RESEARCH IN WRITING, SECONDARY SCHOOL, 1984-2003

GEORGE HILLOCKS, JR

The University of Chicago

Abstract This review covers what is known in the U.S. as "secondary school," generally encompassing grades 7-12. The author frames the review by looking at the broader assessment context, particularly state-wide writing tests that often trivialize writing by requiring writing within severe time restraints on topics that may be of little interest to students and that may benefit students with from privileged social backgrounds. Further, these assessments reduce writing to limited forms such as the five-paragraph theme, even when the genre called for (e.g., narrative) may not be amenable to such forms. The review finds that assessment mandates in turn affect classroom writing instruction in what the author characterizes as negative ways, emphasizing the mastery of a generic form over the generation of ideas. The review concludes that, in spite all of the attention given to writing instruction, writing is not necessarily improving, in large part because of mandates for how writing is assessed.
Key words: writing research, writing assessment, teacher education

Chinese 《中学写作的研究，1984-2003》George Hillocks, Jr.

撮要

本回顾探讨的范围包括了美国一般所说的中学，一般就是指第七至十二级。作者的回顾架构，着眼于较阔的评估内容，特别是全国性的写作测试。这测试往往把写作简单化，要求学生在几个不同的时限内，就几个学生可能不太感兴趣，而且可能对享有特殊社会背景的学生较有利的题目来写作。再者，这些评估把写作局限于某些型式，例如五段式写作主题，即使有些文体（例如：记叙文）并不适合使用这种写作型式。本回顾发现，这种作者定性为负面的评估型式，反过来影响了课堂写作教学，强调写作语体的掌握多于产生写作点子。总的来说，尽管所有注意力都放在写作教学，写作能力并不必然地进步，原因主要是受制于写作的评估方法

L1 – Educational Studies in Language and Literature (2006) 6 (2), p. 27-51.

Correspondence concerning this article should be directed to George Hillocks, Jr., 2012 West 110th St. Chicago IL 60643,. Electronic mail may be sent to ghillock@uchicago.edu.

（Abstract translated into Chinese by Shek Kam Tse.）

French résumé. Cette revue de question couvre ce que l'on nomme aux Etats-Unis " l'enseignement secondaire " incluant généralement les classes 7 à 12. Elle s'articule autour de l'examen des procédures d'évaluation, en particulier des tests d'écriture nationaux qui banalisent souvent la production écrite en demandant d'écrire dans des contraintes de temps sévères sur des sujets qui peuvent présenter peu d'intérêt pour les élèves et qui sont susceptibles de profiter davantage à ceux issus des milieux sociaux favorisés. De plus, ces évaluations réduisent l'écriture à des formes limitées comme la réaction à cinq paragraphes, même quand le genre attendu (par exemple le récit) ne se prête pas à de telles formes. L'auteur constate que, en retour, ces évaluations affectent l'enseignement de la production écrite de façon négative, favorisant la maîtrise d'une forme générique au détriment de la génération d'idées. La revue de question conclut que, malgré toute l'attention donnée à l'enseignement de la production écrite, l'écriture ne s'améliore pas nécessairement, en grande partie à cause de la façon dont elle est évaluée.
(Abstract translated into French by Laurence Pasa.)
Mots clefs recherche en écriture, évaluation de l'écriture, formation d'enseignant

Portuguese resumo. Este levantamento cobre aquilo que nos Estados Unidos é o ensino secundário, entre o 7º e o 12º ano. O autor enquadra o seu levantamento numa análise do contexto alargado da avaliação, em particular das provas nacionais que geralmente banalizam a escrita através do pedido de produção sob constrangimentos temporais estritosd e sobre assuntos que podem não ser interessantes para a generalidade dos alunos, podendo acabar por beneficiar os alunos de meios sociais privilegiados. Além disso, estas avaliações reduzem a escrita a modalidades limitadas, como a "redacção de cinco parágrafos", mesmo quando a tipologia solicitada (por exemplo, a narrativa) não se enquadra nessas modalidades. O levantamento conclui que esta avaliação, por sua vez, afecta a instruções de escrita em sala de aula de um modo que o autor considera negativo por enfatizar o domínio de tipologias genéricas em vez de privilegiar a geração de ideias. O levantamento conclui que, apesar de toda a atenção dada ao ensino da escrita, a qualidade da escrita não está necessariamente a aumentar, em grande parte devido ao modo como ela é avaliada. (Abstract translated into Portuguese by Paulo Feytor Pinto)
Palavras-chave: investigação sobre escrita, avaliação da escrita, formação de professores.

1. INTRODUCTION

Research produced on writing in the secondary school from 1984 to 2003 has changed in many ways from that produced in the decades of 1963 to 1983. From '63 to '83 many studies focused on the syntax of student writing, its changes across age levels and under varying conditions, its relationship to quality of writing, and the impact of certain kinds of instruction (i.e., sentence combining) on the syntax of student writing. Many quasi-experimental studies focused on the impact of various foci and modes of instruction on student writing. Some of these were carefully designed, meeting fairly rigorous criteria for useful research, but many more were not. Many studies focused on the general processes of writing. Many more rigorous studies focused on cognitive processes involved in planning, drafting, and revising.

The theoretical bases for these studies were varied. Because of the many studies of syntax, various linguists and applied linguistic researchers names appeared, among them Chomsky being the most frequent. Other studies referred to the sources of an instructional idea being investigated and to Dewey and Piaget.

The quantity of research in writing appears to have reached a pinnacle by the early eighties if we judge by numbers of studies published in Research in the Teaching of English (RTE) but to have diminished thereafter. In 1984 RTE published 16 articles, 13 of which deal with writing (81%). In 1999, fifteen years later, it published 13 articles of which two focused on composition (15%). Research interests

have also shifted. Interest in the syntactic features of writing as a primary focus has all but disappeared. Langer's study of writing development (1986) uses measures of syntactic complexity as one of many measures that include the complexity of reasoning, the elaboration of responses, the degree of self-awareness, and so forth. Quasi-experimental studies have sharply diminished. Studies of cognitive processes are fewer, though notable publications are of important value, e.g., Bereiter & Scardamalia, 1 and Hayes, In place of such studies at the secondary level, we have studies that look in detail at response to student writing, at teacher-student writing conferences, peer group conferences, the idea of role in class discussion, how teachers acquire knowledge of writing, the education of writing teachers, the concept of composing across disciplines, and many other dimensions of teaching and learning that were not investigated earlier.

Further, many more studies in the recent decades use qualitative or ethnographic methods that are able to provide levels of detail not possible in earlier research. Quantitative studies have not disappeared, but they no longer dominate studies of writing as they once did. The growth of qualitative research in writing appears to go hand in hand with researchers', if not teachers', perceptions of writing as socially situated and with the theorists they use to elucidate their ideas, primarily Vygotsky (1978) and Bahktin (1981). Indeed, these two theorists seem to have had enormous impact on some of the most important studies of literacy, e,g, Nystrand et al (1997) and Langer (2001), Johnson, et al (2003), and Lee and Smagorinsky (2000).

The developmental theory of Piaget which was so important 20 years before has all but vanished from the pages of research. The disappearance indicates a profound shift in thinking about writing. In the sixties and seventies many researchers and theorists, adopting the theories of Piaget and others, assumed that "if a child's mental functions (intellectual operations) have not matured to the extent that he is capable of learning a particular subject, then no instruction will prove useful" Vygotsky, 1978, p. 80). But Vygotsky makes a strong case that "learning results in mental development" and makes development possible (p. 90). He argues that learning takes place in the "zone of proximal development," defined by Vygotsky as "the distance between the actual developmental level as determined by independent problem-solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" (p. 86). The trick, then, becomes not one of waiting for the learner to develop and for learning to appear naturally but of finding ways to promote learning in the zone of what the student is capable of doing with help so that development may take place. While this may be the single most important shift in the psychological thinking about learning to write, few researchers and curriculum makers have attended to the shift. It appears to be beyond their ken.

