

**Organization and Development Features of Grade 8 and Grade 10 Writers:
A Descriptive Study of Delaware Student Testing Program (DSTP) Essays**

Bonnie R. Albertson, Ph.D.

University of Delaware

Author note: The following information comes from a doctoral dissertation study. An expanded version of this article has been submitted to a juried journal for publication consideration. Therefore, please do not publish or post this document in its entirety without permission of this author; however, you may distribute a copy to colleagues for their personal information and/or post a summary of these results. Correspondence regarding this work should be addressed to Bonnie Albertson, DCTE, 011 Willard Hall, University of Delaware, Newark, DE 19716. Email correspondence should be sent to bonniea@udel.edu

INTRODUCTION

Over the last decade, the rhetoric/composition community and the cognitive science community have moved toward a situated theoretical perspective of the composing process. Constructivism, as an umbrella theoretical perspective, promotes the idea that writing both conveys and creates meaning. This perspective acknowledges how writers balance the rhetorical needs of audience, writing purpose, and message. Writers are seen as active members of a “discourse community;” that is, they view content in context. However, much classroom instruction and many writing textbooks remain grounded in the positivist tradition, which is reflected in the “Current Traditionalist” pedagogical approach to composition – an approach that “fixes” meaning within text and privileges a prescribed organization. Teachers who view writing through this lens often think about writing with a narrow and presumptive set of criteria in mind. This approach also privileges a Euro-centric or Euro-American perception of “good” writing – linear, and thesis-driven. Furthermore, from this perspective, teachers can justify an instructional approach that promotes part-to-whole instruction, even though composition and rhetoric scholarship belies this approach. With such an approach, form dictates content rather than the other way around. In many classrooms, this approach manifests itself in the formulaic “five-paragraph theme or essay” (FPT).

With the advent of high stakes testing, enthusiasm for formulaic writing appears to have intensified as teachers seek quick-fix solutions for students’ perceived inability to organize ideas effectively in high stakes writing tests. Teachers want to provide a prescription that will guarantee students’ success on these high stakes, state-mandated,

writing assessments. Many teachers seem to believe that having students become proficient at writing FPTs will help assure them a passing score on the DSTP. Teachers also claim that students, especially reluctant and novice writers, need the scaffold of a prescribed organizational framework in which to operate. Therefore, they believe that teaching the five-paragraph essay helps students become more confident writers; teachers believe they are making the process of writing more accessible to students by providing a "recipe" of sorts. Some teachers also believe that the FPT does, indeed, represent "good," "clear" writing (Albertson, 2001, 2002).

Clear criteria describing high performances are good, but do formulas teach clear criteria or do they mask them by not allowing students to discover them on their own? Furthermore, whose idea of "clarity" defines these criteria? Are they teaching a blueprint or a straightjacket; a roadmap or a roadblock; a recipe or a cookie cutter? Does the formula enable students to be better writers or does it limit them? This study acknowledges the importance of these questions.

THE PROBLEM: FORMULAIC WRITING

The fray surrounding the FPT debate is not new. Thirty years ago, Janet Emig (1971) raised concerns about the proliferation of five-paragraph essays. Since then, many have joined her in loudly decrying the formula for a host of reasons – ranging from privileging form over content to being blatantly hegemonic. As testament to the FPT as a phenomenon, in September of 2000, *English Journal* devoted an entire issue to the subject, and there have been many follow-up articles since (including encore publications of these articles in the 2003 NCTE *INBOX*). In one of the articles from the 2000 *English*

Journal, Gregory Shafer captures the spirit of the debate by likening students schooled in writing FPTs to the “behavioristic bird who pecks the right button to receive a reward.” He continues by noting that the FPT is yet another attempt to treat students like “subjects that need to be processed” (p. 30-32). Similarly, Bruce Pirie (1997) devotes an entire chapter of his book *Reshaping High School English* to what he calls “‘Mind-Forged Manacles’: The Academic Essay.” Pirie likens the process of composing an FPT to the Cinderella story with hopeful ladies of the kingdom trying to jam their feet into Cinderella’s glass slipper, just as we force students “to make unwilling material fit into three obligatory body paragraphs” (p. 77). Of even more concern to Pirie and others (Halasek, 1999; Hillocks, 1995, 2002; Shafer, 2000; Wesley, 2002) is the fact that finding material that will fit the form keeps students from discovering the “shapes of ideas.” Instead of learning to balance content and form, students learn to avoid such challenges. Adding fuel to the fire, in *A pedagogy of possibility: Bakhtinian perspectives on composition studies* Kay Halasek (1999) states that privileging the form of FPTs “...demonstrates the paternalistic, even repressive, nature of a pedagogy that only allows students one way of writing...” (p. 99).

Despite the huge amounts of anecdotal “evidence” for and against formula writing in the classroom, there is almost no research to support either side of the debate. What little mention there is, has often been an “aside” to larger research questions. Nonetheless, Hillocks reports in *The Testing Trap* (2002) that of the ten states in his study of high-stakes writing assessments, teachers who said they emphasized writing FPTs ranged from a low of 5% (New York) to a high of 72% (Illinois). In a more modest study of 36 middle school students writing to a prompt, an action research team from

Connecticut found that *every* student in their corpus wrote an FPT (Avon Writing Team, 2000). They note that this “dependency” probably helped some students meet Connecticut’s minimum writing standard, but that it also was “distinctly limiting” to others. This team concluded, “writing to formula is neither a cure nor a curse” (p. 62-63).

More recently, psychometrician James A. Penny (2003) participated in a holistic scoring project and made the following observations from a measurement perspective: “...of the adequate papers [observed], one form stood out from the rest, and that was the 5-paragraph essay. I found substantial evidence that many teachers of writing promote that classical form, and I think there is a certain safety in the strategy. Even a poorly done 5-paragraph essay usually contained structure sufficient to warrant a 3 [of 4]. ...I [also] noticed that of all the essays, at least all those I read, written using this classical structure, none received higher than a 3[of 4]” (p. 205). In another recent study of early career teachers, Johnson, Smagorinsky, Thompson, & Fry (2003) write about the pressure on these teachers to teach the FPT, especially in preparation for state testing.

The present study purports to look at the efficacy of teaching the FPT for the purpose of helping students earn higher scores on the writing portion of the DSTP. However, an unpublished pilot study revealed that looking at organization schemes told an incomplete story; development of ideas appeared to be a feature of equal interest (Albertson, 2001 & 2002). Therefore, the purpose of this study is two-fold: to challenge folk wisdom about the efficacy of teaching formulaic writing, at least for the purposes of passing state writing tests, and to look at other development features that characterize writing at various score categories.

STUDY QUESTIONS

The central question framing this study concerns any association between formulaic writing and DSTP writing scores. This study aims to provide some data that may help settle the debate about the efficacy of teaching the FPT to “pass” mandatory writing assessments such as the DSTP, with the additional goal of helping teachers make informed instructional decisions. However, there are underlying questions about the association between DSTP scores and level of development. Therefore, there are two study hypotheses:

- o The organization pattern employed in a piece of writing predicts overall score (with essays employing formulaic organizational patterns earning higher scores than essays organized in other ways).
- o The level of development used in an essay can predict overall score (with highly developed essays earning higher scores and inadequately developed essays earning, lower scores).

Related questions include

- Do those highly formulaic essays earn merely passing scores (combined scores of 6) or do they also earn higher scores?
- Do essays that are organized by some other, non-formula, scheme also earn high and/or passing scores?
- What development strategies characterize writers whose essays earn particular score points?
- Of the essays earning top scores, what percentage follows the FPT formula? What other characteristics describe these “exemplary” papers?

To address these questions, this study looks at data from two different years – 2001 and 2002, two different grade levels – grade 8 and grade 10, and three different discourse categories – expressive, informative, and persuasive.

LITERATURE

The professional literature that informs writing assessment comes from diverse fields including composition and rhetoric studies, cognition and development sciences, and measurement and statistic disciplines. Less theoretical but equally important is the literature on teacher beliefs and pedagogical implications of DWAs. The paper does not purport to cover the literature for these fields, but grounding principles include the following:

Composition and Rhetoric Perspectives

- Writing both creates and conveys meaning;
- There is little agreement among composition scholars, researchers, and teachers about what criteria describe “good” writing;
- Most rhetoric and composition scholars endorse a view of writing that recognizes the importance of having writers balance rhetorical elements. They also recognize that all writing is situated in a particular context. However, most writing assessments (and formulaic writing) is grounded in the “Current Traditionalist” pedagogical approach to composition. Such a positivist tradition attempts to “fix” meaning and discourse categories. Hence we often see a mismatch between theory and pedagogy

Cognitive Perspectives

- Composition studies have shifted to include recognition that writing is a situation-dependent and recursive process;
- Successful writers view concepts in context through their participation in a “discourse community”;
- Much classroom practice and writing assessment strips writing of context;
- Preferred discourse patterns are culturally determined. Linear, thesis-driven discourse privileges a Euro-American perspective.

