

# Evaluating Ninth Grade Essays

Edward T. Hill and Robert Wesolowski  
Churchill High School  
Livonia, Michigan

Many of us are concerned about student writing skills, or the lack of them. We cite verbal scores on the Scholastic Aptitude Test which have declined for 18 years; we quote *Readers' Digest* (April, 1981), *Media and Methods* (March, 1980), and the *English Journal* (December, 1980) which call for the improvement of student writing skills. Although we are able to agree that student writing skills have declined, we are unable to agree how to measure those skills whose weaknesses we bemoan. Some of us who are classroom teachers in Livonia, Michigan, shaped our own effort to evaluate student writing samples from 348 ninth graders in an effort to diagnose our students needs and to provide them with meaningful instruction to meet those needs.

Our work began when we read research on the composing process by Moffett, Britton, Emig, Cooper, Odell, and others. From our reading, we developed a comprehensive writing curriculum (remedial, regular, advanced) for tenth-, eleventh-, and twelfth-grade students. Then we worked to develop a screening process through which all our ninth-grade students would pass before being assigned to an appropriate writing class.

We required each ninth grader to write an essay which was read holistically by two teachers from a pool of six — four English teachers and two social studies teachers. If there were disagreements between scores given by the first two readers, a third reader would read the essay and resolve the difference. Based upon the evaluations of these essays and recommendations to counselors, students were placed in writing classes.

We based our evaluation procedure on the evaluation practices of the English Composition Board (ECB) at The University of Michigan and Paul Diederich's belief that trained readers, spending three or four minutes per paper, can provide reliable decisions when rating student papers as demonstrating low-, middle-, or high-quality writing. (Fig. 1).

We adapted Diederich's scale, using a nine-point scale instead of a five-point scale of evaluation, because it served our needs better. If a student scored 1-3 in several categories on the scale, the student was recommended for placement into a remedial section. With a score of 4-6, the student was recommended for a regular section. Students who scored in the 7-9 range were recommended for advanced writing sections. (Fig. 2).

If one of our raters gave a paper a 5 for organization, and another gave the same paper an 8 for organization, we considered that a disagreement because the scores are more

Figure 1

Diederich's Rating Scale*					
Topic	Reader _____		Paper _____		
	Low	Middle	High		
Ideas	2	4	6	8	10
Organization	2	4	6	8	10
Wording	1	2	3	4	5
Flavor	1	2	3	4	5
Usage	1	2	3	4	5
Punctuation	1	2	3	4	5
Spelling	1	2	3	4	5
Handwriting	1	2	3	4	5

\*Teachers circle one number after the name of each topic. Teachers give double weight to ideas and organization by doubling the numbers representing them on the chart (Diederich, p. 54).

than two digits apart. In cases of disagreement a third rater read the paper, rated it, and assigned the paper to a cate-

Figure 2

Churchill Rating Scale*										
Date _____	Reader _____									Student I.D. _____
Ideas	1	2	3	4	5	6	7	8	9	
Organization	1	2	3	4	5	6	7	8	9	
Clarity	1	2	3	4	5	6	7	8	9	
Style	1	2	3	4	5	6	7	8	9	
Wording, Phrasing	1	2	3	4	5	6	7	8	9	
Grammar, Structure	1	2	3	4	5	6	7	8	9	
Punctuation	1	2	3	4	5	6	7	8	9	
Spelling	1	2	3	4	5	6	7	8	9	
Handwriting	1	2	3	4	5	6	7	8	9	

\*Because we worked with a ninth-grade population, not separated into academic levels, we expanded the rating scale to run from 1 to 9; we added the term "clarity" to help explain Diederich's term "wording." We used the term "style" in place of Diederich's term "flavor." Our teachers were happier with the term "grammar" instead of "usage."

gory determined by the two closest scores. Remarkably, we agreed 80 percent of the time which meant that only 19 percent of the essays had to be read a third time.

We realize, of course, that there are limitations to analytic writing scales. For example, depending on their mental attitudes, students, on any given day, may write well or poorly. In an effort to minimize some of the limitations of our testing, we asked all ninth graders at Churchill to list four or five topics they felt most comfortable writing about. The topics students contributed suggested that they wanted to write in a narrative mode about their own experiences.

Guided by students topics and materials developed at the ECB, we developed a writing stimulus to be administered by every English teacher who taught ninth graders during the second week of February, 1981. We were able to collect about 90 percent of the samples (348) in 12 class meetings. Students did not place names on their papers. Instead, they placed their five digit student identification numbers in the upper right hand corner of the paper so no rater could identify the student during the reading. Later we were able to identify the students for placement in appropriate writing classes by the students' identification numbers. Students and teachers were told that no papers would

be returned to them, but they were invited to discuss their scores, by making an appointment to review their work if they chose to do so.

In our efforts to evaluate student writing as a first step to providing our students the instruction they need, we were strongly supported by our principal, Bill McFarland. He provided us three half-day substitutes at a cost of \$315 so that the six-member team might rate papers and an additional \$192 to compensate teachers for working after school to complete the project. The total cost of the project was \$507, about \$1.45 per student.

What did we discover during the project? We discovered (1) that after brief instruction, it does not take long for experienced teachers to use a rating scale effectively to evaluate students' essays; (2) teachers who are trained will agree 80% to 90% of the time when they read the same paper; (3) 13% of our ninth-grade writing samples were judged low-quality; (4) 72% were considered middle-quality; and (5) 14.6% were considered high-quality.

During the next three years we hope to conduct follow-up studies to discover the effect of assessing student writing at the outset of their high school careers in order to provide students the instruction in writing which they require.