The review that follows is selective because of limitations imposed by the original plan and by the current publication. This review confines itself to research on assessment, teaching of writing in the schools, approaches to teaching, and finally education of teachers of writing.

2. WRITING ASSESSMENT

In the two decades from 1963 to 1983 researchers were dealing with the problem of how to judge writing fairly, so that a writer's score would not be dependent on the reader. In some early studies, researchers found that most papers scored received every possible rating from different raters. Because the raters had different experiences and backgrounds, they had different responses to student writing. Some seemed to care most about mechanics, some about organization, others about style or content (Diedrich, 1974). It quickly became clear that if assessors of writing were to have agreement, they would have at least to share the same criteria for judging writing. By the late seventies several different kinds of scales were adopted for assigning ratings to student writing.

Three kinds of scales have been used with high rater agreements. Some researchers see the drive for agreement as problematic in that it may obscure qualities that researchers and teachers need to attend to (Huot, 1990). One kind of scale is called holistic because it represents the quality of a piece of writing as whole. It is usually guided by a scoring rubric listing criteria for assigning scores for each level of quality. The National Assessment of Educational Progress (2002), for example, presents the following description of an excellent eighth grade response to its narrative prompt:

- Tells a clear story that is well developed and shaped with well-chosen details across the response;
- Is well organized with strong transitions;
- Sustains variety in sentence structure and exhibits good word choice;
- Errors in grammar, spelling, and punctuation are few and do not interfere with understanding.

The description is quite general. It does not, for example, explain what counts as "well developed and shaped with well-chosen details." But ordinarily, the descriptions are accompanied by compositions that exemplify each level of competence.

White (1985) argues that holistic ratings have the virtue of representing the piece as a whole as opposed to analyzing the parts in a reductionist way and that such scoring has "satisfied reasonable demands for both economy and reliability and ha[s] led the way to restoring the role of writing in testing" by allowing us to avoid the reductionism of multiple choice tests. On the other hand, he points out that such scoring can only rank order the writing in a given test, relative to the test and the context in which it was administered. It provides little specific information about the characteristics of a particular piece of writing.

More specific information can be provided by primary trait scales. White suggests that holistic scales are closely allied to primary trait scales, the difference being that the primary trait scales define the criteria more explicitly in terms of whatever dimensions the assessor wishes to examine. In fact, White states that the two are "conceptually the same"(p. 142). However, a primary trait scale allows for the consideration of particular traits rather than an overall impression. For example, it might allow scoring the use of specific imagery in narrative writing and exclude attention to mechanics, sentence structure, and spelling and other matters not relevant to specific imagery. For teachers, primary trait scales have the advantage of

allowing concentration on certain aspects of writing while ignoring others. As White points out, writing is so complicated that many teachers concentrate on the specific features of writing, one at a time, e,g., the use of evidence in making a case or argument.

Most scales in state assessments are holistic. Few use primary trait scales. Some use what are called analytic scales. These differ from holistic and primary trait in asking raters to judge several important traits of the writing individually, usually on sub-scales. Illinois (1994) used an analytic scale with ratings for organization, elaboration, mechanics and so forth. The paper score is the sum of ratings on the subscales. Such scales tend to be more difficult to use, more time consuming, and therefore, more expensive.

2.1 State Writing Assessments

In the 1980s, in the United States, certain states began the direct testing of writing, using samples of writing rather than objective tests. At present nearly all of the fifty states require such writing exams. While it is arguable that this nation-wide testing has had a major impact on the teaching of writing across the country, little research has been conducted to examine its effect. Hillocks (2002) studied the impact of state assessments intensively in five states: Illinois, Kentucky, New York, Oregon, and Texas through examination of state documents, curricular materials, teacher handouts, and published materials for the purpose of helping schools do better on the exams. In each state, he conducted interviews with over sixty teachers and over 20 administrators or supervisors in two urban, two suburban, one small town and one rural district.

According to Hillocks (2002), nearly all of the state assessments require that students write on demand in a limited time period. A few states refer to portfolios of writing, but only in Kentucky is the portfolio score a major component in the formula for assessment of the individual school.

Nearly all states announce that students at various levels should be prepared to write narrative, expository, or persuasive prose. Some add descriptive. Most offer no or only very shallow rationales for the choices. Oregon offers no rationale, for example, while Illinois provides a general statement that three types represent all three domains of writing (ISBE, 1994). New York does not present a theory of kinds of writing so much as a theory of language that includes speaking and listening. As a result, the Regency exams at the eleventh grade includes one section in which students listen to a passage read aloud and write a response.

Both Texas and Kentucky present more elaborate theories of writing. Texas presents an analysis of writing based on a theory of discourse developed by Kinneavy (1971), which posits four purposes (informative, persuasive, literary and expressive) and four modes of discourse (narrative, descriptive, classificatory, and evaluative) for a total of sixteen possible types of writing. Texas focuses on four types for testing. The Texas Education Agency limits these even further. For example, Texas fourth and eighth grade teachers refer to the “informative narrative” prompt as the

"How-to," because it always asks students to write about how to do something, e.g., how to make a gift for a parent.

The Kentucky portfolio assessment is based on the theory of discourse advanced by Britton and his colleagues (1975). The result is that writing for the portfolio must include several types of writing including literary (poems, stories, children's books, plays, etc.), personal (narratives, memoirs, etc.), transactional (arguments, proposals, historical pieces, research focused papers, and so forth), and a reflective piece addressed to the reviewer addressing the writer's views of his or her development as a writer.

With the exception of Kentucky, testing conditions are similar in most states. In each grade students have a limited period of time (e.g.,forty minutes) to respond to a prompt which indicates an issue such as extending the school year. The prompt typically gives a little background, suggests an audience to write to and no more. Students are on their own. Most of the state writing assessments use very general prompts chosen because the examiners assume that students will have an opinion that they can support on the basis of general knowledge. Directors of state assessments in Illinois, Oregon, and Texas say that they cannot hold students responsible for content, that is, specific knowledge. Therefore, specific knowledge is not considered in the ratings.

In contrast, the New York Regents exams, in three of the four essay prompts, provide material for students to draw upon in their writing. For the fourth, the exam asks students to write about a literary work they have read. The Oregon writing exam is spread over three days, suggesting the possibility for students to gather information. In Kentucky, the problem is obviated by the nature of the assessment. Students are supposed to choose their own topics and they have plenty of time to collect adequate information.

For all assessments, scoring of compositions is based on a rubric similar to those for the NAEP. Hillocks argues that the benchmark papers used to illustrate the rubric, reveal that very low levels of writing suffice to pass the exam, especially in states that ask students to produce a piece of writing with no access to information in a restricted time period. Yet the vast majority of teachers interviewed by Hillocks indicate that they used the models from the scoring guides to teach writing. It is obvious that students are receiving a diet of poor writing that cannot provide appropriate nourishment for their growth as writers.

