college, than I have ever read before in my life. I am in school almost all day, then when I leave school I go to work, so when I do get home it is too late to do anything, but study. I only get to see my friends on weekend, before I started college I would see them everyday. My family and I only get to see one another at nights. I spent more time in school than in any other place. It seem like they is always something for me to do in school, or work that must be done for one of my classes. When I was in high school, there used to be alot of time for me to get into sports. As soon as I get home from work, I rush to eat, wash up, and do my homework for the next day. I find myself spending alot of time in the library, than ever

before, doing all different kinds of reports. The only time I have to enjoy myself is on the weekend, and then sometimes I am not in the mood to go out. I also use to spend a lot of time with my family, but now we dont even see each other very much. I use to hang out in the park at night with the fellow, but now all my nights or spent at home studying for exam. I also have to leave my house must sooner than before. Because of the long ride to school.

Now let me share some reasons for our policies and some things we have learned from our activities. The process of evaluating student writing has led us, perforce, into evaluating ourselves as teachers, our assumptions, our methodologies, and our standards. The process itself has been so informative that although we constantly modify it, I doubt that we will ever abandon our exit exam and screening committee.

First, and most important, our procedure is based on trust. Our instructors administer the exit exam in their own classes. They appeal failures with student essays written in class and we assume the essays were written without help and that they are originals, not rewrites. And, our instructors trust us. We have taught enough sections of ENG 100 ourselves to know that the percentage of passing or failing students does not reflect absolutely on the abilities of our instructors. They are not accountable for the number of students who fail the exit exam nor do we keep tabs on their track record. An instructor with 16 passing students one quarter might have another class in which 16 students fail the exit exam. Too many other variables preclude using student performance on the exit exam as a way of evaluating teacher performance. We use other means to evaluate our instructors.

For an instructor new to our system, the exit exam comes as something of a shock. However, as time goes on, grateful instructors learn to use the process constructively. An amorphous "they" fails students. Thus, instructors can inform students of their need to repeat the course without sacrificing or destroying the very important relationship between them which it often takes a whole quarter to establish. Instructor and student unite against the system; they can point to the progress they have made together and vow to continue to make more progress together. Many students who fail the exit exam demand to be placed in the same instructor's course during the new quarter. When we began the exit exam we did not anticipate this extra dividend, but it certainly helps student motivation and morale.

When we began using the exit exam, each instructor was given a number to place on each student's exam to ensure anonymity and to help

us return essays to the proper class. Eventually, we asked instructors to write "graded by #   " and to supply their number on each essay they graded. Since we were apprehensive about the idiosyncratic grading methods of some instructors, we wanted to be able to identify and speak with instructors who had problems.

Occasionally, instructors made inappropriate stylistic judgments about student writing. Perfectly good sentences were modified unnecessarily to conform to individualistic preferences. Once or twice, the method of grading made us suspect an instructor's ability to explain concepts to students. Sometimes the standards applied were too stringent; sometimes too easy. The master list of numbers was confidential. Any issues which arose with individual instructors were dealt with privately between the instructor and the Director of Composition.

During one quarter (and probably by mistake), we published the master list. Again, an unforeseen but valuable result occurred. Individual teachers sought out each other to argue about differences in grading and standards and helped us evolve more standardized and generally acceptable procedures. Instructors discussed everything from obscure grammatical points to the quality of the content of a particular paper. They debated the merits of experiential as opposed to analytical content. They argued about how one evaluates a student's ability to write well and to improve his/her writing.

Regardless of how many training sessions are scheduled during a rushed and crowded quarter, it is difficult to be absolutely sure that all instructors are using the same criteria to grade exit exams. The *ad hoc* individual conversations described above evolved spontaneously. Over the course of time, we developed another means of standardizing our grading. In the beginning, we distributed exit exams for grading in a random fashion. Fairly quickly, however, we decided to use our system to our advantage. Instead of a random distribution, we began pairing experienced instructors, who had gone through the exit exam procedure several times, with new instructors. This simple modification ensured some kind of quality control and helped integrate new instructors.

Although we set up our exit exam so that instructors would not be required to grade additional papers, the appeal process does demand more from its participants. The Screening Committee consists of the most experienced full-time instructors in our department. The screening process, coming as it does during the final week of the quarter, creates a hectic final week, but yields an education in evaluation for all involved.

Our procedure is simple. At his or her discretion, an instructor may fail a student who somehow passes the exit exam but whose work during

the quarter suggests that the student could not cope with more advanced writing assignments. Bitter experience has taught us not to believe in miracles. Marginal students rarely rise to the level of the next course but rather need the reinforcement of skills and content development which taking ENG 100 over again will give them. On the other hand, many students fail the exit exam (through nervousness, for example) whose work during the quarter suggests that they can successfully take and pass ENG 101. The Screening Committee exists as an appeal board for instructors who believe their students should pass ENG 100.

We tend to pass a reasonably high percentage of students on appeal. But the process is more difficult to capture. We agonize over each student together. We review a term's work. We discuss in minute detail the student's writing. We grope for a fair decision. We try to second guess a student's abilities, responses, and reactions to challenge. We try to be just. And we continue to do it because we believe the system works.

Elizabeth Metzger

MEASURING GROWTH IN COLLEGE WRITING

INTRODUCTION

The Learning Center, State University of New York at Buffalo, has a scheme for evaluating the writing program. The scheme encompasses student evaluation of courses, reports by each teacher on the progress of one student, the coordinator's evaluation of teachers, measurement of growth in writing samples, and examination of the students' writing anxiety. Since the entire scheme is rather lengthy, I will describe the measurement of writing and of writing anxiety in a one or two semester writing course.

PURPOSE AND PROCEDURE OF THE EVALUATION

In the writing classes students discussed the importance of audience, purpose, and persona to a piece of writing, completed writing tasks, wrote essays, combined sentences from the Strong workbook,[1] and did exercises in sentence structure, grammar and usage.

As the supervisor of six graduate teaching assistants of the writing courses, I wanted to see if our courses were having any effect on improving students' writing in terms of syntactic fluency and overall quality of the writing. In addition, I wanted to know if we as teachers were reducing students' anxiety about writing, believing that for many students their high anxiety about writing was partially responsible for their poor writing.

GETTING READY

I adapted Diederich's[2] and Cooper's[3] schemes to measure growth in writing and used Daly and Miller's[4] instrument to determine the writing anxiety level of students (see Appendix A).

Elizabeth Metzger is an Assistant Professor of English at Youngstown State University. At the time of writing this article, she was Coordinating Instructor of Writing at the Learning Center, State University of New York at Buffalo.

1. William Strong, *Sentence Combining* (New York: Random House, 1973), pp. 10-205.

Before the Spring 1977 semester began, several teachers and I met and compiled a list of ten writing tasks in the expository mode, a mode in which many students would be required to write in other college courses. The tasks provided students with a purpose, audience, and role. The directions for the task, along with sample tasks, follow:

University Learning Center

Writing Evaluation Plan

A. Rehearsal and Pre-writing

Choose a topic and begin thinking about it. Research it if you want, talk with others about it, make notes and jottings, make an outline, or do anything else that will be helpful to you when you write the essay.