Measurement Perspectives

- There is no acknowledged or accepted theory of writing assessment *per se*.
- Research paradigm wars have dominated the literature of writing assessment for more than a decade. The most debated issues include
 Qualitative versus quantitative methodologies,
 Traditional conceptualizations about validation and reliability constructs
 versus current ideas about ecological or consequential validity;
- Direct Writing Assessments (DWAs) such as the DSTP are preferable to indirect measures of writing skill.

Perspectives on Pedagogical Implications

- When assessments have high stakes, they influence teacher practice and determine what will be taught;

- Students whose writing experiences center around “practice prompts” are denied the opportunity to learn how to tackle literacy problems; consequently, high stakes versions of DWA can interfere with effective instruction rather than enhance it

METHODOLOGY

Research Paradigm

The setting for this study is not a “natural” writing environment, but paradoxically, it is an authentic setting in which writing occurs; indeed, writing assessments are an increasingly common occasion for writing. Delaware provides the perfect setting for such a study; because of its small size, a statistically acceptable sample size is also a manageable size for analysis.

In an age when political agendas drive education, when the *No Child*, test-crazed era of “scientific” research determines what studies can and cannot be funded or used to drive instruction, it is easy to avoid researching questions of interest simply because the research questions do not “fit” a social agenda or a privileged research paradigm. When the words “robust” and “conclusions” cannot be used simultaneously, researchers may shy away from investigations. However, investigating any association between score and “organizational scheme,” requires the use of nominal data, which are not amenable to inferential statistical analysis. Refusing to eschew the “question of interest” regarding formulaic writing because of the aforementioned measurement limitations, this study relies on descriptive statistics.

Descriptive studies “describe” and interpret relationships observed between and among narrative-laden categories. Such studies do not, however, attempt to posit causality. Rather, the trends are interpreted, with frequent references to the context in which these writings were produced. The best way to capture the associations between

and among the variables without misusing data is with a descriptive study. Even two decades ago, when direct writing assessments were relatively new, descriptive studies were recognized as useful for composition studies (e.g. Cooper, 1983).

THE DATA

Sample Size and Population

This study analyzes both 8th and 10th grade writing samples taken from two different years, 2001 and 2002, yielding 4 different data sets. The prompts for each of those grades and years appear in Appendix B. In order to answer the research questions, this study ultimately required three data sets: one larger corpus from which a second subset was identified, and a third separate data set.

The Corpus

The larger corpus was a random sample of 300 essays for each grade, for each study year (Because some of the samples were illegible, the numbers are less than 300). The smaller corpus was selected from the larger corpus as a modified quota sample (Laurer and Asher, 1988). Nearly 60% of the student writers at this score point get a “6” (of 10 possible points), so I pulled a random sample of 60% of all 6s for the smaller corpus to be analyzed in depth for development strategies. Because there are fewer essays – radically fewer- at the scoring continuum ends or poles, I wanted *all* the 2s, 3s, 8s,9s, and 10s and *most* of the 4s, 5s, and 7s to have enough to validate conclusions. In other words, I took between 60% and 100% of the essays at each score point for the smaller corpus. For the top scoring essays, every legible essay receiving a score point 9 and score point 10 essay was accessed and analyzed.

Table 1. *Study Corpus (essays drawn from DSTP data base)*

	State Population	Larger Corpus (Coded for Development & Organization)	Smaller Corpus (Coded Analytically for Dev. Strategy)	Essays Scoring 9 & 10
Grade 8, 2001	8530	252 (3.0%)	134	28
Grade 8, 2002	9133	266 (2.9%)	152	
Total - Grade 8	17663	518 (2.95%)	286	
Grade 10, 2001	7840	252 (3.2%)	170	21
Grade 10, 2002	8129	239 (2.9%)	136	
Total – Grade 10	15969	491 (3.0%)	306	
Total Both Grades		1009 (3%)	592	49

Study Variables

There were three variables used in this study: the Delaware Student Testing Program (DSTP) scores, organization schemes, and development levels.

DSTP scores. The dependent variables, score points ranging from 2-10 (based on a 5 point rubric with each essay read/scored twice; essays that could not be scored were eliminated from the study)

Dependent Variables. For this study, the independent variables are two writing traits: organization and development, which are indeed dimensions, or features, of

writing. These two variables have been selected for scrutiny, in lieu of others, because Delaware considers organization and development the two most important variables that raters consider when scoring DSTP essays.

Organization Scheme: The organizational plans or schemes selected were the primary variable of interest for this study. These categories are not intended to be hierarchal; hence, the measures are nominal:

Too Short to Code – “essay” has insufficient text to determine organization plan
Unorganized – essay has no discernable or describable organization plan
“Other” Organized – essay with a recognizable and describable plan that is not formulaic *

“Modified” Formula – essays that seemed to adopt the FPT except that the body of these essays might have 2 or 4 (or some other number) of supporting paragraphs. These essays, like the “FPT” below, relied primarily on reason/example for development.

“FPT” – classic, highly formulaic responses, with an introductory paragraph that usually “announces” the 3 supports to be used. This is followed by 3 body paragraphs, each with its own topic sentence and (usually 3) supports utilizing reason/example. Finally there is a concluding (usually a summary) paragraph.

* When noting the organizational scheme utilized, it became apparent that some writers organized their essays by prompt “bullet points.” That is, every DSTP writing prompt is followed by a series of questions intended to scaffold students’ planning for their essays. On occasion, these probes became the default organizational plan for the remainder of the essay.

Note: Interrater reliability checks included 2 raters and 2 measures of reliability (Pearson correlation and Cohen’s Kappa)

Figure 1. Coding for Organization Scheme

Development level. The coding for development identifies salient features of four development levels described in Figure 2.

Well-developed: reasons, main ideas, topics or supports are fully explained and illustrated in-depth through the use of a variety of appropriate, relevant development strategies*

Adequate development: at least some of the principle supports or details (depending on the prompt) are developed with some variety of relevant strategies. However, some of the development *may* leave some questions, be general and therefore unconvincing, *uneven* or repetitive at times (e.g. relying on restatement rather than elaboration. Nonetheless, from a holistic perspective, the development is *sufficient* to fulfill the purpose and audience.

Inadequate Development: attempts at development are inadequate for one of two reasons: either there are an *insufficient number* of details to achieve the purpose for the specified audience or the supporting details are *limited in variety* – mostly reasons/examples/ explanations that are general; none of the key supports (or details) is adequately supported with relevant elaboration.

Undeveloped: Details or supports are minimal and are listed with nearly no relevant elaboration or explanation.

* Strategies defined include examples, explanations, narrative, commentary (opinion), background information, dialogue, suggestions, facts/statistics, definitions, figurative language, and descriptive details. These strategies informed coding choices for the larger corpus and specific use of them was recorded in depth for essays from the smaller corpus. Further notes were also made in the smaller corpus (e.g., relevant/irrelevant, general/specific, cause & effect, logical/illogical, repetitive). In addition, the number of support sentences was recorded.

Note: Interrater reliability readers helped develop

Figure 2. Coding for Development Levels

FINDINGS AND INTERPRETATIONS

Organization Scheme

Overall there was a decided preference for formula papers among both the 8th and 10th grade writers (see Figures 3 and 4). The preference for FPTs was more pronounced in the 8th grade corpus, and the trend was less evident among 10th grade writers.

Organization Schemes (8)

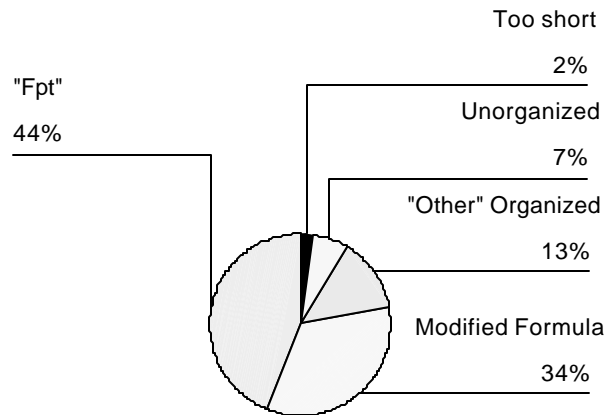


Figure 3. Pie chart shows percentage of organization schemes for aggregate grade 8 corpuses.