Hillocks concludes that the state assessments affect the standards for good writing adopted by teachers, the kind of instruction offered, and the writing curriculum available to students. Few teachers had had advanced training in writing and largely accept the standards set by the state. When asked how well they thought their state assessments supported a writing program that was desirable in their state, some teachers had no response. However, in each state a majority of teachers responding thought that their state assessment did support a desirable writing program. This is particularly true in Illinois and Texas, with the kind of assessment most likely to encourage formulaic writing. In Texas, for example, 61% of the teachers interviewed thought the assessment provided support for a desirable writing program; only 24% thought it did not.

3. STUDIES ABOUT ASSESSMENT RESULTS

A number of studies have been based on data collected in assessments, either on the scores attained or on the writing samples themselves. One group of studies has examined the various test score gaps in attempts to determine what the causes might be and whether they persist. Others have attempted to determine what factors might account for the quality ratings of holistic and primary trait scores.

3.1 Test Score Gaps

The assessments themselves and other researchers have identified three kinds of gaps between various segments of the general population; race and ethnicity, social class, and gender. These differences are substantial, on the order of a full standard deviation or more. According to Jencks and Phillips (1998), "the typical American black.... scores below 75% of American whites on most standardized tests. On some tests the typical American black scores below more than 85% of whites" (p. 1). In a footnote Jencks and Phillips state that similar results hold for Hispanics and Native Americans.

Scholars have attempted to investigate the variables responsible for the differences in test scores. According to Jencks and Phillips (1998), these variables have included segregation, poverty, the mother's background including her test scores and years of schooling, school resources including teachers' test scores, various kinds of test bias, and many others. These attempts have been largely unsuccessful in explaining much of the variance. The test score gap holds for writing as well. The National Assessment of Educational Progress Writing Report Card for 1998, for example, indicates that while only 10% of white eighth graders score at the lowest level of writing skill, 28% of blacks and 31% of Hispanics do. Conversely, while 32% of white eighth graders score at the "proficient " level, only 8% of black and 11% of Hispanic eighth graders do. Hedges and Nowell (1998) argue that while the test score gap has narrowed it remains large. In examining the data from NAEP writing tests from 1984 to 1994, they found that the differences between blacks and whites in writing ranged from .86 to .67 standard deviations. According to Cohen (1977), differences of .2 standard deviations are small, of .5, medium and .8 and over, large. By that criterion, the differences in writing scores are substantial.

A second kind of gap appear in relation to variables that attempt to measure some aspect of socio-economic status (SES). The NAEP uses three indices of SES: eligibility for free or reduced price lunch, participation in Title I schools, and student-reported highest level of parent education. NAEP (2002, pp. 53-56) clearly indicates that writing performance is strongly related to SES, whether estimated by eligibility for the free/ reduced-price lunch, participation in Title I schools, or by level of parent education. In both eighth and twelfth grades, students in the free or reduced-price lunch group or who participate in the Title I schools fall substantially short of the levels of achievement in higher SES groups. Eighth graders and twelfth graders not classified in the lower SES groups score at proficient or above nearly two and a half times as frequently as students in the lower groups. Even so, it seems remarkable that only 43% of eighth graders and only 32% of twelfth graders who

say their parents graduated from college, score at the level of proficient or above. If their reports are true, these students are likely to be in the highest income brackets, suggesting that a large proportion of our most advantaged students do not respond well to the writing assessment.

A third startling sub-group comparison is that between males and females. In NAEP writing assessment for 2002, 21% of eighth grade boys scored at or above proficient, while twice as many females scored at the same level, 42%. Results for twelfth graders are similar, but the gap is greater: 14% of boys score at proficient or advanced while 33% of girls do. Purves (1992) in reporting the IEA results states that "there is a widespread gender bias favoring girls that cuts across languages, cultures, and stages of economic development" (p. 146).

Hedges and Nowell (1995) studied gender differences in mental test scores, their variability, and the numbers of high scoring individuals. They "performed secondary analyses on six large data sets collected between 1960 and 1992" each of which "used a nationally stratified probability sample of adolescents" (p. 42). One of the data sets was from the NAEP for those years. Across the areas of reading, mathematics, science, and writing, they examine the difference between boys' and girls' scores as effect sizes (the difference divided by the standard deviation of the population). They comment, "Average sex differences were small except for writing, in which females performed substantially better than males in every year. Although average sex differences in mathematics and science have narrowed over time, differences in reading and writing scores have not" (p. 44). They view this situation with more alarm than do most composition teachers of my acquaintance.. "The large sex differences in writing ability suggested by the NAEP trend data are alarming.... The data imply that males are, on average, at a rather profound disadvantage in the performance of this basic skill."

As far as I know, no one has compared these data across all indices of race/ethnicity, SES, and gender, perhaps because the results would be overwhelming. My guess is that White female off-spring of college graduates would have far and away the highest scores and that Black males from poor families would trail far behind. It is not possible to exaggerate the danger of this disparity.

4. THE TEACHING OF WRITING IN SECONDARY SCHOOLS

Applebee (1981, 1984) studied writing in American secondary schools across subject matters. The study consisted of three parts: observations in two Midwestern high schools, a nation-wide questionnaire study of a stratified random sample of ninth and eleventh grade teachers, and case studies of students in the schools observed.

Although writing was a major presence in all subject matters, taking up "an average of 44% of the observed lesson time" (1981, p. 30), researchers found that

> students were spending only about 3% of their school time – in class or for homework – on writing of paragraph length or longer. On the other hand, students were engaged in a variety of related activities that involved writing but not composing: fill-in-the-blank exercises, worksheets requiring only short responses, translation from one language to another, and the like.

> Even in those contexts where students were being asked to write at some length, the writing often served merely as a vehicle to test knowledge of specific content, with the teacher functioning primarily as an examiner (1984, p. 2).

For the most part, as Applebee explains, "even when students were asked to write an essay, the essays were treated as tests of previous learning. The task for the student was one of repeating information that had already been organized by the teacher or textbook, rather than of extending or integrating new learning for themselves" (1984, p.3). Applebee points out that the topics assigned are a good indication of this approach to writing. "In many cases students were asked to write on topics that were in a real sense impossible" (1984, p.3). He provides an example from social science that asks students to "describe the political, economic, social and cultural changes that Europe was going through at the time of the Reformation." Applebee comments, "Books could be written in response to such a question. It becomes a possible topic for school writing only because it serves to index bodies of previously presented information" (1984, p. 4).

Analysis was the most frequent kind of school writing both in teacher reports and in samples submitted. In the national survey, 41.7% of English teachers, 42.9 % of science teachers and 49.5% of social science teachers report using analysis frequently. The reported frequent use of theory drops to 20.3 for English and 16.8 for social science teachers while remaining a fairly high 41.7% for science teachers. On the other hand, only 3% of the writing samples submitted were categorized as theorizing.

The teaching of writing in most cases studied by Applebee appeared to be little more than the making of assignments. "In the observational studies, the amount of time devoted to prewriting activities amounted to just over three minutes. That included everything from the time the teacher started introducing the topic until the first student began to write" (1981, p. 74). The "most popular technique of helping students get started was to have them begin their writing in class, so that they could ask questions about what was expected if they found themselves in difficulties" (1981, p. 78). Model pieces of writing were reportedly used in 29% of the classes as a means of "introducing new forms of writing" (p. 78). "Finally," according to Applebee, "brainstorming... was reported in use by some 37% of English teachers, and by no more than 14% in any of the other subject areas" (p. 80).

Applebee indicates that activities designed to help students during writing were "almost nonexistent" (p. 90). Just under 33% of teachers asked for more than a single draft. Applebee suggests that may be because so much of the writing functioned as tests of subject matter knowledge rather than explorations of new material.