Any notes you make to bring to class with you will be examined by the instructor. The instructor will also take up the notes along with the finished essay. Put your name on each page of any notes you submit.

B. Writing

The actual writing will be done in class on specifically identified paper given you by your instructor.

DIRECTIONS

1. Your instructor will provide you with a 3 by 5 index card and an identifying number. Print your name, date, class, and instructor *on your card*. Write the identifying number in the right-hand corner of your card and paper.
2. Write the *number* of your topic in the *left-hand* corner of your card.
3. Write on the topic you have rehearsed. You will have *only* fifty minutes to write the essay.
4. Write the final essay on *one* side of each sheet of paper.
5. Submit the final draft of your essay by the end of the period.

Explanatory Writing Tasks

1. In an attempt to improve her teaching technique, one of your professors this semester has made a somewhat surprising request of her class. She has asked you to recall past school experiences and, after some thought,

2. Paul B. Diederich, "How to Measure Growth in Writing Ability," *English Journal, 55* (1966), 435-499.

3. Charles R. Cooper, "Measuring Growth in Writing," *English Journal,* 64(1975), 11-120.

4. John A. Daly and Michael D. Miller, "The Empircal Development of an Instrument to Measure Writing Apprehension," *Research in the Teaching of English,* 9(1975), 242-249.

to describe the ways in which you feel you learn best. *Her* objective is to find out the best ways to "reach" her students; *your* objective is to describe your most productive learning environment.

2. There has been a growing public controversy over the advantages and disadvantages of T.V. viewing as compared with reading. Some people feel that the information from television is of greater importance and provides more entertainment than the information from books. Other people feel the opposite way: that books are more important than television. *Newsweek* magazine is interested in the perspective of the college student and has chosen you to respond to a particular question: If you were asked to give up one of these two experiences (T.V. viewing or reading) for the rest of your life, which would you give up more easily and why? The *Newsweek* people are interested in the reasons for your choice and plan to publish your essay in a forthcoming issue.

3. The professor in your writing class has just announced that he/she is interested in student opinion and suggestions about a grading system to be used in his/her class. He wants to devote the next class period to a discussion of what students have to say. Each student will get a chance to speak. You, as a student, are vitally concerned with the way grades are decided. Write what you intend to say to the professor and the rest of the class when your turn comes.

4. The "Energy Crisis" has made people more conscious of preserving natural gas and electricity. Imagine that you have lost the comfort of both these resources. Explain to someone who has not experienced this loss how you have had to readjust your life. Is the adjustment harsh or just different?

Students wrote on four different tasks for the two pre and two post writing samples. Students selected their writing task several days before the actual in-class writing. We believed, as did Sanders and Littlefield,[5] that students needed time to think about, research, take notes, and rehearse what they planned to write.

OBTAINING RELIABLE RATER JUDGMENTS

According to McColly,[6] the proper training and orientation of the rater is essential to obtaining reliable judgments about the quality of a paper. In his article, he reported on a study in which readers judged essays averaging four hundred words in length at the rate of one essay per

---

5. Sara E. Sanders and John H. Littlefield, "Perhaps Test Essays Can Reflect Significant Improvement in Freshman Composition," *Research in the Teaching of English* 9(1975), 145-153.

6. William McColly, "What Does Educational Research Say About the Judging of Writing Ability?", *The Journal of Educational Research,* 61(1970), 148-156.

minute. McColly explained that "if a reader is competent and if he has been well trained and oriented, his instantaneous judgment is likely to be a genuine response to the thing for which he is looking. But if he is given the time to deliberate, he is likely to accommodate his judgment to tangential or irrelevant qualities which will introduce bias into the judgment."[7]

During the semester, the teachers read pre and post samples written by students in a previous semester. No names or dates were on the paired papers which averaged less than four hundred words each. Each teacher independently read and chose the better paper in three minutes. These papers served as practice for members of the group who would evaluate pre and post writing samples at the end of the Spring 1977 semester. As a group, they agreed about which paper of the set was better; of twenty-five sets of practice papers, they agreed that the post sample was better in twenty sets.

COLLECTING AND CODING THE SAMPLES

1. Before instruction began I assigned each teacher a block of twenty-five identification numbers (I.D.) for her students to use. Each class was limited to twenty-five students.
2. In the first two and last two weeks of instruction, each student selected a number from the teacher's block and used it for the two pre and two post writing samples.
3. Teachers collected the samples and gave them to me. I locked them up. This procedure was repeated during the last two weeks of instruction when the students again wrote.
4. I then recorded the teacher's name, block of identification numbers and each student's name and identification number on a form like the following:

Instructor:		Semester:				
Course:		I.D. Numbers:				
Name	Better Paper-First Impression			Reader/Rater		
	I.D.#	Pre	Post	1	2	3

5. If two students had used the same number (e.g., number 3), I assigned an A by one of the student's numbers. Therefore, his new number became 3A.
6. Any student who wrote his name on the paper or mentioned the teacher's name, the date, or season was discarded from the group. This procedure reduced the samples to sixty-one paired sets.
7. I separated pre samples from post samples and removed the I.D. cards. By flipping a coin, I assigned an A or B in the right hand corner of the second pre writing samples. When I flipped the coin and "heads" occurred, I assigned an A; when "tails" occurred, I assigned a B. Using the student's I.D. number, I clipped the second pre and the second post samples together in random order. I gave each post writing sample of a paired set the remaining letter; for example, if a student's pre sample was 4B, his post sample was 4A. I carried out this procedure to prevent all pre samples from being assigned an A and all post samples a B—a natural ordering of the letters. Without a randomizing procedure, the ordering might have caused the rater to select the better paper by the letter code and not the quality of the paper. I coded the second pre and second post writing samples because I believed that students would have become more familiar with the rehearsal period and the writing procedure. I reasoned that students' second pre and post would reflect their writing ability to a greater extent than the first pre and post "warm-up" samples.

EVALUATING THE PAPERS

The papers were ready to be judged by the six raters. I paired raters and told them to select the better paper of a set. I gave one rater of the team five sets of papers to read in fifteen minutes, about three minutes for each set. Since they had read papers for practice throughout the semester, discussing and examining good and bad pieces of writing, I gave no instructions to the raters concerning what they should look for in the paired sets. No teacher/rater evaluated papers from her own class.

After reading the five sets of papers, the rater checked column A or B on a sheet like the following:

---

7. McColly, p. 150.

Reader/Rater:				Semester:		
	Better Paper				Better Paper	
Paper I.D. Number	A	B		Paper I.D. Number	A	B

Next, the rater passed the papers back to me. I gave them to the second rater who followed the same procedure. In an attempt to prevent reader fatigue[8] and to maintain efficiency in judging, I limited the rating of the papers to ten sets a day.

