Learning to Write in a Genre: What Student Writers Take from Model Texts

Davida H. Charney
Richard A. Carlson
The Pennsylvania State University

This study investigated the effects of writing models on students' writing of research texts. The models used by participants varied in quality and in labeling cues. Ninety-five psychology majors were given basic facts, including relevant and irrelevant information, for writing a Method Section for one of two experiments. The control group (N = 22) saw no models. The models groups (N = 73) saw three student-written Method sections—either 3 good models (AAA) or 1 good, 1 moderate, and 1 poor model (ABC). Half of each quality group saw the models labeled with grades; the other half saw them unlabeled. Following holistic ratings of the students' texts, the texts were analyzed for content. The models groups' texts were rated as better organized than those of the control group. The models also influenced text content. Seeing a proposition in the models increased the likelihood that students would include it in their texts, with the effect being smaller for propositions that appeared only in moderate or poor models. For the writing topic deemed more difficult, the models group included more topical information than the control group, including more essential propositions but also more unnecessary propositions. No systematic benefits emerged from labeling the models or from providing only good models. Students seemed able to judge the relative quality of the models, even without labels. Overall, providing models seems to increase the salience of the topical information considered by student writers for inclusion in their texts.

Every discipline, profession, and business has its own genres or conventional forms of written communication—from legal briefs and résumés, to computer manuals, and journal articles. Genres arise from the frequent recurrence of similar situations in which writers seek to move readers to some specific end or in which readers need specific kinds of information from writers in order to take some action (Miller, 1984). Adopting a genre relieves writers and readers from having to invent a new form of communication for each recurrence of a situation; topic selection, argumentative strategies to emphasize the importance of the new work in the context of the existing literature; structural arrangements, and stylistic patterns that proved useful for conducting such business in the past are incorporated into a readily recognizable pattern for communication. This pattern includes assumptions about the goals, prior knowledge, tastes, and power relations of the writers and readers. For example, Bazerman's (1989) insightful historical analysis of the development of the experimental research article shows how this genre was shaped by the growth of shared knowledge and the shifting of power relations between scientific writers and readers in the 18th century.

To write successfully in a genre, a writer must be familiar with its conventions of content, structure, and style, as well as understand the assumptions underlying these conventions. A writer must also know how to adapt these constraints to fit the peculiarities of the task at hand. For example, a researcher reporting an experiment must know much more than the conventional written outline of Introduction, Method, Results, and Discussion sections and the aims and details of the experiment itself. The researcher must also use rhetorical strategies to emphasize the importance of the new work in the context of the existing literature; to recognize when departures from standard practice are great enough to require explanation or justification; to anticipate readers' rival interpretations of the data. As studies of expert composing processes have revealed, even prominent scientists with numerous publications to their credit often struggle when writing experimental articles (Law & Williams, 1982; Rymer, 1988).

The task is obviously even more challenging for undergraduates who are often novice writers, novice researchers, and novices in the discipline. Students have difficulty not only learning the structure of the genre itself but also adopting the "voice" of a researcher (Faigley & Hansen, 1985; Herrington, 1992). They have little of the tacit experiential knowledge that full-fledged scientists rely on for adapting the genre to a particular experiment or for anticipating what readers will find interesting or controversial. For example, while students writing a Method section know that they must include "enough detail to allow a reader to replicate the experiment," they often cannot determine which details are required to assure an acceptable replication. Often students treat generic conventions as a Procrustean bed, distorting their material to fit the outline, rather than bending the rules.

Yet, writing in the genre of the research report is often an important means of progress in a discipline. As Herrington (1992) has argued, in order for students to internalize the language, values, and strategies of their future discourse community, they must "insert" themselves into that community, trying out the characteristic ways in which members of that discipline think and express themselves to one another. The anthropology students in Herrington's study found that writing experimental research reports—and rewriting them after feedback from instructors—

made them more aware that tacit disciplinary conventions suffused the language they were trying to adopt. Their efforts to sound like anthropologists changed their perceptions of their own projects. Assigning research reports has become a central way to teach students about a discipline, not only in advanced composition classes, but also in research methods classes in a variety of scientific and technical disciplines. Given the importance of this writing task, and the notorious difficulties it poses to beginning students, facilitating the process would obviously be beneficial.