Applebee concludes that "the major vehicle for writing instruction, in all subject matter areas, was the teacher's comments [on] and corrections of completed work. Errors in writing mechanics were the most common focus of these responses; comments concerned with the ideas the student was expressing were the least frequently reported" (pp. 90-91).

Even the textbooks appear to support these findings. Applebee examined three of the most popular of the textbooks in each of the subject areas studied. He found that

> The writing experiences provided in high school textbooks are narrow and limiting, whether one examines the role of the activity within the learning process or the kind of writing task the student is being asked to undertake....
>
> The types of activities suggested were also limited. Textbooks in all subjects seemed to be constructed around a base of exercises that required only minimal writing: fill-in-the-blank exercises, short answer responses, and the like. Some subjects-literature and the social sciences in particular-supplemented this base of restricted activity with more extensive writing tasks (1984, p. 35).

However, even these more extensive writing tasks are no more than assignments. One cited by Applebee, asks students to write their own blues song, presumably after they have read some in the literature text. Another suggests that students write a modern version of a story in which a character sells his soul to the devil. The problem with such assignments is that they assume that reading the examples is sufficient preparation for all students to do the assignment.

In short, a quarter of a century ago, when Applebee conducted the study, writing was widely treated in very superficial ways. Teachers appeared to assume that very general knowledge of writing would suffice for most purposes. The tasks of learning remain simple and uncomplicated.

Two and a half decades after Applebee conducted this study, is the teaching of writing in secondary school any different? It is difficult to tell. There has been no comparable study. But there are some indications that the teaching of writing has changed in some ways and remained the same in others.

The goal of Hillocks (2002) was to determine the impact of state writing assessments on the teaching of writing. He did not observe classes as Applebee did. Nor did he include teachers in disciplines other than English. But he did interview teachers at every tested level and at levels not tested as well as administrators and supervisors. Most teachers and supervisors described practice in considerable detail, and it appears that they were quite honest and straightforward about their practices.

In the approximately 20 years or so between the time of the Applebee study (1981, 1984) and the Hillocks study (2002), there has been significant change. While Applebee indicates that only three 3% of the time spent on writing was devoted to work on pieces of a paragraph or more, nearly all teachers interviewed talk only about the writing of multi-paragraph compositions, even at the elementary level. In many districts, the focus was on writing five-paragraph themes (5P), thus imposing a limit of sorts. But it is very clear that, as a result of the state assessments, students in the states studied by Hillocks and probably in all states that collect writing samples as part of their assessments, students are writing far more than they did 20 years ago.

Second, the detail provided by teachers in the Hillocks study indicates that they spend far more time in preparation for writing than the three minutes for the teachers in the Applebee study. Indeed, several teachers describe strings of activities that precede writing. Some teachers describe reading and studying several models of writing, analyzing their characteristics in class, brainstorming for ideas, and organizing those ideas, all before writing. Some interviews make it evident that such activities may take several class sessions.

Third, there is greater attention to audience, or, at least greater seeming attention to audience. Most of the state assessments use topics that allude to some audience (your principal, your senator, the mayor of your town, other students) and teachers use these in their assignments. But what exactly students make of such audiences named in their assignments is less than clear.

Fourth, teachers appear to be preparing teachers more for writing than Applebee found. Across the five states Hillocks examined, an average of 78% of the language arts teachers interviewed used model pieces of writing or more abstract descriptions of the kinds of writing students were to do, nearly two and a half times the number Applebee found. Further, in the national survey Applebee asked about instructional procedures that teachers felt were important. Only slightly over 37% of English teachers mentioned brainstorming. Hillocks found that over five states an average of slightly over 71% mentioned it. Applebee indicates that 26.4% of English teachers reported using class time for students to read each other's writing. Hillocks found that 65.8% report using peer response regularly.

On the other hand, while Applebee found that 59.3% of English teachers thought that writing more than one draft was important in teaching, Hillocks found that an average of 60.4% talked about revising as an important instructional technique, suggesting no real change at all. However, the percentages differ widely by state. In Kentucky, which has a portfolio assessment, the percentage of teachers emphasizing revision was 81 and in Oregon, with a three day, three class period time for the assessment, 84% emphasized revision. The three other states with assessments calling for students to sit-down-and-write in a single time period, the average percentage of teachers emphasizing revision was only 45.7, an indication of the impact of state assessments.

Despite these apparent advances, there is an underlying similarity in the way writing is taught during the two periods. In both periods, teachers and curriculum makers assume that the knowledge necessary for effective writing is general knowledge of a few principles that are applicable to all or most writing.

In this regard, it is interesting to note three case studies conducted more recently. The first was conducted by Anagnostopoulos (2003) in 1996-1999, about 20 years after Applebee conducted his study of writing in the secondary school. She observed classes in a Chicago school that had been placed on probation as a result of low test scores. In order to hold teachers and principals accountable for student outcomes, Chicago had instituted its own Chicago Academic Standards Examination, known as the CASE exam. The test was accompanied by criteria for scoring, a rubric which, according to Anagnostopoulos, "constructs readers as not only uncritical, but minimally skilled as well." She comments on the criteria: "According to these criteria, 'good readers' can cite details from a text. They do not, however, use these details to identify key ideas or 'significant concepts.' In fact, the details that students include in their essays could be 'irrelevant' to the simplistic interpretations that students develop" (p. 191).

In order to prepare students for the test, the teachers she observed consistently conduct recitations about a few chapters of the novel (cut short because of time pressures) and show the film version to insure that students would know the major events of the story. Anagnostopolous analyzed the questions asked and found that

nearly all were literal with some leading to simple inferences, but no complex inferences about characters or events, the thematic content, or the structure of the book. Teachers would predict possible exam questions, pose a thesis statement, and use recitation to draw students into supplying the evidence to support it. She describes how one of the teachers suggests that the question on the exam might ask about the title of the novel or who the mockingbirds in the novel might be. In one classroom segment examined in detail, the class considers Boo Radley as a possible mockingbird, a person who does nothing but good for others. The teacher asks for examples of the good things that Boo Radley does. Students provide several examples. Anagnostopoulos calls their responses "text reproduction." She comments that "The students fulfilled their role in text reproduction by recalling details that fit into Mr. Jones' framing of the novel. After they offered a series of plot details, Mr, Jones congratulated the students on 'gettin it all,' and stated that if the question appeared on the CASE, the students would do well because they '[understood] it'" (p. 198). Sometimes, after such recitations, he explained how to put it all into a 5P. Anagnostopolous interprets this teaching behavior as an effect of the high pressure testing. However, it is highly reminiscent of what Applebee observed 20 years earlier, and he reports no pressure from placing schools on probation.

A second case study indicates that standardized tests alone are not responsible for such teaching. Kahn (2000) studied the materials for evaluation of students used by a cohort of grade 10 English teachers in a suburban high school that had been recognized as a Center of Excellence by the National Council of Teachers of English. The school had a population of mostly White (78%) and Asian (13%) students 71% of whom planned to attend either a two or four year college. Kahn states that "the school's average scores for reading and writing in the state testing program tended to be significantly higher than the state and national averages.