When all of the papers had been read by two raters, I recorded their responses on the teacher's class list under the column "Better Paper— First Impression: Reader/Rater 1,2,3." For each paper I recorded the response of raters one and two. If their responses showed that they agreed that a paper was the better of the two, I put the paper into an "agreement" pile; if they disagreed, I put the paper into a "disagreement" pile. I submitted these papers to a third rater. Thus, I sought agreement from two raters on which paper was the better of a set; raters were judging the overall *quality* of the paper. I believed that students would write better by the end of the course and that raters would confirm my belief by selecting as the better paper the post sample more often than the pre sample.

COUNTING T-UNITS AND ERRORS IN THE SAMPLES

Teachers counted T-units and total number of words in the writing samples. They paired the first two and last two papers, took fifty T-units from each pair, and made the count. They obtained T-unit lengths by dividing the number of words by the number of T-units.[9] They then entered the information on the following form:

---

8. Richard Braddock, *Research in Written Composition* (Champaign: National Council of Teachers of English, 1963), p. 11.

9. Frank O'Hare, *Sentence Combining: Improving Student Writing Withough Formal Grammar Instruction* (Urbana: National Council of Teachers of English, 1973), p. 149.

Teacher:				
Course:		Semester:		
	Pre		Post	
Student's Name/I.D. Number	Total Words	Words/ T-Unit	Total Words	Words/ T-Units

In addition, teachers counted deviations from standard English usage [10] in the coded samples and subtracted the pre score from the post score. They used the following form to record the information:

Student: I.D. Number: Semester:

Deviation from standard English

                                       Pre   Post   Difference

A. Spelling
B. Punctuation
C. Capitalization
D. Wrong Word
E. Fragment
F. Verb Tense
G. Run-on
H. Plural
I. Verb agreement
J. Possessives
K. Case pronouns
L. Pronoun agreement
M. Misplaced modifier
N. Other:......
    Total

---

10. Samuel A. Kirschner and G. Howard Poteet, "Non-standard English Usage in the Writing of Black, White and Hispanic English Students in an Urban Community College," *Research in the Teaching of English,* 7(1973), 351-355.

Why the T-unit and error counts? I wanted to see if students were using a larger quantity of words in the post samples than they were in the pre samples. After exposing students to intensive writing instruction, I expected them to elaborate and to use more words to clarify an idea through examples, illustrate an object by naming its attributes, and point, to detail using prepositional phrases.[11] Hence, I expected longer sentences with a variety of embeddings and fewer deviations from standard English usage.

SCORING THE WRITING ANXIETY SCALE

Daly and Miller designated the twenty-six items of their anxiety scale as negative or positive. They offer a formula for determining writing anxiety level (see Appendix B). The lowest possible score is 26 and the highest is 130; the higher the score is, the less anxious the student.

INFORMING STUDENTS

Teachers could report information to students on a profile summary sheet like the following:

	Paper	Date	Total Words	Words per T-Unit	Deviances from Standard English	Writing Anxiety
Student's Name:				**Semester:**		**Course:**
Teacher:						
Pre 1						
2						
Post 3						
4						

This profile provides the student and the teacher with important information. First, the results are diagnostic and permit students to know where they are weak or strong in several categories. If weak in certain areas, as revealed by the pre writing samples, they could work on remedying these areas during the semester. If students are still weak by

---

11. Francis Christensen, *Notes Toward a New Rhetoric: Six Essays for Teachers* (New York: Harper and Row, 1967), pp. 27-33.

the end of the course, they could work independently in self-help books without taking another semester-long course. Secondly, the profile sheet is helpful in that the teacher and the student can compare the two papers written before instruction began with the two papers written at the end of instruction to ascertain whether the student progressed. For example, if words per T-unit, and total number of words increased while deviances from standard English usage decreased in the set of post writing samples, then the student will have improved in his ability to produce longer, more error-free writing.

CONCLUSION

I have presented a simple scheme for measuring growth in writing that any English teacher or writing program can use. The scheme encourages the collection of qualitative and quantitative measures near the beginning and near the end of the course; the scheme encourages efficient analysis that teachers can perform far in advance of the ensuing course to examine and revise, if necessary, their existing approaches to teaching and measuring growth.

The need for such a scheme is this: (1) We must diagnose early in the course the writing skills that students are weak in and try to remediate those weaknesses; (2) as competent teachers, we must be able to demonstrate that our courses had an effect on students—that students become better writers as a result of taking the courses.

In any writing class or program we accept the fact that writing is a difficult task and growth comes slowly. However, we should be not only *willing,* but also *able* to measure growth in the writing of our students. Using this simple scheme I have presented, the overwhelming task of measuring growth in writing becomes manageable.

APPENDIX A

QUESTIONNAIRE (Daly and Miller)
Learning Center
State University of New York at Buffalo

DIRECTIONS: Below are a series of statements about writing. There are no right or wrong answers to these statements. Please indicate the degree to which each statement applies to you by circling whether you (1) strongly agree, (2) agree, (3) are uncertain, (4) disagree, or (5) strongly

disagree with the statement. While some of these statements may seem repetitious, take your time and try to be as honest as possible. Thank you for your cooperation in this matter.

1. I avoid writing.   1 2 3 4 5
2. I have no fear of my writing being evaluated.   1 2 3 4 5
3. I look forward to writing down my ideas.   1 2 3 4 5
4. I am afraid of writing essays when I know they will be evaluated.   1 2 3 4 5
5. Taking a composition course is a very frightening experience.   1 2 3 4 5
6. Handing in a composition makes me feel good.   1 2 3 4 5
7. My mind seems to go blank when I start to work on a composition.   1 2 3 4 5
8. Expressing ideas through writing seems to be a waste of time.   1 2 3 4 5
9. I would enjoy submitting my writing to magazines for evaluation and publication.   1 2 3 4 5
10. I like to write my ideas down.   1 2 3 4 5
11. I feel confident in my ability to clearly express my ideas in writing.   1 2 3 4 5
12. I like to have my friends read what I have written.   1 2 3 4 5
13. I'm nervous about writing.   1 2 3 4 5
14. People seem to enjoy what I write.   1 2 3 4 5
15. I enjoy writing.   1 2 3 4 5
16. I never seem to be able to clearly write down my ideas.   1 2 3 4 5
17. Writing is a lot of fun.   1 2 3 4 5
18. I expect to do poorly in composition classes even before I enter them.   1 2 3 4 5
19. I like seeing my thoughts on paper.   1 2 3 4 5
20. Discussing my writing with others is an enjoyable experience.   1 2 3 4 5
21. I have a terrible time organizing my ideas in a composition course.   1 2 3 4 5
22. When I hand in a composition I know I'm going to do poorly.   1 2 3 4 5
23. It's easy for me to write good compositions.   1 2 3 4 5
24. I don't think I write as well as most other people.   1 2 3 4 5
25. I don't like my compositions to be evaluated.   1 2 3 4 5
26. I'm no good at writing.   1 2 3 4 5