Model Texts

A common technique for teaching genres such as the experimental research report is to present students with model texts that can be imitated or drawn on while students are writing their own texts. Although types of model texts vary, we will define a model as *a text written by a specific writer in a specific situation that is subsequently reused to exemplify a genre that generalizes over writers in such situations.* Such models are often used to supplement explicit guidelines or "rules" (provided in a textbook or style guide) for spelling out some of the conventional features of the genre; in a recent survey as many as 76% of university-level composition instructors reported using models regularly (Stolarek, 1994).

Models are as common in the workplace as in the classroom. Beginning engineers rely heavily on company files of technical reports when writing their reports, and it is quite common for supervisors to provide such reports as models (Winsor, 1990a; 1990b).

But while the use of model texts is widespread, Hillocks (1986) points out that very little is known about their effectiveness. In fact, his meta-analysis of a series of studies suggests that models may well be less effective in writing instruction than some other kinds of intervention. Further, as many teachers know first-hand, students often misuse models, imitating their weaknesses as well as their strengths, or applying the model inappropriately or too literally to their particular rhetorical situation (Smagorinsky, 1992; Werner, 1989).

The mixed record produced by model texts is somewhat surprising, given that learning-by-example has proven such a powerful pedagogical strategy in other skill domains (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Nitsch, 1977; Pirolli, 1991; Sweller & Cooper, 1985). The difficulty may lie in the nature of writing tasks. Writing tasks are less well-defined than other problem solving tasks (like playing chess or solving algebra problems) that have been extensively studied by cognitive and educational psychologists. The conventions of a genre are not hard-and-fast rules, although they are often presented as such in style guidelines or publication manuals. Thompson (1993) contrasted the "rules" for writing an experimental Results section with the actual practice of a senior biochemist, finding the biologist's frequent and systematic "violations" rhetorically well justified. Furthermore, writing problems have no definitively correct or incorrect solutions. Any given writing task will elicit a range of acceptable responses, none of which is ever absolutely perfect. No two writing tasks are ever exactly alike; a text that is an effective response in one rhetorical situation may be completely inappropriate in a similar situation.

As a result of the nature of the task, a model text is not an algorithm for writing a new text in the way that a worked-out algebra problem may be a model for solving new problems of the same type. Writers who have a model text nevertheless face some complex interpretive challenges. First they must decide which aspects of the model are truly representative of effective texts in that genre. Second, they must consider the situation in which the model was written and assess how it differs from their own. Finally, their judgments may not be confined to what they actually see in the model; what is properly excluded in one situation may be essential in another. So any single model may not be very useful when compared to a variety of models that better represent the range of acceptable variation.

Apart from differences between the model and the writing topic, the way that models are selected may also influence their usefulness. Smagorinsky (1992; in press) reviews the mixed record on models for writing, pointing out that in several studies in which models failed to improve performance, the models were available to participants only before, and not while, they wrote. In some cases, the writers themselves may have lacked sufficient content knowledge to take advantage of the models. Smagorinsky also found that more successful models targeted one or two writing features rather than a whole complex of features. Overall, however, the number of studies investigating models is surprisingly small. Little research is available on the specific effects of models on the writing process or on the effects of various kinds of models.

The Present Study

This study investigated the usefulness of models to students learning a new genre in their discipline and the effects of different kinds of models. We asked undergraduate psychology majors to write a Method section for a simple experiment, allowing some students to consult student-written Method sections while writing. We focused on the Method section because its structure and content are relatively well-defined, while leaving sufficient room for variation. Writing a Method section also requires an understanding of sophisticated disciplinary concepts.

This study addressed three main research questions: 1) Do writers provided with models produce better texts than writers working without models? 2) Do writers benefit more from seeing several high-quality models or from models representing a range of quality levels? and 3) Do writers benefit from explicit cues about the quality of the models? The first question addressed the basic issue of whether models help or hurt student efforts to write in a genre. The second question is of interest because writing tasks are by nature ill-defined. Because no one model can adequately represent the range of acceptable responses, one might expect that students would benefit most from seeing several good models and from observing the range of variation among them. But research has also shown that learners benefit from seeing counterexamples, examples of unsuccessful or wrongheaded efforts (Nitsch, 1977; Tennyson, 1973; Tennyson, Woolley, & Merrill, 1972). Therefore, comparing models of different quality may help students identify the strengths of the models and avoid the weaknesses. The last question addressed the issue of how best to encourage active analysis of the models. Labeling the models may allow students to passively accept everything in the "good" models as good rather than taking a critical stance. On the other hand, students lacking familiarity with the discipline and the genre may be unable to judge unlabeled models appropriately; in this case, the models may simply serve to confuse them. Given these different theoretical possibilities, our study sought to identify what kinds of models—if any—are most helpful for student writers.