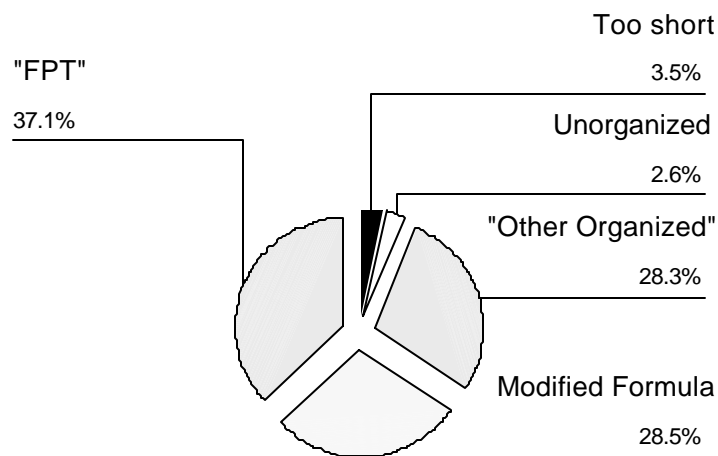


Figure 4. Organization schemes for 2001 and 2002 (Grade 10 aggregated data).

It would be unwise, however, to assume that this trend is due to any preference among 8th grade writers for formulaic papers (or any aversion to the same on the part of

10th grade writers). Indeed, the four prompts were very different, despite the fact that both grade clusters had one of two prompts in the informative discourse category.

Weigle (2002) also notes that some prompts do indeed “elicit traditional five-paragraph essays that were similar in form to one another” while others do not (p. 66).

But, do highly formulaic essays receive better scores than essays organized in other ways? This study suggests that although there may be an association between organization and score, the only conclusions that can be drawn about this relationship is that *more* writers in this study chose to organize their writing using the five paragraph formula than any other organization scheme, and that *most* students earn scores of “6” on the DSTP stand-alone writing prompt. One can also conclude that writers whose essays are unorganized can expect low scores. Figures 5 and 6 show the flat association between score and organization scheme.

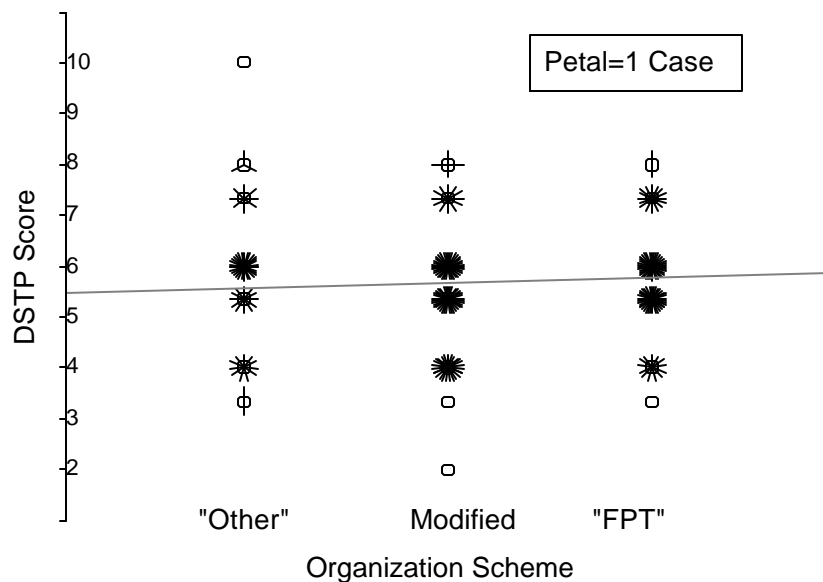


Figure 5. Organization schemes plotted by DSTP scores for aggregate grade 8 data, illustrating the flat association between score and organization.

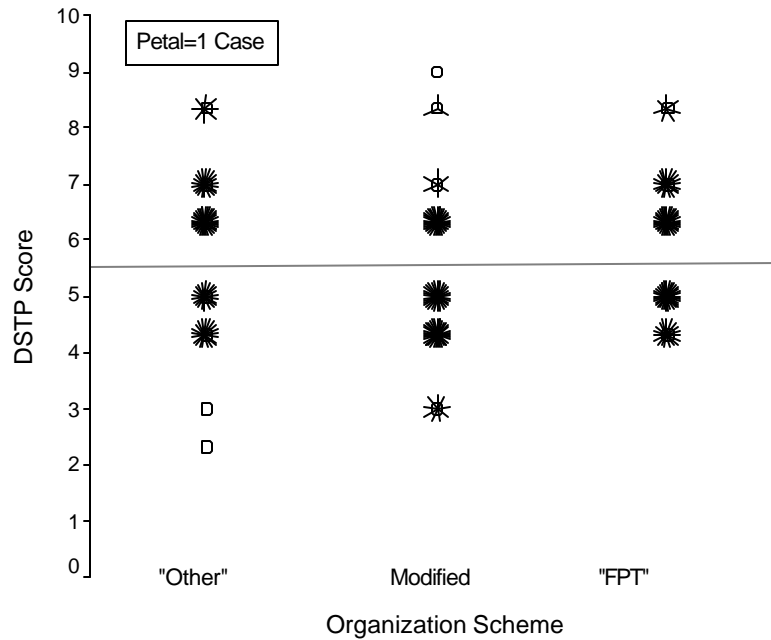


Figure 6. Organization schemes displayed by DSTP scores for aggregate grade 10 data. The flat fit line further emphasized the lack of association between score and organization scheme.

When analyzing the percentages of essay types in each scoring category (“lower,” “midrange,” “higher”) a different – and less simplistic – interpretation emerges. Table 4 shows the distribution of other organized and FPT-organized papers within low, midrange, high, and top-scoring score categories. The table shows that except for grade 10 in 2002, the lowest scoring essays were “other” organized. However, it also shows that for all four data sets, the higher scoring essays were “other” organized. This trend would suggest that experimenting with other forms may be a high-risk and high-reward undertaking.

Table 4

Frequency of Occurrence for Organization Scheme at Score Categories

Prompt/Grade	DSTP Score Category	4-5 lower	6 midrange	7-8 higher	9-10 top-scoring [separate corpus]
“AFTER SCHOOL” (8)	FPT	24%	66%	10%	FPT – 36%
	“Other” Org.	30%	53%	14%	
“GOOD TEACHER” (8)	FPT	22%	71%	6%	“Other” - 57%
	“Other” Org.	50%	22%	22%	
“EXPERIENCE” (10)	FPT	25%	55%	20%	FPT – 38%
	“Other” Org.	30%	43%	26%	
“SIGNIFICANT EVENT” (10)	FPT	36%	53%	11%	“Other” - 62%
	“Other” Org.	33%	50%	18%	

Recall that in order to be coded “other organized,” the organizational scheme of the essay must be describable. A review of these descriptions reveals that the most frequent choice of “other” organization scheme was a mix of narrative with exposition. However, the discovery that many students who selected an “other” organization scheme organized their essay by prompt “bullet point” is another important study finding. A qualitative study interviewing students about their choice to use these probes as a default plan of organization would yield more information about the rationale for this decision.

Development Level

Development and Score Relationships. Is level of development in an essay a good predictor of overall score (with highly developed essay earning higher scores and inadequately developed essays earning lower scores)? As anticipated, there is a strong

positive correlation between well-developed essays and higher scores. Although the differences were sometimes inconsistent in the middle scoring essays, there was a distinct upward trend for development level and score (see Figures 7 and 8).

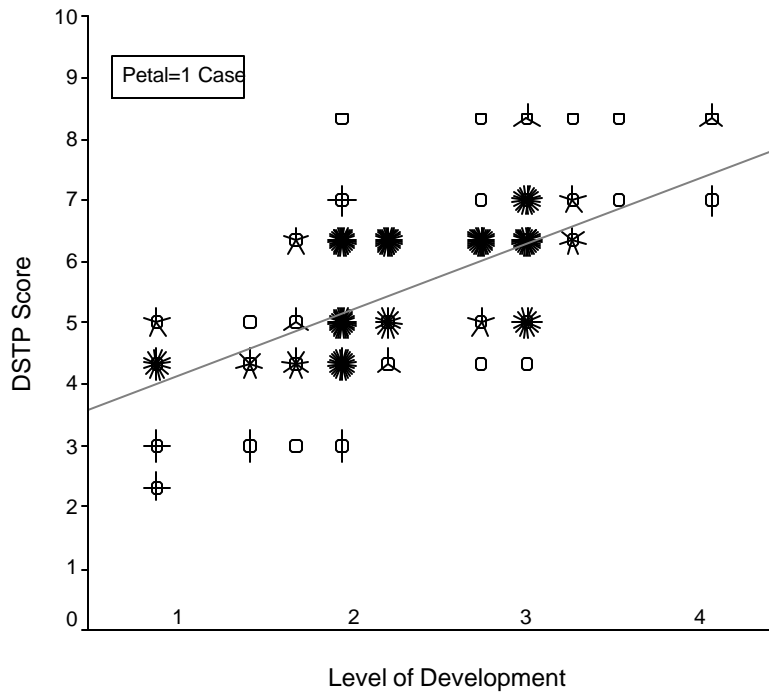


Figure 7. Development levels plotted by DSTP score for aggregate grade 8 corpuses.