APPENDIX B

ANXIETY SCALE SCORE SHEET

Student's Name (Print)　　　　　　　　　　　Semester　Year　Pre　Post

Writing Apprehension =
78 + Positive Score -
Negative Score

SCORE

Question　　　　　　　　　　　　　　　　Question

	Positive	Negative			Positive	Negative
1.		▦		14.	▦	
2.	▦			15.	▦	
3.	▦			16.		▦
4.		▦		17.	▦	
5.		▦		18.		▦
6.	▦			19.	▦	
7.		▦		20.	▦	
8.		▦		21.		▦
9.	▦			22.		▦
10.	▦			23.	▦	
11.	▦			24.		▦
12.	▦			25.		▦
13.		▦		26.		▦
				TOTAL		

81

Muriel Harris

EVALUATION: THE PROCESS FOR REVISION

It is becoming increasingly clear that teacher evaluation of student writing, offered as a final judgment on a finished product, is only minimally useful as a tool for learning. We, of course, find student writers who can abstract and apply to their next writing what they have learned from the list of errors, deficiencies, and successes noted on their finished papers, but for too many basic writers there is little retention and even less interest in the contents of such post mortems. Even when we evaluate students' papers and ask for revisions, we are entering into the act too late if the first comments a student receives are directed toward a draft which is already, to some degree, suffering the onslaught of *rigor mortis.*

What we need, then, for truly useful evaluation is a continuing program of offering feedback to student writers as they move from the initial chaos of the unrefined subject to a well articulated written product. Moreover, we need to provide students with different purposes and methods for each stage of evaluation to fit their needs as they develop each piece of writing and as their general skills improve. In addition, the student's own evaluation skills should develop as the semester progresses so that his initial responses give way to more mature judgments. Finally, the instructor needs a format or strategy for evaluating the writing skills the student has acquired by the end of the course. The program of evaluation offered here aims at achieving these goals.

We should first appreciate that the acquisition of evaluation skills through on-going critiquing is essential for the student who has not yet adequately developed his own skills as the primary critic of his writing. To move students beyond that passive waiting to see "what's wrong," what The Teacher wants corrected, we cannot be the sole graders during a semester or two of composition courses and then suddenly turn the

---

Muriel Harris is Director of Purdue University's Writing Laboratory and the editor of the Writing Lab Newsletter.

student loose to become a self-regulating editor who can effectively spot the need to reorganize, revise and correct. We must wean the student so that he or she becomes not only an independent writer but an independent critic as well. We can accomplish this by helping students acquire the very different kinds of evaluation skills appropriate to each stage of a piece of writing by providing models for evaluation. and opportunities for extensive practice in different kinds of evaluating. Our first task, then, is to differentiate the types of feedback needed in the pre-writing, writing, and revision stages and at the same time to consider how strategies such as peer criticism and evaluation forms can help the student learn how to become his own best critic—and not incidentally, a critical reader of other writing.[1] Ideally, in the best of all possible writing courses, students should be able by semester's end to grade their own papers with some degree of accuracy.

Evaluation begins where any writer begins, with the pre-writing stage which, as Donald Murray so succinctly describes it, "is everything that takes place before the first draft. Prewriting usually takes about 85% of the writer's time. It includes the awareness of his world from which his subject is born. In prewriting, the writer focuses on that subject, spots an audience, chooses a form which may carry his subject to his audience."[2] Well said, but how can the inexperienced basic writing student who has either been ignored or forced to write for a lone "Teacher-Grader" spot his audience if he has not yet developed a clear sense of the distinctions between different audiences, their interests, and their varying needs for information. Feedback on these matters from a real audience is the first need of the inexperienced writer, and it can be offered easily in small groups who come together to react to each other's suggestions or proposals for a paper.

When this initial pre-writing exploration proceeds orally, an apprentice writer can test his ideas aloud by "talking it out" or reading from jotted beginnings in a journal or roughed out notes of a preliminary planning draft. In whatever way he chooses to proceed, the writer who is not yet comfortable with the idea of writing as communication needs to

---

1. Surveys of studies showing the effectiveness of peer grading can be found in Ross Jerabek and Daniel Dieterich's "Composition Evaluation: The State of the Art," *College Composition and Communication,* 26 (May 1975), pp. 183-186, and an ERIC/RCS Report by Turee Olsen, "Grading Alternatives," *English Journal,* 64 (March 1975), pp. 106-108.

2. Donald Murray, "Teach Writing as a Process Not Product," *Rhetoric and Composition: A Sourcebook for Teachers,* ed. Richard L. Graves (Rochelle Park, New Jersey: Hayden, 1976), p. 80.

try out his first formulation of his subject in a small group setting. It is here that he will gain that initial sense of audience which, as Murray reminds us, is so necessary in the pre-writing stage. While there are a few excellent texts which offer beginning exercises in varying the audience (describe a party first to a close friend and then to a parent, etc.), not all basic writers succeed merely by being reminded of the varieties of audiences that exist, for the egocentric writer continues to see the world from his own perspective. It is the live, questioning, reacting audience which most effectively jars the writer into an evaluation of whether he has appropriately communicated.

I have found that students who meet in small groups in the classroom to send up their trial balloons do several useful things in the act of talking out or reading their first suggestions.[3] They often embellish on or continue to create content as they talk, adding to or rejecting what they are offering not only because the mental juices are beginning to flow but also because of their changing perceptions of the audience's reactions. Verbal or non-verbal reinforcement from another student who really begins to listen suggests that they may have some very real reader interest; a question from another student makes the writer aware of the need for more information or the need to develop another aspect of the topic. In one way or another, if the members of the group are actively engaged in helping each other to begin their papers, the writer will start to gather useful information about who his audience is. The instructor's role in this stage of evaluation is really that of a facilitator who establishes a comfortable level of openness in the classroom, brings the groups together, and offers only minimal structure for the groups' task, perhaps no more than a rehearsal of some open-ended suggestions or a vocabulary for useful responses.

Before the small groups begin their listening and responding, the instructor can also remind the class that they may emerge from their sessions with somewhat altered conceptions of the direction or emphasis of their original topic. In one of my basic writing classes, I can vividly recall a small group session early in the semester in which a shy and very inarticulate student from a farm hesitantly offered his group the possibility of writing an explanation of the high cost of raising a calf to the stage of being sold. The student was generally a reluctant writer, unable to produce more than a paragraph or two on a given theme that

---

3. This process can go on in a writing lab tutorial or in instructor-student conferences though, of course, the audience is more limited.

he had chosen because of his interest in or knowledge of the subject. His manner when presenting the topic suggested that he expected silence, disinterest, or laughter from his group—the kind of feedback that would relegate this topic to a bare-bones repetition to himself of what he already knew. Instead, a student from an inner city area responded vigorously that she really wanted to know why beef was getting so expensive. Startled, the agriculture major began his explanation but found himself being interrupted by others in the group who needed this or that bit of information in order to follow this explanation that was going to tell them why meat was disappearing so rapidly from their dorm menus. Finally, the prospective writer realized that he really wanted to write a paper for city-bred consumers of meat, persuading them that farmers would have to ask even higher prices for their beef cattle in the future in order to survive.