Method

Participants

Ninety-five undergraduates (mainly sophomores and juniors) enrolled in Psychology 201 (a required research methods course for psychology majors) participated in this study. A primary objective of Psychology 201 is to familiarize students with the genre of the experimental research report by means of both reading and writing. At the time the study was conducted, about half-way through the term, students had already received initial instruction in the purpose and components of the Method section. And they had previously taken a general introductory psychology course that broadly surveyed areas of psychological research. Students who volunteered to participate in the study received extra class credit.

Materials

Writing Topics. The students' task—to write a Method section for a re-enacted psychological experiment—is a common assignment in introductory research methods courses. Because writing tasks are known to be sensitive to details of topic, we prepared materials for two experiments so that we could check for topic effects. We based the two topics on published studies of the effects of prior knowledge on memory for verbal information. In our re-enactment, we simplified and modified the studies to resemble each other more closely. Apart from the experimental materials, most of the essential features of the two studies (the number of subjects, the treatment structure, and the major events in the experimental procedures) remained identical.

The first experiment ("Washing") is based on a familiar study by Bransford and Johnson (1972). In this study subjects are presented with a vague passage describing the process of washing clothes. One group of subjects sees the passage with a title that activates prior knowledge about doing laundry, clarifies the passage, and leads to more accurate recall of its contents. The other group sees the passage untitled. In our re-enactment of this study, we added a within-subjects practice variable in which both groups have two chances to read and recall the passage.

The second experiment ("Hierarchies") is based on a study by Bower and his colleagues (1969). Here the task is to memorize sets of common nouns. One group is presented with the words organized into hierarchical, tree-structure diagrams with the most general term at the top (metals) and branches of more specific categories (rare, common, and alloys) followed by specific instances (gold). The other group sees the same words randomly scattered in the tree-structures. The complication in this study is that each group studies two sets of hierarchies, one of minerals and one of plants. Like the Washing study, the design included a within-subjects practice variable in which each group has two chances to read and recall the two hierarchies. The design is formally identical to the Washing study, a two-level, between-subjects variable for the materials (organized or random) and a two-level within-subjects practice variable (first or second study-recall trial).

Content Information. To enable students to write Method sections about these experiments, we provided them with sufficient detailed content information. We could not provide the information in connected prose, however, without obscuring the students' own decisions about what to say and how to say it. Therefore, we produced brief, video-taped "documentaries" that explained and dramatized the events of the experiments. The documentaries were similar to the kinds of in-class simulations of experiments that are used to teach students about various research methods.

We also provided a fact sheet, relieving students of having to memorize the details of the experiment while watching the videotape. Like the videotape, the fact sheet included irrelevant as well as relevant information—some facts that belonged in the Method section, some that belonged in other sections of an experimental report (e.g., results or hypotheses), and some that were inappropriate altogether (e.g., trivial details such as the name of the store where the paper was purchased or that a coin-toss was used to randomly assign subjects to conditions). We randomly ordered the items on the fact sheets. We also gave students other materials to help them understand their experiment. For example, students assigned to write about the Washing study were given the passage in both its titled and untitled versions. Students writing about the Hierarchy study received the plant and mineral tree structures in both organized and random forms.

Providing this material focused the writing task on the processes of selecting, arranging, and expressing information, rather than on retrieving information from memory or generating new ideas. This focus seemed most appropriate for students who are still learning the necessary concepts and strategies for designing their own experiments.

Writing Models. To provide all participants with roughly the same information, we used texts about one topic as models for the other. That is, participants who wrote about the Washing study saw models that were about the Hierarchy study; participants who wrote about the Hierarchy study saw models about the Washing study.

Each participant received three models, but some participants got three good models while others got one good, one intermediate, and one poor model. We therefore needed a total of five models for each topic. To create naturalistic variations in quality, we based the models on experimental reports written by students in previous semesters of Psychology 201, selecting papers that had received high, intermediate, and low scores for the Method section. We used these papers as "templates" for our models, closely matching their content, sentence structure, organization, and style. We wrote two versions of each model, one describing the Washing study, the other the Hierarchy study. Appendix A includes one of the A models, the B model, and the C model for the Hierarchy topic.