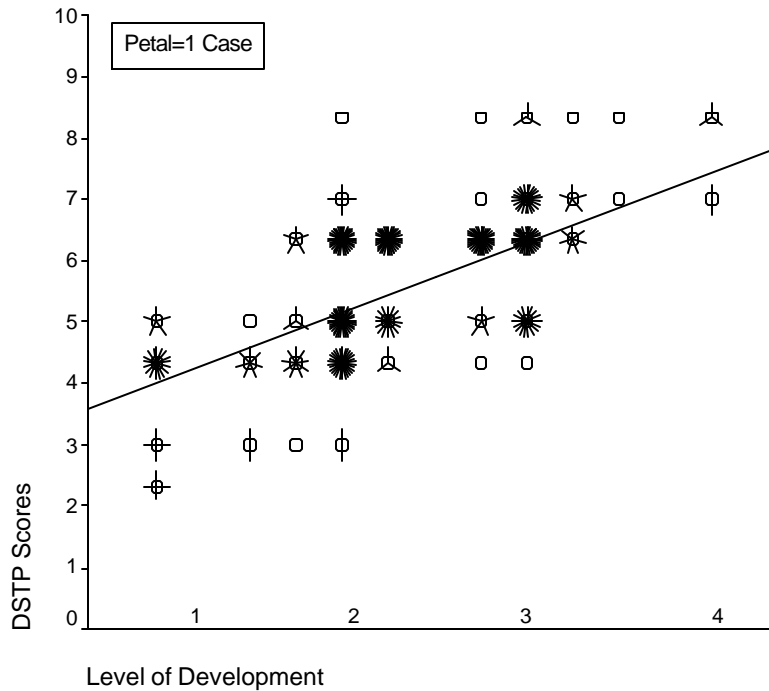


Figure 8. Development levels displayed by DSTP score for 2001 and 2002. The scatter plot shows the large number of inadequate to barely adequate development levels, even for essays earning scores of “5” to “7.”

Overall Trends in Development. There were several consistent patterns across the grade levels. Within all scoring ranges, development by what many (e.g., Decker and Schwegler, 1998) call “reason and example” was the dominant mode of development (see Figure 9) . This preference was even greater among grade 8 writers and among all lower-scoring essays. At the low to mid-scoring levels, irrelevant, illogical, and repetitive explanations and examples were pervasive. Opinion/personal commentary was used at both grade levels, throughout the corpus, but score category often reflected the extent to which the commentary was reflective and appropriate as opposed to “rant-like” and unhelpful in communicating ideas. At the higher scoring ranges, the regular – if not frequent – use of narrative (and sometimes accompanying dialogue), figurative language, and descriptive detail was also observed.

<u>Development Strategy</u> Reasons/Main Ideas & Explanations/Examples (Limited and general)	<u>Development Strategy</u> Reasons/Main Ideas & Explanations/Examples (Often general, adequate number)	<u>Development Strategy</u> Reasons/Main Ideas & Explanations/Examples (Mostly general but some specific elaboration)
2-4	5-7	8-10
<ul style="list-style-type: none"> • Some commentary (e.g., opinion, but often “rant-like”) • Often repetitive, irrelevant &/or illogical 	<ul style="list-style-type: none"> • Occasional use of varied strategies including commentary (opinion), narrative, descriptive detail, suggestion, figurative language, background information, fact & statistic. • Some repetition, irrelevancy 	<ul style="list-style-type: none"> • Regular use of multiple strategies including narrative (& dialogue) commentary, descriptive detail, background information, fact and statistic, figurative language, and/or suggestions and commentary (opinion)

Figure 9. Most frequently noted “development strategies” observed at different score categories for both grade levels

Indeed, both 8th and 10th grader writers from this study utilized the same full range of development strategies. However, as one might expect, the 10th grade writers availed themselves of a greater variety of these strategies more frequently.

One noteworthy finding from the qualitative portion of this study concerns the often indistinguishable role or indefinite function of support statements, especially in the use of examples and explanations, and especially with the lower-scoring essays and among grade 8 writers. These writers seemed to develop ideas by association rather than a reasoned consideration of rhetorical needs. One current criticism (e.g., Hillocks, 2002; Pirie, 1997) of writing instruction is that teachers pay more attention to form than to idea development in writing. Such a belief offers a possible explanation as to why students’

connections among reasons, examples, and explanations seemed to be so muddled, especially with the lower scoring essays. There are other potential explanations as well.

Meade and Ellis (1970) found that 56% of the professional paragraphs they analyzed defied traditional mode of development classifications; most paragraphs represented, at best, a mix of developmental strategies. Of those paragraphs that had an identifiable strategy for development, Meade and Ellis found that 23 % were developed by example, 9 % by reason, and 6% by chronology (cited in Cooper, 1983, p. 291). While this is an old study, it reflects the diversity as well as the preference for example (and explanation), and to some extent the murkiness found in these essays.

A related explanation is that it is difficult to categorize the “roles” (function) of text chunks used to develop ideas in expository writing. Another older but relevant study, conducted by Cooper and Matusushai (1983), categorized the roles of sentences, such as sequencing, rhetorical, generalizing, relationship, development. Cooper and his colleagues were able to agree on roles in narrative writing (97%) but had low interrater reliability (67%) for explanatory writing, which they interpreted as discourse-specific difficulty determining the function of the supports (cited in Cooper, 1983, p. 303). Because the 8th grade prompts invited more exposition and the 10th grade prompts elicited more narrative responses, Cooper’s and Matusushai’s conclusions could help explain the difference between these 8th and 10th grade essays.

Another possible explanation for the observed proliferation of seemingly unconnected reasons and examples/explanations may illustrate difficulty among some of the lower scoring 8th grade writers who were not transitioning well from what Bereiter and Scardamalia (1987) call “knowledge telling” to the kind of “knowledge

transforming” required of higher scoring papers. Furthermore, Bereiter and Scardamalia (1982) speak about students’ tendency to develop ideas through the “what next” strategy, adding associative text without regard for its relevance or rhetorical purpose. Likewise, Goodin and Perkins (1982) note that students often develop ideas by stringing together “afterthoughts and digression,” a trend frequently evident among these essays (p. 70, in Larson, 1987). While the early but seminal work of Kellogg Hunt (1965) suggests that students from grade 4 on have at their command a full range of structures for developing essays (although he was addressing structural complexity), others have expanded his conclusions to argue that it is the students’ abilities to “manage” structures effectively and not the knowledge of these structures that increase with maturity (Applebee, 2002, p. 5). This study suggests that it is possible that middle and high school students do not understand how to develop or elaborate their ideas based on the rhetorical function of available strategies.

THE RELATIONSHIP BETWEEN DEVELOPMENT AND ORGANIZATION

Knowing what effect, if any, organization choices and development levels have on score is the focus of this study. One of the study’s major findings is the interrelationship between development level and each organizational strategy. Essays that were “other” organized were rated significantly higher on the development continuum for 10th grade, but moderately lower in the 8th grade. Interpretations of this observation must remain tentative, although this trend could have important implications. It could be that those who write formula essays are accustomed to a prescribed, “fill-in-the-blanks” process for elaborating their supports. In 8th grade, where the prompts elicited exposition, this seemed to provide an adequate framework for development. However,

those comfortable with allowing content to dictate form – those employing some “other” organizational strategy, especially among more mature writers – may have more freedom to develop their ideas as they see necessary, without feeling limited to a prescribed number of sentences within a prescribed structure. Regardless, this information does suggest one possible answer to the question “Does the FPT function more as a ‘scaffold’ or a ‘restraint’ for writers?” While the FPT may help writers establish a minimum for supporting their ideas among less mature writers, it may also restrict the development of ideas, especially among older writers responding to prompts that elicit more narration.