Kahn studied the teachers' quizzes, unit tests, final exams and composition assignments. Teachers use a point system with each test item worth a particular number of points and each test is worth the total of the points per item, e.g., the final exam is worth 200 points. She found that over 65% of the points available for the semester were based on multiple choice, matching, or true/false items. Most of these involved highly literal information that had been presented in the textbook or by the teachers in class. Likewise, questions that might have called for interpretation actually expected students to recall the interpretations presented in class. For example, one question for the novel, A Separate Peace, "students were asked to label 12 objects, situations, places, and characters as (a) a symbol of peace or (b) a symbol of war..... Students did not have to judge whether something was a symbol, explain why something is a symbol, or analyze the effect or meaning of the symbol" (p. 284). Obviously, these answers had been determined and represented substantial knowledge that students were supposed to have acquired. In short, these students appear to be doing less processing of content for writing than even those in Applebee study of 20 years previously.

Further, there was one unit on writing during the semester, lasting two weeks and focusing on the 5P. At its conclusion the evaluation involved writing the 5P to very exacting specifications, down to the positioning of certain sentences in the composition. The final exam questions that deal with writing, are multiple-choice asking

students to recall the rules of the 5P. For example, one question reads as follows: "In the last paragraph, the thesis statement is always the _____ sentence." The choices are "a. first, b. second, c. last, d. doesn't matter": (p. 282). The correct answer, by the way, is a. It appears that the teaching of writing in at least some schools has deteriorated to well below the levels of instruction that Applebee found.

Finally, a detailed case study of a teacher's learning to teach the 5P by Johnson, Smagorinsky, Thompson, and Fry (2003) helps to explain the deterioration. The study details how a neophyte teacher learns to teach the 5P. Johnson et al had access to interviews with and observations of the neophyte (Leigh) from student teaching through the first full year of teaching and including communication after a few years of teaching. The authors contend that four of the six possible reasons for the persistence of the 5P are implicated in Leigh's learning to teach it and accept it. They see her as an outstanding teaching candidate and present evidence of her doing a better job of teaching the 5P than many other teachers appear to, helping students to generate content, to revise, etc. They see her working conditions as good. But they contend that the other four reasons come to bear. Her own deeper belief system, developed during her apprenticeship of observation, predisposed her to favor the 5P. Her teacher education program did not help her develop a "strong conceptual framework for critiquing the five-paragraph theme or developing a rationale for teaching writing in other ways" (p. 167). Instead, she was left to "develop her conception of writing through apprenticeships with mentors and colleagues she found in the field" (p. 168), and they focused on highly specified and rigid forms of writing, not on learning strategies for examining the content which might dictate form. In addition, the pressures experienced from the school, community and state to do well on the exam leads to a belief that the 5P is necessary and, in some eyes, the epitome of strong writing.

5. RESEARCH ON TEACHING WRITING

5.1 Writing Assignments

Applebee (1981, 1984) indicates that when he conducted his survey, giving the writing assignment virtually stood alone. Since then, there has been considerable research and writing about assignments and the best ways to frame them for student learning and assessment. Ruth and Murphy (1987) present a very thorough analysis of assignments for assessment and the research about them. The specific topics or writing prompts exemplified in the book are comparable to topics in the various volumes published by NCTE for use by teachers. The book discusses the problems with topics for writing especially as they stand alone for assessment. Before use, they need to be pretested for such problems as ambiguity, level of information required to write about them, and the impact of the level of rhetorical specification. It also reviews research related to such issues. But there has been no research indicating that the selection of topics for writing assignments has any impact on learning to write.

5.2 Responding to Student Writing

Applebee (1981) called response to writing "the major vehicle for writing instruction, in all subject matter areas." Freedman's (1987) study of response to writing is one of the most important in terms of both breadth and depth. Freedman set out to examine the nature of response teachers make and the impact that response has on student writing. She does this by conducting a survey of 560 teachers nominated as successful teachers of writing by National Writing Project site directors, a survey of selected students of those teachers, and ethnographies of two skilled teachers of ninth grade academic writing. She redefines response to writing to include teacher and peer written and oral response not only to final products but to drafts in progress and most importantly to the thinking that students do as they participate in discussion and generation of ideas in preparation for writing. The latter is not usually thought of as response to writing, but it may be the most important kind. The surveys and ethnographies illustrate and support for three conditions for successful response to student writing: "successful teachers... resist taking over the writing of their students"; they "communicate high expectations for all students"; and they provide plentiful help and support for students during the writing process (p. 160).

Sperling and Freedman (1987), in a case study of a student culled from Freedman (1987), examine the question of why even the most promising students misunderstand and or misconstrue the written comments that teachers write on their papers even when they are accompanied by conferences, peer group response , and whole class discussion of responses. They study the responses of a high achieving ninth grade girl to her teacher's comments on segments of text (what they call "response rounds") of a character study developed over several drafts. They categorize the teacher comments as either reflecting information made explicit by the teacher in class or not. In all of the student's drafts, many of the teacher's comments had no in-class referents. In every such case, the student's attempts at revision failed in some way. Of the teacher's comments with in-class referents, about half are positive reinforcements and half indicate a need for revision. Sperling and Freedman report that the student has no problems processing the former. When the teacher compliments word choice or the use of detail, the student notes it and tries to produce more in the next piece of writing. The student's attempts to revise for comments indicating a need for change, however, even when they have in-class referents are often complicated by differences between the teacher's and student's values and knowledge, suggesting that when responding to written drafts is the major vehicle for composition instruction, it is probably not very effective.

Sperling (1990) provides an analysis of one teacher's conferences with students about their writing. She says that "participating in the explicit dialogue of teacher-student conversation, students collaborate in the often implicit act of acquiring and developing written language" (p. 282). She indicates that the conferences she examined had a range of purposes: "to plan future text, ...to clarify the teacher's written comments..., to give feedback on texts on which there were no written comments..." and "to cover concerns tangential ... to those above"(p. 289). Sperling presents an analysis of the number of units of discourse initiated by the focal students or teacher and completed by the students or teacher to show that the conferences are collabora-

tions and represent a “context for dialogic learning to blossom”(p. 318). While it is easy to imagine that the concerns might be matters of form, every conference quoted in the study has a substantive focus. That is, they develop the potential content of the writing.

5.3 Instruction as Preparation for Writing

Hillocks (1986a) reviewed nearly 500 quasi-experimental studies of writing instruction conducted between 1963 and 1983 and selected those that met criteria for strong research design principles (Campbell & Stanley, 1966) in order to conduct a.meta-analysis or research synthesis, which permits the comparison of results across studies and to deal with the variability of results (Cooper & Hedges, 1994).

The application of several criteria for inclusion in the synthesis (See p. 109) eliminated over 80% of the studies and resulted in the inclusion of 60 studies with a total of 73 experimental treatments. Researchers coded the salient features of all experimental and control treatments along several dimensions: duration of study, level of study, mode of instruction, and focus of instruction. The duration and level of the study (elementary, high school, college) had no significant impact on the results. Both mode and focus of instruction revealed significant differences. Mode of instruction "refers to the role assumed by the classroom teacher, the kind and order of activities present, and the specificity and clarity of objectives and learning tasks." In contrast, focus refers to the dominant content of instruction, for example, sentence combining, grammar, or the study of model pieces of writing (Hillocks, 1986, p. 113).

For the meta-analysis, Hillocks computed effect sizes for each experimental treatment. These are the difference between the control and the experimental group gains, from pretest to post test, divided by the pooled standard deviation for all groups’ post test scores. The results indicate by what proportion of standard deviation the experimental groups outperformed or underperformed the control groups in the same study.

5.3.1 Mode of Instruction

Coding on this dimension indicated four clear modes of instruction, each including five to ten studies. In the first of these, called presentational, the teacher dominates the classroom, presenting information in lecture and recitation and from textbooks, setting assignments, explaining objectives to students, and so forth. This is the mode of teaching that Nystrand with Gamoran, Kachur, and Prendergast (1997) found to dominate the 451 lessons (class periods) in 112 eighth and ninth grade language arts classes they observed.