The A models were all competent student papers, displaying a clear understanding of the design of the experiment and the purpose of each part of the Method section. While each of the A models conveys the gist of the experiment, they differ in amount of explanation and formality of the writing. Each one contains a few flaws, such as missing or misplaced information. The B model displays good control of sectioning; even though it lacks explicitly labeled subsections, it is organized into fairly coherent paragraphs corresponding to subjects, materials, design and procedure. It describes the design of the experiment less clearly than the

A models, omitting more of the necessary details and including a good deal of unnecessary detail (e.g., where subjects sat in the room). The C model has several serious deficiencies, including information appearing in the wrong section, uneven and inappropriate level of detail (both insufficient and excessive detail), and some mechanical and stylistic infelicities. A detailed analysis of the content of all ten models will be presented below (Table 2).

Design and Data Analysis

Participants were randomly assigned to a topic, Washing or Hierarchy, and to a model condition—either to a No-Model control group (N = 22) or to 1 of 4 Models groups (N = 73) as summarized in Table 1. The 4 Models groups represented a 2 × 2 factorial design which varied the quality of the models (AAA or ABC) and their labeling with a grade (Labeled or Unlabeled). In the AAA group, participants saw the 3 good models; participants in the ABC group saw 1 good model, 1 intermediate model, and 1 poor model. The three A models were interchanged in the ABC sets so that each appeared equally often. Half of each group saw models labeled with letter grades (A, B, or C), and half saw them unlabeled.

With 4 model conditions and 1 control condition, the design of the study is not completely orthogonal. We structured the analyses of variance (ANOVAs) to treat Model and No Model as levels of a Presentation factor, with Quality (AAA and ABC) and Label (Labeled and Unlabeled)

Table 1

Experimental Design with Number of Subjects Per Condition and Mean Final Grade in Psychology 201*

	NO MODELS	MODELS			
		Labeled		Unlabeled	
		AAA	ABC	AAA	ABC
Washing Topic					
N	11	10	9	9	10
Grade	834	862	776	823	816
Hierarchy Topic					
N	11	9	9	9	8
Grade	859	829	837	847	879

*Max. grade = 1000, not significant at p = .05.

crossed within the model level. These factors were all between group. The ANOVAs tested for differences between the No Models group and the combined Models groups. They also tested for main effects and interactions of Quality and Label within the Models group. T-tests compared individual Models groups to each other (where significant interactions were found in the main analysis) and to the No-Models group. By crossing Topic (Washing and Hierarchy) with Presentation (No Models and Models), we also tested for topic effects.

Finally, we factored the students' final grades in Psychology 201 into our analyses. Students in this course are graded on a point system with a maximum of 1000 points. We found the median grade for our participants (855 on a 1000-point scale) and assigned students above the median to a high grade group ($N = 47$) and students at or below the median to the low grade group ($N = 48$). In the results reported below, we used these groupings to examine whether models had the same effect for students of high and low proficiency in psychology. We also used students' grades to check the reliability of our procedure for randomly assigning participants. We conducted an analysis of variance with grade as the dependent measure to see if students in different treatment conditions were roughly equivalent in proficiency; no significant differences were found (Table 1).

Procedure

All participants had the same task: to write a Method section for an experiment. Participants in the Models groups began by reviewing the writing models. We told the participants that the models might help them write their own Method sections but that the models were not perfect guidelines because they described a different experiment and because they were chosen at random from papers written by previous Psychology 201 students. The students could keep the models and refer to them while writing. All participants (control and models groups) then reviewed the fact sheets and materials for their topic. They were told that the fact sheets contained both relevant and irrelevant details listed in random order. Participants then watched the videotape; they were allowed to take notes. At that point, they were given an hour to write their Method section. We encouraged participants to use whatever composing process they preferred: outlining, multiple drafts, and so on. All participants finished within the allotted time.

Prior to analysis, the students' texts were typed to eliminate handwriting and neatness effects. The students' texts were then rated holistically and analyzed for specific content, with the goal of tracing the effect of models on students' selection and arrangement of information.