Other Features Associated with Score Categories

Length of Support Portion of Essay. For both 8th and 10th grade, this study suggests that there is a positive correlation between the mean number of sentences used in the support portion, or body portion, of the essays and DSTP score. Hillocks (1986) reports that word count, in and of itself, accounts for only 9 - 25 percent of variance in score. However, other studies (e.g., Bereiter and Scardamalia, 1987; Breland and Jones, 1984) have found that essay length is positively correlated with writing scores. This study only deals with number of sentences in the support section of the essay – a different measure of a limited slice; however, the results complement previous conclusions about length and perceived essay quality.

Use of Cause and Effect Statements. A noteworthy outcome of the study is the degree to which these writers relied on cause and effect statements for support of their ideas. Most often associated with explanations and/or examples, there were many statements that expressed causal relationships. Often these were also coded illogical, such as in the following example: *[If students are required to participate in after school*

activities]” ...Students will not have enough time to do our chores at home. If we don’t do our chores, our house will look obscene. Then our parents would blame us and we could get into trouble. If we get into trouble, our parents would also accuse the school which would give out school a terrible reputation,” Although this reliance was more pronounced with the 8th grade corpus, both groups of essays exhibited a similar pattern: the percent of cause and effect statements peaked in the mid-range score categories (4-7) and declined with the higher scoring papers. Figure 14 shows the distribution for both grades.

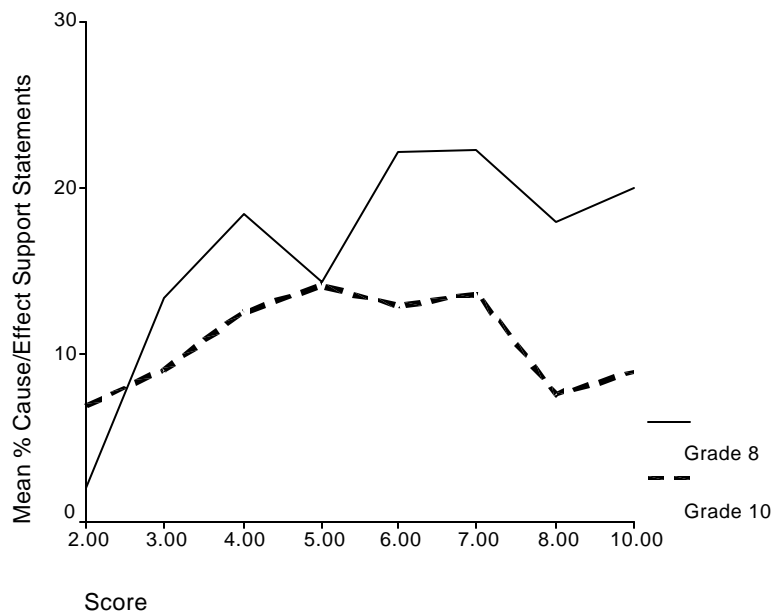


Figure 13. Comparison of cause and effect supports utilized by grade 8 and 10 writers.

There are several possible explanations to consider regarding this observation. According to Piaget (1974), children’s understanding of causality develops as they enter what he called the formal operations stage. Lavoie and Good (1988) identify the age at

which most children enter this stage as 14, usually around grade 9, which is roughly the stage these 8th grade writers are preparing to enter. Langer (1986) also specifically mentions causal relationships as among those that students first try at low levels of sophistication, which would indeed seem to be the case with many of the cause/effect statements used by these 8th grade writers. It is possible that grade 8 writers actually do rely more on cause and effect statements for support. Any such conclusion, however, would be tempered by a consideration of two alternative explanations regarding this phenomenon: first, it is possible that 10th grade writers are more adept at embedding statements about causal relationships in more subtle ways. The 8th grade cause/effect statements seemed very obvious, and often illogical, which may have made them more noticeable. It is also possible that, the 8th grade prompts actually invited more cause/effect statements in younger writers because they had less prior knowledge to call upon for support. Regardless, the observation is worthy of further investigation.

Trends among Top Scoring Essays

At the top scoring levels (“9” and “10” score points), for both grade levels, several strong trends emerged. First of all, nearly all of the grade 8 essays and virtually all of the 10th grade top scoring essays used some narrative to develop at least some of the essay. In essays coded “2” – “other organized” essays – writers frequently wrote full-blown narratives, usually with accompanying exposition. Even in those essays that were essentially highly formulaic FPTs, the writers managed to use narrative at least in the introduction or conclusion if not embedded within the body paragraphs.

Overall, according to this study, the highest scoring essays were more likely to be “other organized” as some version of a formula paper, for both grade 8 and grade 10 (see Figure 7 and 8).

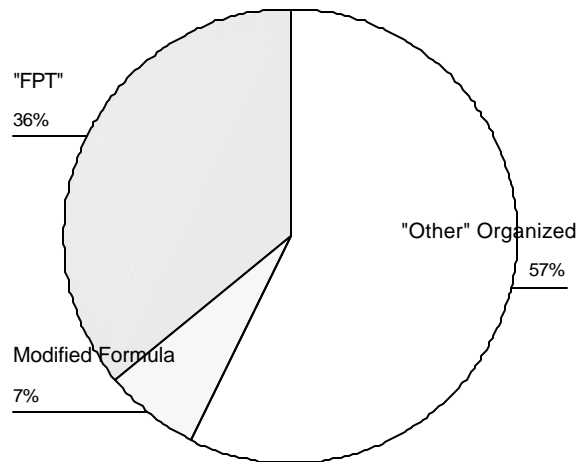


Figure 5. Pie chart for organization scheme for top-scoring grade 8 essays showing there were no short or unorganized essays among the essays earning “9s” and “10s.”

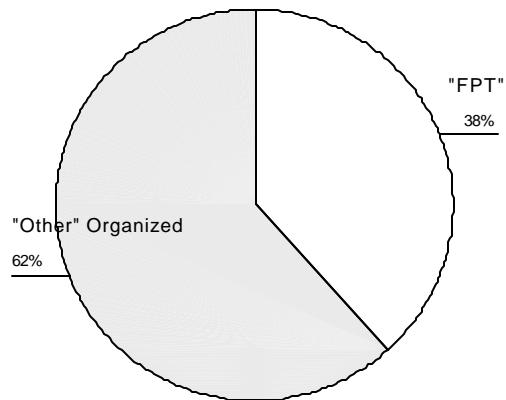


Figure 8. Organization schemes for 10th grade essays earning scores of “9” or “10.”

Another striking observation, particularly at the 10th grade level but also at the 8th grade level, was the fact that these writers introduced the main idea for *each* body paragraph in a subtle manner. Rarely did these writers place a direct statement of thesis, main idea, or topic in the preeminent position in the paragraph. More often, the idea was preceded with relevant and effective background information, a narrative example, commentary, or – less frequent but particularly effective – figurative language. One possible explanation for this observation comes from a seminal, albeit often abbreviated, 1974 study, by Richard Braddock. After analyzing professional writing, Braddock found that only 13% of the paragraphs had explicit topic sentences at all let alone topic sentences in the first sentence (Braddock also pointed out that many of these professional samples could have benefited from the addition of an explicit topic sentence). In light of Braddock’s still-unchallenged findings, one might theorize that these top-level writers are moving toward the stylistic sophistication of professional writers.

The use of figurative language, broadly defined to include literary (or other) allusion or metaphor, was the third most common characteristic of these top scoring essays, at both grade levels. The 10th grade writers’ use of figurative language (and the occasional use of facts and statistics) oftentimes demonstrated an exceptional amount of prior knowledge. Most of the essays in this top scoring category also made references to the global significance of their topic. These writers also frequently chose “different” subjects to explore, perhaps sensing that a well-written response that dealt with an unusual subject might receive a favorable score. Finally, the judicious and subtle use of humor also seemed to be common among many of the top scoring essays.

The use of specific details did not turn out to be as predictive as expected. Technically, this is a rubric requirement in Delaware for essays in the “4” or “5” category. However, this study suggests that not all writers in these categories (combined scores of “8”, “9” or “10”) consistently incorporated any specific details. One explanation could be that the anchor papers selected by Delaware teachers, which scorers rely on for scoring, do not reflect the specificity described in the rubric. Nonetheless, the most effective writers managed to weave both specific examples and explanations with with more general and global statements of significance; this strategy was not, however, a universal.

STUDY LIMITATIONS

Generalizing Results

The conclusions set forth in this study have credible applications only to the study situation: high-stakes, statewide, writing assessment such as the DSTP. Because assessment drives classroom instruction, conclusions do hold some promise for classroom application, but they are cautious conclusions. Related confounding factors include what McNamara (1996) calls a low sense of performance expectation and motivation to fulfill the purpose of the prompt when writers know they are writing to an inauthentic task (cited in Weigle, 2002, p. 46-47). In “real” classrooms, where many students feel free to take risks and experiment with form as well as content, the findings could be quite different.