The second mode of instruction, natural process, is quite different. Rather than presenting rules, criteria, and models to guide writing, teachers in this mode encourage students to write on topics of their own choice, receive feedback from peers, and revise writing as they wish. Most of these make use of small, student led discussion groups but avoid structured problem solving. Several of the studies examining this

mode of instruction refer to Piaget and the idea that development precedes learning. Therefore, the idea is that learning will occur as the child matures and pursues her or his own goals.

The third mode of instruction is individualized writing conferences between teacher and student. Generally, the studies do not explore the nature of the conferences. The major differentiating factor is that the teacher meets with one student at a time outside the classroom, at least during some parts of a course.

Hillocks calls the fourth mode of instruction environmental because it places student, materials, activities, teacher, and learning task in balance. To be included in this category, a treatment had to stipulate the use of student-led small group discussions focused on solving problems that involve specifically stated dimensions such as judging pieces of writing according to specific criteria and revising some or all of them according to suggestions generated through use of the criteria; discussing materials in order to make an analysis or classification, or interpretation. Hillocks (1999) presents a long profile of an eleventh grade teacher using such teaching methods along with transcripts of group and class discussion that meet the criteria set by Nystrand et al (1997).

Students in the environmental groups (mean effect = .44) outperform those in the natural process groups (mean effect = .19) and those in individualized treatments (mean effect = .17). The progress by students in presentational groups is nearly nonexistent (mean effect = .02). Further, each of these mean experimental/control effect sizes is homogeneous. None exhibits statistically significant variation.

5.3.2 Focus of Instruction

Focus of instruction refers to the dominant content of instruction occurring prior to assignment. Six foci were the subject of five or more studies each: grammar, study of model pieces of writing, sentence combining, the use of scales for judging and revising writing, inquiry, and free writing.

Studies in the grammar category concentrate on teaching grammatical concepts from traditional school grammar (TSG), except in one case that made use of generative grammar. Grammatical concepts include such TSG concepts as parts of speech and parts and kinds of sentences, as well as prescriptions about items of usage. The stated goal of understanding how language works is frequently assumed to be instrumental in learning to write.

The category labeled "models" includes studies of the effect of asking students to read and learn the characteristics of finished pieces of writing or abstract representations (e. g., five blocks representing the 5P) so that students will be able to use the models as guides to their own writing. Applebee (1984) and Hillocks (2002) indicate that this focus is common in schools.

Sentence combining treatments ask students to combine sets of usually prewritten sentences in certain ways. (See Strong, 1986, for a complete description of a variety of approaches to sentence combining.) While sentence combining is related to grammar, it does not focus on naming the parts of speech or the identification of

kinds of sentences and clauses. Rather, it focuses on the procedures of putting phrases, clauses, and sentences together in a variety of ways.

Studies in the scales category made use of criteria to help students judge and revise pieces of writing by others. In one such study, Sager (1972) taught fairly simple writing scales to sixth graders which they used to rate pieces of writing by others not known to them. Following the ratings, they revised the low rated pieces to meet higher level criteria.

Inquiry appears in several treatments and was operationally defined for this meta-analysis as focusing on sets of data and "activities designed to help students develop skills or strategies for dealing with the data in order to say or write something about it" (Hillocks, 1986, p. 211). Inquiry is discussed at length or exemplified in several chapters of Hillocks (1995).

Free writing is a technique that asks students to write whatever they have on their minds in journals, which may remain inviolate, or as preparation for sharing ideas and experiences with others. This approach views free writing as a means of helping students discover what they have to say and their own voices for saying it.

The results for focus of instruction indicate that the mean effect size for grammar is -.29, well below the impact of any other focus of teaching. The negative finding results from the fact that students focused on TSG or generative grammar made only tiny gains that were not statistically significant while their counterparts in control groups made fairly substantial gains. The study of models results in gains (mean effect = .22) smaller than the average gain for all experimental treatments. Free writing makes even less headway (mean effect = .16). The most powerful treatments are sentence combining (mean effect = .35), scales (mean effect = .36), and inquiry (mean effect = .56). All mean effect sizes are homogeneous.

The treatments with the largest gains for sentence combining, scales, and inquiry all focus on systematically teaching procedural knowledge, knowledge of how to do things. While free writing engages students in procedures that they already know, it does not systematically help students learn new, specifiable procedures. Both grammar and models focus on learning what Hillocks (1995 and 1999) calls declarative knowledge, knowledge that teachers hope may result in procedures, but they do not aim at the procedures.

For this review, few experimental or quasi-experimental studies turned up in searches of literature on secondary school writing. Some that do appear to confirm the findings of the meta-analysis. Yeh (1998), for example, studied the teaching of argument to minority middle school students in two experimental groups and two comparison groups with a total of 110 students. It is a carefully designed study in which the differences between the experimental and comparison groups are precisely laid out. All groups worked on writing argument and shared a book that involved issues (e.g., “throw[ing] “toxic wastes into the ocean”) and related information for debate. All students read the materials and were engaged in debate teams to make presentations to their classmates, after which they each wrote an essay on the issue, writing eight essays in the course of ten weeks. All groups also pursued the writing process from prewriting to revision and final draft. The difference between the groups was in the kind of prewriting that students were taught. In the experimental groups students were taught explicitly how to use a “heuristic” for developing an

argument based on Toulmin's model of argument. The comparison groups were encouraged to develop their ideas through the use of a web, "(Concept map), with the opinion (main claim) in the middle and branches for an introduction, three supporting reasons and a conclusion" (p. 62), which sounds suspiciously like a 5P. The experimental outperformed the comparison groups from pre to post tests by a margin comparable to those in the inquiry group of the meta-analysis. If we disaggregate the data in Table 2 (p. 66), it is possible to calculate the experimental control effect size. The mean gain for the experimental group is .65, for the comparison group .11, with a pooled sd of .73. Thus, the experimental control effect size is .74, comparable to results for studies in the environmental and inquiry groups where it would be coded.

5.4 The Question of Specificity of Knowledge as Students Learn to Write

In 1992 Smagorinsky and Smith examined the question of what kinds of knowledge were being taught and studied in the name of knowledge of composition and literary understanding, particularly in regard to specificity. They relate the question to a long standing controversy in educational psychology concerning the kinds of knowledge likely to transfer. This controversy is particularly concerned with the questions, "To what extent is knowledge specific to particular situations? To what extent can learners transfer knowledge from one context to another? Can people learn general skills that help them solve problems in a variety of fields?" (p. 280).

Smagorinsky and Smith observe that three positions have emerged among theorists and researchers in composition and literary understanding in relation to transfer of knowledge. The first is that of general knowledge, the advocates of which believe that a few general strategies suffice for any sort of writing. Smagorinsky and Smith cite, for example, Warriner and Griffith's (1977) English Grammar and Composition which states that "No matter what you are writing about the basic steps involved in writing are almost always the same." According to Smagorinsky and Smith the steps outlined in the book "include selecting and limiting a topic, assembling materials, organizing and outlining ideas, writing a draft that follows a particular form (usually including five paragraphs), revising, and preparing a final draft" (p. 282).