It is here in that all-important stage of pre-writing where, as James E. Davis explains, "The talker may work himself toward a stance or a commitment on a subject."[4] Certainly, the writer may find his own stance, but the interaction with the audience is what helps the basic writer learn how to sharpen or define it, particularly when he has not yet developed a sense of writing as public communication. When a basic writer is writing not for self-discovery alone, but for that public beyond himself, he can learn how to evaluate his initial judgments, to base the writing not on his intention of what an audience might want, but on their real reactions. The more the writer is exposed to this kind of feedback, the better able he is to begin building some generalizations about the future audiences he will write for.

This discussion of small group pre-writing feedback may sound like nothing more than a re-warmed version of "class discussion," but it isn't. Faced with speaking up in a large classroom, students rarely compose orally or react to someone else's composing process with the same ease that they do in small groups, and the feedback in a large class often has to be encouraged or provoked by the teacher, the result being that the student who responds to another student's talking too often has one eye on the instructor's reaction to his comment. In a small group having about five members for optimal effectiveness,[5] the likelihood for

---

4. James E. Davis, "The Blockhead Writer: A Confessional," reprinted in *Rhetoric and Composition*, p. 219.

5. Ernest and Nancy Bormann, *Effective Small Group Communication* (Minneapolis: Burgess, 1972), p. 6.

useful interaction increases greatly. I have noticed also that students tend, at first, to write to their particular peer group, but as the sense of audience becomes firmer, there is the urge to branch out to other audiences (provoking a need for the group to role-play another audience). However, when there is negative feedback, the writer begins to make choices. Is he writing primarily for himself; or, if it is public discourse he is engaged in, is there another kind of audience who would be likely to be more receptive to the topic? These important questions and distinctions can be discussed in the abstract, but they seem to flow more naturally after group reaction time.

After the writer has had some pre-writing feedback and some time to turn his suggestions into a first draft, he is ready for a different kind of evaluation, a more structured critiquing by a group in which the writer may or may not be present (though I find that both situations should be tried). Again, the evaluation is offered primarily from the writer's peers, though the instructor can be a more active participant in offering models for evaluation by means of evaluation forms. The questions to be answered on these forms are a way of giving direction to the group's task, but more important, they are an aid to basic writing students who usually do not, at first, have a clear idea of what they should be looking for in trying to judge whether a piece of writing is good. I have found that evaluation sheets for the group to fill out early in the semester are best kept very general, seeking mainly for some of the more easily arrived at holistic responses, e.g.:

Did the panel of readers enjoy reading this paper?

If so, what contributed most to the enjoyment—intersting topic, vivid details, etc.?

If not, what could make the paper more effective—more description, clearer focus on the subject, etc.?

To suggest to the writer the range of audience reactions, I usually leave spaces after each question on the sheet for the readers on the panel to respond separately if there is no clear consensus among them. Because a basic writer also needs to realize that some parts of a paper can be more successful or less effective than others, I include on early evaluation forms questions such as:

Which is the best part of this paper? Why?

What should be left out, changed, or expanded?

These kinds of questions are encouragingly easy to respond to as the writer-reader starts to flex his critic's muscle; similarly, such questions are fairly easy to internalize as guides for the writer's next writing. Another very useful question on an evaluation sheet used early in the semester is one that asks the panel of readers to state what they think the main point or thesis of the paper is, thus seeking out the degree of overlap between the writer's intention and the reader's perception.

As the semester progresses, the evaluation sheet questions for this second or rough draft stage (after the initial pre-writing feedback) become more precise to include new concerns that are being discussed in class, such as effective use of introductions and conclusions and paragraphing. To help "test the effectiveness of a student's piece of writing as a whole," Richard Larson offers four questions to ask which, though intended for use by teachers, can and should become students' criteria as well:

1. Does the writer perform felicitously the act he promised?
2. Are the conclusions, the judgments, consistent with and supported by the data and arguments that precede them?
3. Is it possible for the reader to see, from beginning to end, in what direction the piece is moving, what steps are taken to reach the writer's goal, and why?
4. Who is talking to us? Are we in the presence of a faceless speaker or a distinctive identity? Is that identity consistent within the paper, and is it suitable to the writer's goal in coming before us? [6]

On the students' evaluation forms we may not be able to ask all of Larson's questions as fully as they are presented here, but we ought to be moving the class toward an understanding of these criteria.

We ought also to listen to the students' sense of what they consider to be important standards by which to judge their writing. If the evaluation sheets have been working effectively, the questions originally suggested or structured primarily by the teacher should give way, later in the semester, to the class's suggestions. When the evaluation sheet is made up of criteria which the students have chosen as their goal for the assignment and have themselves written in their own phrasing, it has a validity which no textbook list of recommendations could ever hope to achieve.

----

6. Richard Larson, "The Whole Is More Than the Sum of Its Parts: Notes on Responding to Students' Papers," *Arizona English Bulletin,* 16, No. 2 (February 1974) pp. 176-177.

Before I proceed, I ought to include another rationale for these evaluation sheets because for some they surely seem like that bureaucratic approach to life we prefer to avoid. Filling out these forms is an excellent oportunity to practice conciseness, clarity, and accuracy in writing, for the act of answering the questions requires that the critic select from the flow of the group's conversation the relevant words that need to be recorded. It quickly becomes apparent that unclear or partially explained evaluations are less than useful to the writer when he later consults his sheet for suggestions as he proceeds to the next stage of revision. In addition, as I move around the classroom during evaluation sessions, I find that students in their roles as critics may need help in articulating vague impressions. Sometimes I am able to help the group see the connection between what they are groping for and what we may have been discussing in class. At other times the group and I need to examine the sentence or paragraph that the evaluators cannot adequately judge, to see what criteria we can apply.

In sum, the teacher's role during the stage of panel evaluation is, first, to structure the evaluation procedure so that students can practice and refine their critical skills; and second, to be available for help in recording the kind of evaluation that will also be useful to the writer. The teacher's role here is somewhat more structured than Thom Hawkins' suggestion that the teacher's most effective role generally in small groups "is to facilitate learning by questioning, listening, and observing,"[7] but there is a need for models (at least, initially) for evaluation criteria. After the group has done its work and the writer has had a chance to browse through the comments, I usually ask for equal time as yet one more reader of the rough draft, and I react in writing both to the group's comments and to the writer's writing. What is returned to the panel of readers and then to the writer is a set of multiple voices talking to each other—in writing.