Another caution must be made about the feasibility of generalizing results across discourse categories. The two 8th grade prompts were persuasive and informative respectively; the 10th grade prompts were expressive and informative respectively. Yet

the topics were quite different, possibly accounting for some of the choices in organizational scheme, as previously mentioned. It would be necessary to study many more writing samples in all three discourse categories, on a variety of topics, for both grade levels before one could generalize about writers' responses to various discourse categories.

Limitations Regarding Variables

One caution is that the data analysis used for much of this study deals with the interaction between feature and DSTP score. This assumes that the DSTP scores are accurate representations of students' writing performances on these tests, an assumption that cannot go unchallenged despite an assumed error rate of less than 10% for such high stakes assessments.

IMPLICATIONS

Impact on the Rhetoric and Composition Communities

For rhetoric and composition, this study has the potential of contributing to the field in numerous ways. Of preeminent importance is acknowledgment that formulaic writing is a by-product of the outdated current traditional rhetoric and the presentation mode of instruction in the classroom. The overwhelming presence of formulaic writing in response to this assessment does point to the existence of a current traditional rhetorical approach to writing instruction, a suggestion also noted by Penny (2003). For the composition field, this suggests that more of an effort is needed to help students make writing decisions based on an understanding of the rhetorical needs of the writing situation. Only then will assessments achieve the kind of validity we would like to see – validity based on a student's ability to think and construct meaning.

Impact on Pedagogy

Will these findings affect pedagogy? This study was not about writing instruction. However, scholarly research confirms that assessment does influence instruction; in fact, many assessment and composition scholars (e.g. Black and Wiliam, 1998; Huot, 2002; Weigle, 2002) predict that the future of writing assessment lies in its power to inform instruction. Therefore, even if only for the purpose of passing high-stakes writing assessments, this study can provide useful information for classroom teachers. At the least, the information gleaned from this study has the potential to help teachers make more efficacious decisions about their instruction. A lack of positive association between organization and DSTP score will neither persuade nor dissuade teachers from teaching and demanding formulaic writing from their students; however, it may change their rationale for doing so. Stating that the state “wants” or “does not want” to see FPTs would be simply untrue based on these reported results. In fact, these results question the efficacy of insisting upon five paragraph essays, suggesting instead that teaching judicious selection of and thorough elaboration of details may be more worthwhile, at least for the purpose of passing high-stakes writing assessments.

CONCLUSION

In summary, most composition theorists and practitioners agree that good writing is an interpretative act of meaning-making, situated in a particular context and involving, at least to some degree, a discovery process for both the reader and the writer (e.g., Wolcott, 1998). But teaching to formulas does not allow students to develop a repertoire of strategies. They are denied the opportunity to leave the classroom knowing how to tackle various writing occasions. Hillocks summarizes the problem of the FPT

this way: “The five-paragraph theme structure is one that is devised to fit any subject if one is willing to treat any subject superficially” (2002, p. 177). Furthermore, he concludes, that the FPT “...imposes not only a format but a way of thinking that eliminates the need for critical thought” (p. 136).

Potential exists for the descriptive and the qualitative findings in this study to suggest further research and to provide information to guide teachers in their writing instruction. Shirley Brice-Heath (in press) notes, “The making of meaning...is fundamentally, more often than not, an act of reconciliation. We work to reconcile disparities and incongruities just as we strive to fill gaps...” (p. 2). Indeed, despite the tentative nature of conclusions drawn about the writing in this study, the information gathered can, in a very practical sense, be used by teachers to guide their students as they struggle to balance the competing demands of learning-to-write, writing-to-learn, and learning to pass writing assessments such as the Delaware Student Testing Program.

APPENDIX A

DELAWARE STUDENT TESTING PROGRAM – GENERAL WRITING RUBRIC

The following characteristics determine the success of the response in meeting the needs of the audience and fulfilling the writing purpose.

Score of 5	Score of 4	Score of 3	Score of 2	Score of 1
<p>Score point 5 meets all the criteria listed in score point 4.</p> <p>In addition, a paper receiving this score shows an exceptional awareness of readers' concerns and needs.</p> <p>The student may have shown an exceptional use of:</p> <ul style="list-style-type: none"> - Development strategies specific to the purpose for writing - Distinctive style, voice, tone - Literary devices - Compositional risks 	<p>Unified with smooth transitions, a clear and logical progression of ideas, and an effective introduction and closing.</p> <p>Sufficient, specific, and relevant details that are fully elaborated.</p> <p>Consistently complete sentences with appropriate variety in length and structure.</p> <p>A consistent style with precise and vivid word choice.</p> <p>Few, if any, errors in standard written English that do not interfere with understanding.</p>	<p>Generally unified with some transitions, a clear progression of ideas, and an introduction and closing.</p> <p>Specific details but may be insufficient, irrelevant, or not fully elaborated.</p> <p>Generally complete sentences with sufficient variety in length and structure.</p> <p>Some style and generally precise word choice.</p> <p>Some errors in standard written English that rarely interfere with understanding.</p>	<p>Minimally unified and may lack transitions or an introduction or closing.</p> <p>Some specific details but may be insufficient, irrelevant, and/or not elaborated.</p> <p>Some sentence formation errors and a lack of sentence variety.</p> <p>Sometimes general and repetitive word choice.</p> <p>Several kinds of errors in standard written English that interfere with understanding.</p>	<p>Lacks unity.</p> <p>No or few specific details that are minimally elaborated.</p> <p>Frequent and severe sentence formation errors and/or a lack of sentence variety.</p> <p>Often general, repetitive, and/or confusing word choice.</p> <p>Frequent and severe errors in standard written English that interfere with understanding.</p>



APPENDIX B

Grade 8, 2001 (persuasive writing prompt):

WRITING PROMPT

Your principal read an article that said students who participate in after-school activities do better in school. The principal is considering requiring every student to stay after school and participate in at least one activity. Write a letter to your principal telling why you do or do not agree with this idea.

Thinking about the following can help you focus and plan your answer:

- *What are some reasons for after school activities?*
- *What are some reasons against after school activities?*
- *Which position can you support?*

Grade 8, 2002 (informative writing prompt):

WRITING PROMPT

Your school is having a writing contest entitled “What Makes A Good Teacher.” Write an essay for the contest identifying the qualities of a good teacher, and explaining why they are important.

Thinking about the following will help you focus and plan your writing.

- *What are the qualities of a good teacher?*
- *Why are these qualities important?*

Grade 10, 2001 (expressive writing prompt):

WRITING PROMPT

You are completing a job application. As part of the application process, your potential employer requires a writing sample explaining the expression, “experience is the best teacher,” and telling how it applies to you or someone you know. Write what you will present to your potential employer.

Thinking about the following will help you focus and plan your writing.

- *What might the expression, “experience is the best teacher,” mean?*
- *What are some experiences you have had (or someone you know has had) which taught you an important lesson?*
- *What did you learn and why was it valuable?*

Grade 10, 2002 (informative writing prompt):

WRITING PROMPT

The publishers of a new book, “Students View the Twentieth Century”, have invited students to submit articles on significant people or events of the twentieth century. Write an article that you could submit about a significant person or event.

Thinking about the following will help you focus and plan your writing.

- *What twentieth century event or person do you consider significant?*
- *In what ways is this person or event significant?*

REFERENCES

- Albertson, B. (2001). *Writing instruction and the DSTP*. Unpublished manuscript.
- Albertson, B. (2002). *Five-Paragraph Theme Efficacy Study*. Unpublished manuscript.
- American Educational Research Association. (2000). Position statement of the American Educational Research Association concerning high-stakes testing in preK-12 education. *Educational Researcher*, 29, 24-25.
- Amrein, A., & Berliner, D. (2002a). An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing. Education Policy Studies Laboratory EPSL-0211-125-EPRU. Available <http://edpolicylab.org>. Tempe, AZ: Arizona State University.
- Amerein, A. & Berliner, D. (2002b). *The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP test results in states with high school graduation exams*. Tempe, AZ: University of Arizona Education Policy Studies Laboratory (EPSL – 0211-126-EPRU)
- Amerein, A. & Berliner, D. (2003). Re-Analysis of NAEP math and reading scores in states with and without high-stakes tests: Response to Rosenshine. *Edcational Policy Analysis Archives*, 11(25) (ISSN No. 1068-2341)
- Applebee, A.N. (1991). Informal reasoning and writing instruction. J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 225-246). Hillsdale, NJ: Erlbaum.
- Applebee, A. N. (2002). Alternative models of writing development. In R. Indrisano & J. R. Squire (Eds). *Writing: Research/Theory/Practice* (pp. 90-111) Newark, DE: IRA. [On-line] Available: <http://cela.albany.edu/publication/article/writing.htm>.
- Applefield, J. M.; Huber, R., & Moallem, M. (2001) Constructivism in theory and practice: Toward a better understanding. *The High School Journal*, 84(2) 35-53.
- Avon Writing Team. (2000). Scaffold to success or constraint to creativity: A study of students writing to prompt. Avon, Connecticut: Avon School District.
- Ball, A. F. (1999). Evaluating the writing of culturally and linguistically diverse students: The case of the African-American vernacular English speaker. In Cooper and O'Dell (Eds.), *Evaluating Writing* (pp. 225-248). Urbana, IL: NCTE.
- Barksdale-Ladd, M. A., & Thomas, K. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education*, 51, 384-397.