Smagorinsky and Smith point out that "Faith in the sufficiency of general knowledge of text structure is rare among the professoriate and has been replaced by a belief in general procedural knowledge that has begun to transform teaching and textbooks" (p. 283). They point to the development of heuristics as one departure from the more traditional Warriner approach. Even that has been "supplanted in popularity by general procedures for producing texts that rely on nonlinear thinking such as brainstorming, clustering, and free writing" (p. 283). According to Smagorinsky and Smith, the idea of "non-linear thinking" has been most earnestly advanced as general knowledge of the writing process by Murray (1980, 1987) and Elbow (1973, 1981). They cite Murray (1987) as referring to the process approach to writing "as consisting of five steps: collecting, focusing, ordering, developing, and clarifying. Writers can apply this general process to any composing problem and couple the five steps with general strategies such as free writing, brainstorming, and mapping" (p 283).

Smagorinsky and Smith cite Elbow as perhaps "the most passionate advocate of general procedures" (p. 284). They indicate that he advocates two different writing processes. The first is for use when you do not want to do much new thinking, as in memos, reports, somewhat difficult letters, and even essays. He calls this the direct writing process and advocates simply dividing the available time in half, using the first half for writing without worrying about organization, language, correctness, or precision. The second half is for revision. The second process for more difficult tasks is described in similar terms but with the metaphor of a voyage out and a voyage in. Smagorinsky and Smith explain that "He gives several quite disparate examples of possible applications of the loop process: a comparison/contrast of Freud and Jung, an analysis of the causes of the French Revolution, a report on levels of pollution of various chemicals in Puget Sound, an analysis of government expenditures for various kinds of armaments and defense, and a paper on abortion" (p. 284). They comment that “The assumption behind this conception of composing knowledge is that writing consists of a very few simple procedures that one develops and uses effectively through practice”(p. 284).

A second position is that knowledge for writing is task specific and argues that different writing tasks not only require knowledge of different forms of writing but of task specific procedures or strategies for dealing with both content and form necessary to produce the desired product. Smagorinsky and Smith cite research on teaching students to write extended definitions of abstract concepts.. They outline the following enabling strategies for developing the content: "1) to circumscribe the concept generally, 2) to compare examples in order to generate criteria which discriminate between the target concept and related but essentially different concepts, and 3) to generate examples which clarify the distinctions" (p.286). Smagorinsky and Smith present a list of works presenting a comparable approach to teaching writing.

The third position they identify is one that argues that advanced or professional writing is not only task specific but contextualized in highly specific communities and that the communities have conventions and standards that are specific to the communities but that may not be shared by others. That is, those who wish. to become literary critics in the academy must learn the standards and priorities of that community or they will not be able to achieve recognition in that community of scholars. Those who wish to become effective lawyers and judges must have the knowledge and ability to find the precedents that will provide for the backing of the warrants by which they show that their data actually provide evidence in support of their claims. See, for example, Stratman’s (1990) discussion of legal thinking and writing.

Smagorinsky and Smith suggest that these distinctions indicate a "curricular path" (p.299). They summarize: "The general knowledge position is most widely substantiated at the elementary level, the task-specific position is best supported at the secondary and collegiate levels, and community-specific position is most typically investigated at the upper levels of schooling and in the professions" (p. 298-299). The curricular path is apparent. Beginning writers need to know and learn to use the writing process. But as students become older, at some point, they need additional, task specific knowledge to meet the demands successfully for writing they

encounter. Even if students learn task specific critical thinking skills in high school and college, they have to learn how to adapt them to the specific professional communities they enter.

5.5 Studies of Teachers Who Beat the Odds

Some of the most valuable research about teaching writing in secondary schools is that of Langer (2001), examining three groups of teachers in urban schools with diverse populations, some of which consistently beat the odds by helping students to higher achievement in English than socio-economic data would predict. She studied high-performing teachers in high-performing schools, high-performing teachers in typically-performing schools, and typically-performing teachers in typically-performing schools. She finds six major distinctions between the high-performing and typically-performing teachers.

First, she finds that high-performing teachers use a combination of approaches to teaching skills. They separate them out for specific attention, e.g., explaining with examples and exercises how to use certain punctuation marks. They use what Langer calls "simulated instruction" in which students are asked to produce short units of text for the purpose of applying the skills. Finally, they integrate the skills taught with on-going larger curricular goals, tasks in which students are asked to apply the skills in question. Typically-performing teachers, however, very often only use the method of separating out the skill to be taught without integrating it into the on-going curriculum.

Second, in like manner, she found that high-performing teachers integrate preparation for district or state-wide tests into the on-going curriculum, while typically-performing teachers separate it out.

Third, Langer finds that at least 88% of high-performing teachers "overtly pointed out connections... among concepts and experiences within lessons; connections across lessons, classes, and even grades; and connections between in-school and out-of-school knowledge and experiences classes and grades, and across in-school and out-of-school applications." In contrast, "the more typical teachers tended to make no connections at all" (p. 864). They treat knowledge and skills as discrete entities.

Fourth, Langer finds that "All of the more successful teachers overtly taught their students strategies for organizing their thoughts and completing their tasks, whereas only 17% of the more typical teachers did so." For example, "in the higher performing schools, teachers often segmented new or more difficult tasks, providing their students with guides for ways to accomplish them" (p. 868).... "Strategies for how to do the task as well as how to think about the task were discussed and modeled .... Most teachers in the higher performing schools share and discuss with students rubrics for evaluating performance; they also incorporate them into their ongoing instructional activities as a way to help their students develop an understanding of the components that contribute to a higher score" (p. 868). In short, in the higher performing schools there was a clear emphasis on teaching the procedures of what to do to do well, a development of meta-cognitive knowledge. In contrast, "in more

typical schools, instruction focused on the content or the skill, but not necessarily on providing students with procedural or meta-cognitive strategies" (p. 869).

Fifth, Langer finds that all the high-performing teachers adopted

> a generative approach to student learning, going beyond students' acquisition of the skills or knowledge to engage them in deeper understandings. In comparison, all of the more typical teachers tended to move on to other goals and activities once they had evidence that the target skills or knowledge had been learned (p. 870).

Sixth, and finally, Langer finds that high-performing teachers create social contexts for learning. In schools that beat the odds, she says,

> English learning and high literacy (the content as well as the skills) were treated as social activity, with depth and complexity of understanding and proficiency with conventions growing out of the shared cognition that emerges from interaction with present and imagined others.

In contrast, the more typical classrooms, "emphasized individual activity and individual thinking, with students tending to work alone or to interact primarily with the teacher."

Langer continues,

> In the higher performing schools, at least 96% of the teachers helped students engage in the thoughtful dialogue we call shared cognition. Teachers expected their students to not merely work together, but to sharpen their understandings with, against, and from each other. In comparison, teachers in the more typical classes focused on individual thinking. Even when their students worked together, the thinking was parallel as opposed to dialogic (p. 872).

Sperling and Woodlief (1997) also study expert writing teachers who taught in the ways the National Writing project recommends and display characteristics that Langer finds in her successful teachers. They give their students many opportunities to write about literature, current events, and first hand experience, while at the same time devoting plenty of time to preparatory activities, feedback, including whole class interaction and teacher conferences.

6. EDUCATION OF TEACHERS OF WRITING

Researchers and many teachers know quite a bit about what constitutes effective teaching of writing. We have evidence that teachers are paying more attention to writing than they did 25 years ago. Yet the evidence indicates that teachers are either unaware of the research evidence, or they do not care about it. While Langer (2001), Sperling & Woodlief (1997), Sperling (1995), and Hillocks (1999) show exemplary teachers, most appear to know little about the teaching of writing other than what comes to them through their state assessments..