Since I am convinced of the validity of the workshop approach to the composition classroom, the revision that follows after the evaluation forms are returned to the writer goes on for several days in class. It is here (or in conferences) that the instructor becomes most directly involved in helping each individual student. Solutions for weak spots are discussed, alternative organizational patterns can be considered, or rules of grammar that are needed can be explained. Intensive work in grammar

---

7. Thom Hawkins, *Group Inquiry Techniques for Teaching Writing* (Urbana, Illinois: NCTE, 1976), p. 1.

is best left for this stage because errors in earlier drafts may disappear from the page as sentences are discarded or rewritten. Techniques for proofreading can also be offered at this time, if that is what is needed. In this stage of revision, then, the student has a more well-defined sense of what writing problem (or problems) he is trying to solve, and the instructor becomes a consultant who can offer from experience a wider range of suggested solutions than the student may yet have at his fingertips. The effect of this is to reverse the usual grading procedure because help is offered as a solution to a need, not as an *ex post facto* umpire's call. For example, the need for parallel structure in a series is usually marked as an error by the instructor and then revised and perhaps learned for future use by a student. However, compare this order of instruction (and its probable effectiveness) to the situation in which a student searching for an emphatic ending, or peroration, to his paper is offered some instruction in parallel structure, should he care to use it. Like most beginning craftsmen in the middle of coping with a demanding task, students are more receptive to new tools when they are offered in time to solve particular technical problems.

When the paper which results from this second stage of revision is handed in for a grade, the teacher's evaluation is both easy and quick. Rather than being confronted with an unknown, new product, the teacher is working with familiar content in which successful revisions and remaining difficulties are easier to spot. We can and should grade these revised papers throughout the semester to help students evaluate their work, but even these grades can be stages along the way to a final evaluation in a course where students are in the process of acquiring a skill. I have never been comfortable with the concept of assigning a course grade based on an average of those grades given during the semester because no matter what the student's entering skills were, his or her goal is to be a competent writer by semester's end. We can weigh the last few papers more heavily, but this puts undue stress on the writing performance evident in a small sample. One partial solution which, however, does not alleviate the problem of grading a small writing sample, is to allow students to spend the last week or so of the course revising several papers of their choice to submit as a final sample for consideration. By the end of the semester the student who has achieved some skill as a critical reader can go back over old papers to see problems or better solutions that weren't apparent to him earlier. At the end of the semester, when the student submits what he now considers to be his best effort, he is demonstrating the skills he has acquired by the end of the course.

I strongly believe, and am convinced by watching students' progress, that when evaluation is stressed as an on-going tool for revision, the student comes to the realization that not only is writing a process, but evaluation is too. The teacher's role as Super Critic dissolves as he becomes instead what the instructor of composition truly is, a tutor helping students as they learn how to write well. Extensive practice in evaluation through each stage from pre-writing to final draft helps the student to sharpen his skills as a critic of other writing, guides him as he revises, and demonstrates to him that, finally, evaluating his writing is *his* job.

Richard L. Larson

SELECTED BIBLIOGRAPHY OF WRITINGS ON THE EVALUATION OF STUDENTS' ACHIEVEMENTS IN COMPOSITION

I present here a selection from the large, and growing, body of literature on responding to and evaluating students' composing skills in courses and tests. The selection, obviously, is my personal choice—the books and essays that seem to me most conspicuously to contribute to our knowledge of how to judge student writing.*

I group my selections according to the purposes that seem to underlie the authors' work. First, and perhaps most helpful to the classroom teacher, come essays designed to advise teachers on how to react to students' writing as part of the instruction in a writing course. Though the teacher can probably never quite escape the role of authority figure and judge in dealing with students' papers, the essays in this group suggest how the teacher can act more as guide or coach than as judge— can suggest to the student what he or she has accomplished, or not accomplished, and what steps he might take to add to the accomplishments, either in revising the paper being looked at or in attempting the next one. The perspectives from which the guidance is given differ (my own piece, for instance, invites the teacher to look at the student's paper, as a whole, as a completed act of using language, and to suggest ways in which that act of using language for a particular purpose might be strengthened); different users of the bibliography will find different approaches to student papers congenial to them, and that varied response will be entirely appropriate, since none of the suggested perspectives or approaches has been shown by research studies to be the best, or a better, way to respond to students' work.

The second group of essays, quite small, deals openly with the task of making judgments about students' writing. Here the judgments discussed

are essentially administrative acts—decisions made about where to place student papers, and students, on scales that permit assigning the student to a particular class, declaring that he or she has or does not have a particular aptitude for writing or for academic study, determining that he or she is or is not demonstrating required proficiency in written English, or establishing the progress that he or she is making relative to a starting point (particularly if that progress will on some scale or other be deemed adequate, for example, to excuse the student from further work in writing). The emphasis, that is, does not fall in these essays on the teaching that can be done for the student, but only upon giving the student a bit of feedback in the form of a summary score, while enabling school administrators to reach a decision. In this group are two current discussions about testing procedures and procedures for scoring tests. Of these two, the more directly applicable to the work of the classroom teacher is the book by Diederich; the other item discusses and evaluates particular tests that teachers might employ.

The third group of essays listed deals, we might say, with the *processes* of measurement; these pieces investigate the theoretical and practical uses of various means for measuring and various scales or kinds of data that can be derived from testing. This group also contains discussions of national efforts to determine the skills in writing exhibited by students of different ages. The studies in this group do not necessarily advocate, for the classroom teacher or even for the administrator, the general use of the procedures discussed. Instead, as noted, they are contributions to the literature on the theory of testing, or to research on what tests can tell us and on how they can be scored. Much of the information contained here is now used to support, or one day may be used to support, the use of particular techniques to gather information, and thus in the future may affect the lives of teachers who do not now hear these techniques discussed in their schools or their neighborhoods. Knowing about the discussion of these testing procedures, therefore, may be of benefit as teachers talk about the evaluation of students. At the very least, familiarity with these pieces will let teachers see the issues now being debated among those concerned with evaluation, and thus let them sense the state of the art of evaluation.

For the teacher in basic writing, even more than for the teacher in regular writing courses and advanced courses, evaluation is a special concern because of the persistent need to determine when a student is ready to move onward in the sequence of writing courses or has satisfied some minimum standards of accomplishment. The items in this

bibliography will not offer easy answers for teachers who are responsible for these determinations, but they may equip these teachers to make better decisions about how to work with students, and may fortify them against capricious efforts to adopt judgmental techniques that have not themselves been fully investigated and evaluated.

I. SUGGESTIONS FOR CLASSROOM TEACHERS ON RESPONDING TO STUDENTS' WRITING

Bain, Robert, "Reading Student Papers," *College Composition and Communication,* 25 (October, 1974), 307-309.
> Offers suggestions about the judging of student papers based on the premise that the student's work should be evaluated mainly on how well he executes his purpose in dealing with his chosen subject.