- Beach, R. (1992). Experimental and descriptive research methods in composition. In G. Kirsch & P. A. Sullivan (Eds.), *Methods and methodology in composition research* (pp. 217-46). Carbondale, IL: Southern Illinois University Press.
- Belanoff, P. (1991). The myth of assessment. *Journal of Basic Writing*, 10(1), 54-66.
- Bereiter, C. & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in Instructional Psychology* (vol. 2). (p. 1-64). Hillsdale, NJ: Erlbaum.
- Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.
- Berlin, J. (1982). Contemporary composition: The major pedagogical theories. *College Composition and Communication*, 44(4), 765-767.
- Berlin, J. (1988). Rhetoric and ideology in the writing class. In V. Villanueva (Ed.), *Crosstalk in composition theory* (pp. 677-699). Urbana, IL: NCTE.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Phi Delta Kappan*, 80(2), 139-148.
- Borko, H., and Putnam, R. T. (1998). The role of context in teacher learning and teacher education." In *Contextual teaching and learning: preparing teachers to enhance student success in and beyond school. Information Series No. 376* (ERIC Document Reproduction Service No 429 263).
- Braddock, R. (1974). The frequency and placement of topic sentences in expository prose. *Research ion the Teaching of English*, 8, 287-302.
- Bramer, G. (1970). Like it is: Discourse analysis for a new generation. *College Composition and Communication*, 21, 347-355.
- Breetvelt, I., van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction*, 12, (2), 103-123.
- Breland, H. M., & Jones, R. J. (1984). Perceptions of writing skills. *Written Communication*, (1), 101-119.
- Brindley, R., & Schneider, J. J. (2002). Writing instruction or destruction: Lessons to be learned from fourth-grade teachers' perspectives on teaching writing. *Journal of Teacher Education*, 553(4), 328-341.
- Britton, J. (1978). The composing process and the functions of writing. In C. Cooper and L. Odell (Eds.), *Research on composing* (pp. 13-28). Urbana, IL: NCTE.

- Bruner, J. (1986). *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.
- Bullock, R., & Trimbur, J. (1991). *The politics of writing instruction: Post-secondary*. Portsmouth: Heinemann.
- Calfee, R. (in press). Writing assessment in large-scale contexts. In B. Guzzetti (Ed.), *Literacy in America: An encyclopedia*. New York: ABC Publishers.
- Cherry, R., & Witte, S. P. (1998). Direct assessments of writing: Substance and romance. *Assessing writing* 5(1), 71-87.
- Connor, U. (1996). *Contrastive rhetoric*. Cambridge: Cambridge University Press.
- Connors, R. J. (1981). The rise and fall of the modes of discourse. *College Composition and Communication*, 32, 444-55.
- Connors, R. J. (1983). Composition studies and science. *College English*, 45, 1-20.
- Cooper, C. R. & O'Dell, L. (1999). What we know about genre, and how it can help us assign and evaluate writing. In Cooper and O'Dell (Eds.), *Evaluating Writing* (pp. 23-52). Urbana, IL: NCTE.
- Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H. I. Braum (Eds.), *Test validity* (pp. 3-18). Hillsdale, NJ: Erlbaum.
- Crusius, T. W. (1989). *Discourse: A critique and synthesis of major theories*. New York: MLA.
- D'Angelo, F. J. (1987). Aims, Modes and forms of discourse. In Gary Tate (Ed.), *Teaching composition: 12 bibliographic essays* (pp. 131-154). Fort Worth, TX: Texas Christian University Press.
- Daniell, B. (1999). Narratives of literacy: connecting composition to culture. *College Composition and Communication*, 50(3), 393-410.
- Decker, R. E. & Schwegler, R. A. (1998). *Patterns of exposition*. New York: Longman.
- Delpit, L. (1988). The silenced dialogue: Power and pedagogy in education other people's children. *Harvard Educational Review*, 58(3), 280-298.
- Delpit, L. (1992). Acquisition of a literate discourse: Bowing before the master. *Theory into Practice*, 31, 296-302.
- Duffy, G., & Roehler, L. (1986). Constraints on teacher change. *Journal of Teacher Education*, 37(1), 55-58.
- Dyson, A. H. (1993). *Social worlds of children learning to write in an urban primary school*. New York: Teachers College Press.

- Ede, L. (1992). Methods, methodologies, and politics of knowledge: Reflections and speculations. In G. Kirsch & P. A. Sullivan (Eds.), *Methods and methodology in composition research* (pp. 314-332). Carbondale, IL: Southern Illinois University Press.
- Elbow, P. (1996). Writing assessment: Do it better; Do it less. In E. M. White, W. D. Lutz, & S. Kamusikri (Eds.), *Assessment of writing: Politics, policies, practices* (pp. 120-134). New York: MLA.
- Elbow, P. and Yancey, K.B. (1994). On the nature of holistic scoring: An inquiry composed on-mail. *Assessing Writing* 1(1), 91-108.
- Emig, J. (1971). *The composing process of twelfth graders*. Urbana, IL: NCTE.
- Erikson, F. & Gutierrez, K. (2002). Culture, rigor, and science in educational research. *Educational Researcher*, 30 (8), 21-24.
- Feurer, M. J., Towne, L. & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher*, 31 (8), 4-24.
- Fish, S. (1980). *Is there a text in this class? The authority of interpretative communities*. Cambridge, MA: Harvard University Press.
- Freedman, S. W. (1991). Evaluating writing: Linking large-scale testing and classroom assessment (Occasional Paper No. 27). Berkley, CA: University of California, Center for the Study of Writing.
- Gee, J. (1996). *Social linguistics and literacies: Ideology in discourses*. London: Taylor & Francis.
- Goodin, G., & Perkins, K. (1982). Discourse analysis and the art of discourse. *College English*, 44, 57-63.
- Graves, D. (2002). *Testing is not teaching: What should count in education*. Portsmouth: Heinemann.
- Halasek, K. (1999). *A pedagogy of possibility: Bakhtinian perspectives on composition studies*. Carbondale, IL: Southern Illinois University Press.
- Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing* 8(1), 5-16.
- Hamp-Lyons, L. & Condon, W. (2002). *Assessing the portfolio: Principles for practice, theory, and Research*. Cresskill, NJ: Hampton Press.
- Heath, S. B. (in press). Dynamics of completion: The acting artist as viewing other. In M. Turner (Ed.), *Art, brain, and cognition*. Oxford, England: Oxford University Press.
- Heilker, P. (1996). *The essay: Theory and pedagogy for an active form*. Urbana, IL: NCTE.

- Hillocks, G. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: NCTE.
- Hillocks, G. (1995). *Teaching writing as reflective practice*. New York: Teachers College Press.
- Hillocks, G. (2002). *The testing trap: How state writing assessments control learning*. New York: Teachers College Press.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (Research Monograph 3). Urbana, IL: NCTE
- Huot, B. (1990). Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication*, 41, 201-113.
- Huot, B. (1996). Toward a new theory of writing assessment. *College Composition and Communication*, 42, 425-445.
- Huot, B. (2002). *(Re)Articulating Writing Assessment*. Cedar City, Utah: Southern Utah University Press.
- Huot, B., & Schendel, E. (1999). Reflecting on assessment: validity inquiry as ethical inquiry. *Journal of Teaching Writing* 17(1&2), 37-55.
- Johaneck, C. (2001). *Composing research: A Contextualist paradigm for rhetoric and composition*. Logan, UT: Utah University Press.
- Johnson, T., Smagorinsky, P, Thompson, L, & Fry, P. (2003). The five-paragraph theme. *Research in Teaching English*, 38(2), 136-176.
- Kamurski, S. (1996). African American English and writing assessment: An Afrocentric approach. In E. M. White, W. D. Lutz, S. Kamurski (Eds.), *Assessing of writing: Politics, policies, practices* (pp. 1878-203). New York: Modern Language Association.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4) 319-342.
- Kaplan, R.B. (1966). Cultural thought patterns in inter-cultural education. *Language and Learning* 14, 1-20.
- Ketter, J. & Pool, J. (2001). Exploring the impact of a high-stakes direct writing assessment in two high school classrooms. *Research in Teaching English*, 35(3), 344-393.
- Kinneavy, J. L. (1971). *A theory of discourse*. Englewood Cliffs, NJ: Prentice Hall.
- Kirsch, G. (1992) Methodological pluralism: Epistemological issues. In G. Kirsch & P. Sullivan (Eds.), *Methods and methodology* (pp. 247-269). Carbondale, IL: A. Southern Illinois University Press.