Smagorinsky and Whiting (1995) surveyed English education programs in 81 universities, which returned syllabi for methods classes.. The methods courses designated as surveys appear to treat composition in a few sessions because they are structured to address a host of topics, from censorship to computers. Most programs appear to present only the most general knowledge about writing. Even courses devoted to writing tended to be workshops for students to work on their own writing

rather courses in the teaching of writing. Kennedy (1998) examines a variety of teacher education programs, including some secondary pre-service and in-service programs. She classifies them as traditional or reform with traditional focusing on the mechanics of presenting lessons but largely ignoring the subject matter of writing and the reform programs as focusing on the subject matter. Kennedy's goal is to determine if teacher education makes a difference. She presents two descriptions of classrooms, one traditional and one reform. The traditional focuses exclusively on lesson design which is formulaic and the reform professor lectures about the ineffectiveness of teaching grammar and usage but says nothing about how to teach. Kennedy finds that the programs do make a difference in the students but her evidence has to do with little more than teachers' attitudes toward mechanics. One can conclude, however, that in the programs she studied, students learn only the most general knowledge about writing and its teaching. One might conclude that colleges and universities simply do not prepare teachers for the teaching of writing, and therein lies the problem with writing in the schools.

REFERENCES

Anagnostopoulos, D. (2003). Testing and student engagement in urban classrooms: A multi-layered approach. *Research in the Teaching of English, 38*(2), 177-212.

Applebee, A. N. (1981). *Writing in the secondary school: English in the content areas.* NCTE Research Report, No. 21. Urbana, IL: National Council of Teachers of English.

Applebee, A. N. (1984). *Contexts for learning to write.* Norwood, NJ: Ablex.

Atwell, N. (1987). *In the middle: Writing, reading, and learning with adolescents.* Portsmouth, NH: Heinemann.

Bahktin, M. (1981). *The dialogic imagination: Four essays* (Michael Holquist, Ed.) (Caryl Emerson, Trans.). Austin: University of Texas Press.

Britton, J.N., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities* (11-18). London: Macmillan Education.

Campbell, D. T. & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cooper, H. & Hedges, L. V. (Eds). (1994). *The handbook of research synthesis.* Washington, D.C. : Brookings Institution.

Diedrich, P. B. (1974). *Measuring Growth in English.* Urbana, IL: National Council of Teachers of English.

Elbow, P. (1973). *Writing without teachers*. New York: Oxford University Press.

Elbow, P. (1981). *Writing with power: Techniques for mastering the writing process*. New York: Oxford University Press.

Freedman, S.W. (1987). *Response to student writing*. NCTE Research Report No. 23. Urbana, IL: National Council of Teachers of English.

Hedges, L.V. & Nowell, A. (1995) *Sex differences in mental test scores, variability, and numbers of high-scoring individuals.* Science, 269, (pp.41-45)

Hedges, L.V. & Nowell, A. (1998). Black-white test score convergence since 1965. In Jencks, C. and Phillips, M. (1998). *The black-white test score gap*. Washington, D. C.: Brookings Institution Press.

Hillocks, G., Jr. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: National Conference on Research in English/ERIC Clearinghouse on Reading and Communications skills.

Hillocks, G., Jr. (1986). The writer's knowledge: Theory, research and implications for practice. In A. Petrosky & D. Bartholomae (Eds), *The teaching of writing: 85th Yearbook of the National Society for the Study of Education, Part II* (pp. 71-94). Chicago: National Society for the Study of Education.

Hillocks, G., Jr. (1995). *Teaching writing as reflective practice*. New York: Teachers College Press.

Hillocks, G., Jr. (1999). *Ways of thinking, ways of teaching*. New York: Teachers College Press.
Hillocks, G., Jr. (2002), *The testing trap: How state assessments of writing control learning.* New York: Teachers College Press.
Huot, B. (1990). The literature of direct writing assessment: major concerns and prevailing trends. *Review of Educational Research ,60*, 237-263.
Jencks, C. and Phillips, M. (1998). The black-white test score gap: An introduction. In Jencks, C. & Phillips, M. (1998). *The black-white test score gap* (pp. 1-51). Washington, D. C.: Brookings Institution Press.
Johnson, T.S., Smagorinsky, P., Thompson, L., & Fry, P. (2003). Learning to teach the five-paragraph theme. *Research in the Teaching of English, 38* (2), 136-176.
Kahn, E.A. (2000). A case study of assessment in a grade 10 English course. *Journal of Educational Research, 93*(5), 276-286.
Kennedy, M. M. (1998). *Learning to teach writing: Does teacher education make a difference?* New York: Teachers College Press.
Kinneavy, J. L. (1971). *A theory of discourse*. Englewood Cliffs, NJ: Prentice-Hall.
Langer, J.A. (1986). *Children reading and writing*. Norwood, NJ: Ablex.
Langer. J.A. (2001) Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal*, 38, 837-880.
Lee, C.D. & Samgorinsky, P. (Eds). (2000). *Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry. Learning in doing: Social, cognitive, and computational perspectives.* New York: Cambridge University Press.
Murray, D. (1980). Writing as process. In T.R. Donovan & V. W. McClelland (Eds), *Eight approaches to teaching composition* (pp. 3-20). Urbana, IL: National Council of Teachers of English.
Murray, D. (1987). *Write to learn.* New York: Holt, Rinehart, and Winston.
National Assessment of Educational Progress. (2002). *The nation's report card: Writing 2002*. Washington DC: National Center for Education Statistics.
Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.
Plato (1973). *The Phaedrus and letters VII and VIII* (W. Hamilton, Trans.). London: Penguin.
Purves, A.C. (Ed.). (1992). *The IEA study of written composition: Education and performance in fourteen countries.* Oxford: Pergamon Press.
Ruth, L. & Murphy, S. (1987). *Designing writing tasks for the assessment of writing.* Norwood, NJ: Ablex.
Sager, C. (1972). *Improving the quality of written composition through pupil use of rating scale* [thesis] (Boston University, 1971). Ann Arbor: University Microfilms.
Smagorinsky, P. & Smith, M.W. (1992). The nature of knowledge in composition and literary understanding: The question of specificity. *Review of Educational Research, 62*, 279-305.
Smagorinsky, P. & Whiting, M.E. (1995). *How English teachers get taught: Methods of teaching the methods class*. Urbana, IL: National Council of Teachers of English.
Sperling, M. (1995). Uncovering the role of role in writing and learning to write: One day in an inner-city classroom. *Written Communication, 12*, 93-133.
Sperling, M. & Woodlief, L. (1997). Two classrooms, two writing communities: Urban and suburban tenth-graders learning to write. *Research in the Teaching of English, 31*, 205-239.
Sperling, M. (1991). Dialogues of deliberation: Conversation in the teacher-student writing conference. *Written Communication, 8*, 131-62.
Sperling, M. (1990). I want to talk to each of you: Collaboration and the teacher-student writing conference. *Research in the Teaching of English, 24*, 279-321.
Sperling, M. & Freedman, S.W. (1987). A good girl writes like a good girl: Written responses to student writing. *Written Communication, 9*, 343-69.
Stratman, J.F. (1989). The emergence of legal composition as a field of inquiry: Evaluating the prospects. *Review of Educational Research, 60*, 153-235.
Strong, W. (1986). *Creative approaches to sentence combining. Theory and Research Into Practice Book.* Urbana, Il: ERIC/National Council of Teachers of English.
Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, et al., Eds). Cambridge: Harvard University Press.
White, E. M. (1985). *Teaching and assessing writing.* San Francisco: Jossey-Bass.

Yeh, S.S. (1998). Empowering education: Teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English, 33*(1), 49-83.