Beaven, Mary H., "Individualized Goal-Setting, Self-Evaluation, and Peer Evaluation," in Charles Cooper and Lee Odell, eds., *Evaluating Writing* (Urbana, Illinois: National Council of Teachers of English, 1977), pp. 135-156.
> Focuses on formative evaluation of writing, and after outlining six major assumptions underlying the author's approach (including her attention to the links between development of a student as a writer and his or her overall development as a person), describes procedures for teachers to use in making comments on students' writing: to work with students on setting (and achieving) their own goals for improvement; to have students evaluate their own work; and to have students' work evaluated by peers working in groups. Offers a rationale for each procedure, and suggests questions that a teacher can help students to learn in using it. Draws extensively on published research, particularly from educational psychology and psychology of personal development and group processes.

Cooper, Charles R., "Responding to Student Writing," in Walter Petty and Patrick Finn, ed., *The Writing Processes of Students* (Report of the Annual Conference on Language Arts, State University of New York at Buffalo, 1975), pp. 31-39.
> Urges an emphasis in the composition curriculum on regular and frequent writing—in which students go through the complete process of composing—and recommends supportive, helpful re-sponses to students' work (including responses by other students). Particularly advises discussion of the student's rhetorical emphasis

(including audience, voice, and design) in the paper, the "intellectual strategies" employed, and characteristics of the syntax.

Larson, Richard L., "The Whole is More Than the Sum of Its Parts: Notes on Responding to Students' Papers," *Arizona English Bulletin,* 16 (February 1974), 175-181.
Suggests that in evaluating students' writing, teachers look at the success of the piece as a complete work, offers some questions to aid in making this assessment, and shows how the questions apply to some student writing.

Lamberg, Walter J., "Feedback on Writing: Much More Than Teacher Corrections," *Statement: The Journal of the Colorado Language Arts Society,* 12 (May, 1977), 33-38.
Defining "feedback" as "information about performance," reviews various theories about responding to student writing, cites the results of some research on the subject, and asks whether feedback should be different for different purposes and in different conditions.

Odell, Lee, "Responding to Student Writing," *College Composition and Communication,* 24 (December, 1973), 394-400.
Drawing on analytical techniques developed by Kenneth Pike and others, identifies some habits of thinking and responding to experiences that appear in students' writing, and suggests how teachers can help students develop new strategies and procedures for thinking about their subjects.

Kantor, Ken, "Evaluating Creative Writing: A Different Ball Game," *English Journal,* 64 (April, 1975), 72-74.
Proposes six criteria, drawn from psychologists' discussions of creativity, for the evaluation of students' creative writing, and applies the criteria to a brief story written by a student.

Shaughnessy, Mina P., *Errors and Expectations* (New York: Oxford University Press, 1977).
Working from an analysis of papers by thousands of students, classifies and interprets the "errors" found in these students' work, suggesting the sources of these errors—the reasons (or reasoning) that lead students to make them—and proposing ways of responding to them. Not primarily a book about testing or evaluation, but nonetheless a book that can help make the evaluation of papers containing errors become wiser and more humane.

II. SUGGESTIONS ABOUT TESTS AND MEASURES OF ABILITY IN COMPOSITION AND GROWTH IN COMPOSITION SKILLS

Braddock, Richard, "Evaluation of Writing Tests," in Alfred H. Grommon, ed., *Reviews of Selected Published Tests in English* (Urbana, Illinois: National Council of Teachers of English, 1976), pp. 118-126.
Cites the shortcomings of standardized tests of writing, taken as a group, and then comments on four specific tests, at least two of which are useful for their stated purposes.

Diederich, Paul B., *Measuring Growth in English* (Urbana, Illinois: National Council of Teachers of English, 1974).
Proposes procedures for the reliable evaluation of students' writing, and offers advice on how to assure the reliability of evaluation through the use of appropriate statistical procedures. Includes sample examination papers and a statement of criteria by which students' writing can be judged. Important book for those who are involved in testing and want to learn about interpretation of the statistics often given in manuals that explain the scoring of tests.

III. RESEARCH STUDIES, ESSAYS ON THE THEORY OF TESTING, AND DISCUSSIONS OF ISSUES IN TESTING AND MEASUREMENT IN COMPOSITION

Breland, Hunter M., "Multiple-Choice Test Assesses Writing Ability," *Findings (ETS)*, V, 1 (1977), 1-4.
Using data drawn from studies at Educational Testing Service, argues that the Test of Standard Written English is as useful as a written essay, and much easier to administer and score, for purposes of placing students in writing courses and thus for the planning of instructional activities. Suggests also that many students completing composition courses are not, in the judgment of scorers who worked on the study, producing satisfactory essays.

Britton, James, *et al.*, *The Development of Writing Abilities (11-18)* (London: Macmillan Education, 1975).
Reports on an extended research project conducted under the sponsorship of the Schools Council, concerning the kinds of writing done in school (in all subjects) by students aged 11-18. Constructs a fresh procedure for classifying writing according to the relationship

of writer to reader(s) and according to the function served by the writing. An important book for those interested in writing in the schools and in procedures for doing research on writing.

Cohen, Arthur M., "Assessing College Students' Ability to Write Compositions," *Research in the Teaching of English, 7* (Winter, 1973), 356-371.
Reports on how a group of two-year college teachers participated in an experiment to determine whether their students' ability to write well was improving in response to instruction, and on the instrument devised by the group to measure that improvement. Useful discussion of procedures for engaging classroom teachers in the conduct of research.

Cooper, Charles R., "Holistic Evaluation of Writing," in Charles Cooper and Lee Odell, eds., *Evaluating Writing* (Urbana, Illinois: National Council of Teachers of English, 1977), pp. 3-31.
Argues for the superiority of using an essay test rather than a multiple-choice test in the evaluation of writing, and then reviews a number of holistic procedures (to be differentiated from the making of counts of particular elements of syntax, diction, mechanics, and so on), showing how each procedure should be managed. Discusses the use of scales made up of graded complete essays, feature analysis (judging one feature of the writing only), primary trait scoring, general impression marking, and "center of gravity response" (the term is from Peter Elbow's *Writing Without Teachers,* London: Oxford University Press, 1973), before discussing the development and use of analytic scales (where the desired qualities of a piece of writing are enumerated, and the characteristics of high, average, and low papers are specified). Explains the procedures essential to effective use of analytic scales, and includes illustrations of such scales.

Fagan, William T., Charles R. Cooper, and Julie Jensen, "Measures: Writing," in *Measures for Research and Evaluation in the English Language Arts* (Urbana, Illinois: National Council of Teachers of English, 1975), pp. 185-206.
Lists, describes, and gives data on validity and reliability about fourteen measures useful in research on writing, many of them not previously published. Includes tests, evaluation scales, indices of such features as syntactic maturity, analytical tools, and a corpus of

American expository essays. Likely to be of value to persons engaged in serious research about writing and the teaching of composition.