- Knoblauch, C. H., & Brannon, L. (1984). *Rhetorical traditions and teaching or writing*. Portsmouth: Heinemann.
- Kohn, A. (1999). Getting evaluation wrong: The case against standardized testing. *The schools our children deserve* (pp. 73-91). New York: Houghton Mifflin
- Langer, J. (1986). *Children reading and writing: Structures and strategies*. Norwood, NJ: Ablex.
- Larson, R. (1987). Structure and form in non-narrative prose. In G. Tate (Ed), *Teaching Composition: 12 bibliographic essays* (pp. 39-82). Fort Worth, TX: Texas Christian University Press.
- Lauer, J. M., & Asher, J. W. (1988). *Composition research: Empirical designs*. New York: Oxford University Press.
- Lave, J. (1988). *Cognition in practice*. New York: Cambridge University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press.
- Lavoie, D. R., & Good, R. (1988) The nature and use of prediction skills in a biological computer simulation. *Journal of Research in Science Teaching*, 25, 335-360.
- Linn, R., Baker, E., & Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Lloyd-Jones, R. (1987). Tests of writing ability. In Gary Tate (Ed.), *Teaching composition: 12 Bibliographic essays* (pp. 255-176). Fort Worth, TX: Texas Christian University Press.
- Lunsford, A. (1986). The past – and future – of writing assessment. In K. L. Greenberg, H. S. Wiener, & R. A. Donovan (Eds.), *Writing assessment: Issues and strategies* (pp. 1- 12). White Plains, NY: Longman.
- Mabry, L. (1999). Writing to the rubric. *Phi Delta Kappan*, 80, 673-679.
- Madaus, G. F. (1988). The influence of testing on curriculum. In L.N. Tanner (Ed.), *Critical issues in curriculum: Eighty-seventh yearbook of the National Society for the Study of Education* (pp. 83-121). Chicago, IL: University of Chicago Press.
- Marder, D. (1974). The spectrum of rhetoric. *College Composition and Communication*, 25, 181-185.
- McAndrew, D. A. (1982). The effects of an assigned rhetorical context on the syntax
- Meade, M. S., & Ellis, W. G. (1971). The use in writing of textbook methods of paragraph development. *Journal of Educational Research*, 65, 74-76.

- Messick, S. (1988). The once and future issue of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H.I. Braum (Eds.), *Test validity* (pp. 33-45). Hillside, NJ: Erlbaum.
- Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18, 5-11.
- Messick, S. (1989b). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of Performance assessments. *Educational researcher*, 23(2), 13-24.
- Moffett, J. (1968). *Teaching the universe of discourse*. Boston: Houghton Mifflin.
- Moss, P. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62, 229-58.
- Moss, P. (1994). Validity in High Stakes Writing Assessment: Problems and Possibilities. *Assessing Writing*, 1(1), 109-128.
- Mountford, R. (1999). Let them experiment: Accommodating diverse discourse practices in large-scale writing assessment. In Cooper and O'Dell (Eds.), *Evaluating Writing* (pp. 366-395). Urbana, IL: NCTE.
- National Commission on Writing. (2003). *The neglected "R": The need for a writing revolution*. Princeton, NJ: CEEB.
- Needels, M. C., & Knapp, M. S. (1994). Teaching writing to children who are underserved. *Journal of Educational Psychology*, 86, 339-349
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19, 317-328.
- No Child Left Behind Act of 2001*, August 6, 2002.
- North, S. M. (1987). *The making of knowledge in composition: Portrait of an emerging field*. Upper Montclair, NJ: Boynton/Cook.
- Pajares, M. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 303-332.
- Penny, J. A. (2003). Reading high stakes writing samples: My life as a reader. *Assessing Writing*, 8(3), 195-215.
- Pirie, B. (1997). *Reshaping high school English*. Urbana, IL: NCTE.

- Potato, J. A. (1994). Written expression. In Choate, J. S., Enright, B. E., Miller, L.J., Poteen, J. A., & Rakes, T. A. (Eds.), *Curriculum-Based Assessment and Programming*. Boston: Allyn and Bacon.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer.
- Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational researcher*, 19(7), 10-18.
- Rogoff, B. (1990). *Apprenticeships in thinking: Cognition development in social context*. Oxford, England: Oxford University Press.
- Rose, M. (1981). Sophisticated, ineffective books: The dismantling of the composition process in composition texts. *College Composition and Communication*, 32, 65-73.
- Rose, M. (1988). Narrowing the mind and the page: Remedial writers and cognitive reductionism. *College Composition and Communication*, 39, 267-302.
- Rosenshine, B. (2003). High-stakes testing: Another analysis. *Educational Policy Analysis Archives*, 11(24) (ISSN No. 1068-2341).
- Schraver, K. A. (1990). Theory building in rhetoric and composition: The role of empirical scholarship (Technical Paper No. 38). Berkley, CA: University of California, Center for the Study of Writing.
- Schraver, K. A. (1992). Connecting cognition and context in composition. In G. Kirsch & P. A. Sullivan (Eds.), *Methods and methodology in composition research*, 190-216. Carbondale, IL: Southern Illinois University Press.
- Shafer, G. (2000). Composition for the Twenty-First Century. *English Journal*, 90(1), 29-33.
- Shepard, L. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.
- Sisserson, K.; Manning, C. K.; Knepler, A.; & Jolliffe, A. (2002). Authentic intellectual achievement in writing. *English Journal*, 92(6), 63-69.
- Stiggins, R. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. (2002). Where is our assessment future and how can we get there from here? In R. W. Lissitz & W.D. Schafer (Eds.), *Assessment in educational reform* (pp.18-49). Boston, MA: Allyn & Bacon.
- Street, J. C., & Street, B. V. (1991). The schooling of literacy. In D. Barton & R. Ivanic (Eds.), *Writing in the community* (pp. 143-166). Thousand Oaks, CA: Sage.

- Trimbur, J. (1996). Response: Why do we test writing? In E. M. White, W. D. Lutz, & S. Kamusikri (Eds.), *Assessment of writing: Politics, policies, practices* (pp. 45-48). New York: Modern Language Association.
- Wesley, K. (2003). The ill-effects of the five paragraph theme. *NCTE Inbox*, 2 (41). Available: <http://www.ncte.org/pdfs/members-only/ej/0901-sept00/EJ0901/Effects.pdf>.
- Weigle, S.C. (2002). *Assessing writing*. Cambridge, UK: Oxford University Press.
- White, E. M. (1993). Holistic scoring: Past triumphs, future challenges. In Williamson, B. & Huot, B. (Eds.), *Validating holistic scoring for writing assessment* (pp. 79-108). Cresskill, NJ: Hampton Press.
- White, E. M. (1994). "Issues and Problems in Writing Assessment" *Assessing Writing*, 1(1), 11-28.
- White, E. M. (1996). Response: Assessment as a site of contention. In E. M. White, W. D. Lutz, & S. Kamusikri (Eds.), *Assessment of writing: Politics, policies, practices* (pp. 301-304). New York: Modern Language Association.
- Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass.
- Wiggins, G. (1994). The constant danger of sacrificing validity to reliability: Making writing assessment serve writers. *Assessing Writing*, 1(1), 129-140.
- Williamson, M. M. (1994). The workshop of efficiency: Untangling theoretical and Practical considerations in writing assessment. *Assessing Writing*, 1, 147-193.
- Wolcott, W. & Legg, S. M. (1998). *An overview of writing assessment: Theory, research, and practice*. Urbana, IL: NCTE.
- Yancey, K. B. (1999). Looking back as we look forward: Historizing writing assessment. *College Composition and Communication*, 50, 483-503.