Godshalk, Fred I., Frances Swineford and William E. Coffman, *The Measurement of Writing Ability* (New York: College Entrance Examination Board, 1966).
Reports on an extended study to determine which of three kinds of tests—a series of objective tests, a series of objective tests together with an interlinear exercise (in which the student is given a passage of prose with errors, and is asked to indicate corrections between the lines of the passage), and a series of objective tests together with a short essay—gives the most reliable predictor of students' writing ability. Concludes that the objective tests with essay are the most valid predictors of writing ability, but supports the assertion that scores on objective tests alone, if the tests are well chosen, are themselves valid predictors. Discusses the establishment of a criterion against which to judge the validity of the various kinds of tests. A complex study, buttressed with extensive statistical data and tables.

Hunt, Kellogg W., "Early Blooming and Late Blooming Syntactic Structures," in Charles Cooper and Lee Odell, eds., *Evaluating Writing* (Urbana, Illinois: National Council of Teachers of English, 1977), pp. 91-104.
Reports on some recent research which confirms that older children and adults in general have more words per T-unit and more embedded S-constituents (analogous to what were formerly called kernel sentences) per sentence, and enumerates kinds of syntactic structures that appear more frequently in older writers than in younger writers. Comments on the implications of these findings for research and for the making of curricula.

Illo, John, "From Senior to Freshman: A Study of Performance in English Composition in High School and College," *Research in the Teaching of English,* 10 (Fall, 1976), pp. 137-146.
Reports on a study of the correlations between various factors and indices (including the results of a questionnaire concerning home activities, high-school instruction, and family) concerning freshmen's ability at an experience with writing, and measures of performance in writing in college courses. Suggests that the emotion, purpose, and will of the student is more directly responsi-

ble for the success of the student in college courses than other factors, such as intelligence and quality of preparation in writing. Explores the implications of this argument.

Lloyd-Jones, Richard, "Primary Trait Scoring," in Charles Cooper and Lee Odell, eds., *Evaluating Writing* (Urbana, Illinois: National Council of Teachers of English, 1977), pp. 33-66.
Differentiates "holistic scoring" from "atomistic scoring," and then, citing the theories of discourse underlying the procedure, defines "primary trait scoring" as, essentially, the judging of how well the writer responded to the audience, purpose, and occasion stipulated in doing a writing exercise or test. Describes procedures for developing exercises with which primary trait scoring can be used, and describes the preparation of scoring guides for these exercises. Includes illustrative exercises and the (occasionally complex) scoring guides that accompany them. Suggests applications for primary trait scoring in research and teaching.

Mellon, John, "The Writing Assessment," in *National Assessment and the Teaching of English* (Urbana, Illinois: National Council of Teachers of English, 1975), pp. 14-38.
Describes the first round of tests of writing conducted by the National Assessment, reports some of the findings, and evaluates both the testing procedures and the findings. Notes particularly the suggestions in the assessment that students in high school may not be improving greatly their ability at composing. Notes, too, the the imprecision of phrasing in the writing assignments used in the first round and offers suggestions about the effective focusing of assignments. Finally, notes the absence of attention, in the first round of testing, to expressive writing and writing that deals with feelings. Useful introduction to the accomplishments of the National Assessment in writing and to its problems.

Mellon, John C., "Round Two of The National Writing Assessment— Interpreting the Apparent Decline in Writing Ability: A Review," *Research in the Teaching of English,* 10 (Spring, 1976), 66-74.
Examines a report of the National Assessment on "Writing Mechanics: 1969-1974." Notes the deficiencies in the procedures used by the National Assessment, and urges caution in interpreting the data presented, but accepts the general conclusion that there has been a decline in the writing ability of students (in the age groups

considered) in the five-year period. Notes that some of the decline is perhaps due to lessening of attention to writing in the teaching of English in schools. Poses policy questions regarding the role that writing should play in the curriculum.

Moslemi, Marlene H., "The Grading of Creative Writing Essays," *Research in the Teaching of English,* 9 (Fall, 1975), 154-161.
Although reporting an experiment with ninth-graders in which pre- and post-tests surrounding an intensive unit (using audio-visual materials) on creative writing showed no significant improvement in students' ability at creative writing, the article does describe procedures for selecting and training judges of creative writing that resulted in high reliability of scores among the judges.

Odell, Lee, "Measuring Changes in Intellectual Processes as One Dimension of Growth in Writing," in Charles Cooper and Lee Odell, eds., *Evaluating Writing* (Urbana, Illinois: National Council of Teachers of English: 1977), pp. 107-132.
Recapitulates the analysis by Pike and others of the intellectual processes in which people engage, citing the acts of focusing, contrasting, classifying, noting change, relating change, relating events to physical context. Then suggests how students' writing can be analyzed to discover the kinds of processes at work in each piece and the frequency of their occurrence, suggesting that such analysis helps the diagnosis of students' writing problems. Suggests that measures of change in the use of these processes may be important to a comprehensive evaluation of growth in writing.

Purves, Alan, and the Task Force on Measurement and Evaluation in the Study of English. *Common Sense and Testing English* (Urbana, Illinois: National Council of Teachers of English, 1975).
For persons not much acquainted with testing practices and procedures, describes main concepts in testing, kinds of tests, kinds of scores, and uses to which test results can be put, along with ways of defining the uses and limitations of different tests so that persons interested in employing the tests can understand what they are up to, and proceed more wisely than they might otherwise do.

Sanders, Sara E. and John H. Littlefield, "Perhaps Test Essays Can Reflect Significant Improvement in Freshman Composition," *Research in the Teaching of English,* 9 (Fall, 1975), 145-153.

Points out weaknesses in the procedures typically used to test improvements in composition—through impromptu pre- and post-test themes—and reports a study in which students *did* demonstrate improvement when allowed, on both pre-and post-test, to engage in research, undertake pre-writing activities, and revise the first drafts of their essays. Differentiates between the standard "expository" approach to the teaching of writing and the "aims" approach (following Kinneavy's *A Theory of Discourse*), but in this experiment reports significant improvement between pre- and post-test for students working with both approaches.

CALL FOR ARTICLES

VOCABULARY

The editors invite articles which discuss successful methods of teaching vocabulary to basic writing students. Articles should justify the choice of methods, analyze basic writing students' central difficulties with words, and discuss the features of academic language that pose the most serious problems for basic writing students.

Deadline for articles: January 30, 1979.

Articles should be no more than 6,000 words (about 20 pages). Please follow the MLA Style Sheet, second edition, for matters of form. Include all footnotes at the end of the article. Enclose two copies of the article and a self-addressed stamped envelope. Manuscripts and correspondence should be addressed to: The Editors, Journal of Basic Writing, Instructional Resource Center, 535 E. 80th Street, New York, New York 10021.

JOURNAL OF BASIC WRITING