Holistic Assessment of Student Texts

As one measure of overall quality, the texts were assessed holistically. Four graduate teaching assistants from Psychology 201 independently rated each student text on four qualitative scales: inclusion of relevant information, exclusion of irrelevant information, organization into sections, and elaboration. Apart from elaboration, all were 5-point Likert-scales (with 1 representing poor performance and 5 excellence). The elaboration scale had 9 points, 1 representing too little detail, 9 too much detail, and 5 excellence). We have averaged the scores of the four raters in the analyses presented hereafter. The scores of all the raters positively correlated, but overall agreement proved relatively low using Cronbach's alpha (relevant content, 0.49; irrelevant content, 0.52; organization, 0.76; elaboration, 0.57). While the low correlations are disappointing, they do not invalidate further analysis. Because the scores are combined, they represent a lower bound to the true reliability of the total set of ratings (Bohrnstedt, 1983). The ANOVAs provide an additional (typically conservative) test of the trustworthiness of the ratings. Lower inter-rater reliabilities would tend to obscure "real" differences between the variables (i.e., producing Type II errors), rather than producing spurious differences (Type I errors).

Analysis of the Content of Student Texts

To assess the content of the students' texts and the possible influence of the models on content, we needed a more sensitive measure than holistic scoring. We wanted to determine what information students had included, assess how much of it was appropriate, and trace whether the models influenced the selection and arrangement of information. Rather than attempting to account for everything in the students' texts, we scored the texts for the presence of a set of propositions—in particular, those ideas that we had presented to the students as background for writing their texts.[1]

To capture as much as possible of the information we had presented to the students, we created a list of propositions for each topic drawn from the fact sheets, the models, and the transcripts of the videotapes. The resulting 86 propositions for the hierarchy topic are presented in Appendix B under the section headings Subjects, Materials, Design, and Procedure. (The proposition list for the Washing topic included 81 very similar items.) The propositions in each section include both relevant and irrelevant details. For example, the Subjects section draws from the fact sheet irrelevant details about the subjects' professors and places of residence, as well as important details like the number of subjects, their student status, and their native language. Some propositions appear under more than one heading. For example, the low-quality C Model includes some reference to the fact that subjects "were sent to different parts of the room" in its Subjects section, while the A1 model and the B model include this reference in the Procedure section. The proposition is accordingly listed under both sections.

Two independent raters scored the students' texts for the presence or absence of the propositions (1 = present, 0.5 = partial credit, 0 = absent). Texts were scored section-by-section, checking whether or not the student's text contained the propositions listed for that section. A proposition was scored as "present" if the text explicitly conveyed the idea, regardless of whether it was expressed in one connected sentence or in phrases. High levels of agreement between the raters were obtained: Subjects at 0.96; Materials at 0.96; Design at 0.95; and Procedure at 0.89 using Cronbach's alpha.

As described in more detail below, we categorized each proposition in two ways: context and relevance. The *context dimension* reflects which models (if any) included a given proposition. The *relevance dimension* reflects whether experienced psychologists judged a proposition important to include in that section. The relevance and context categorizations are alternative ways of characterizing the propositions available in this experimental setting. These categories permit investigation of some factors that might have influenced students to include certain facts in their texts.

Propositional Context. This analysis investigated whether propositions that had appeared in the models were more likely to show up in the Models groups' texts than in the control group's texts. Further, we sought to determine whether propositions that had appeared in the good models were more likely to be included in student texts than those from the poor models.

First, we scored the models themselves for the presence or absence of each proposition. By examining which propositions occurred in which models, we identified sets of propositions in four contexts: 1) those that occurred across quality levels, that is, in the A models as well as the B or C models;[2] 2) those that occurred only in A models; 3) those that occurred only in the B or C models; and 4) those that appeared in none of the models. Appendix B indicates the context category of each proposition.

For each student, we then calculated the average score for the propositions in each context category. It is useful to think of these scores as the probability that a proposition from that category would be included in the students' text. For example, a score of .90 in the A's Only category would indicate that the propositions in that category were very likely to be included, while a score of .10 would mean that each proposition had only a small chance of inclusion.

The context measure can be considered a measure of the obviousness of the propositions (and whether seeing them in the models made them even more obvious or plausible). All the students—even those in the control group—were exposed to the ideas in all four categories by watching the videotape and seeing the materials. All students also learned in a general way in their introductory psychology class what kinds of information to include in each part of the Method section. Clearly, some propositions were very obvious, and likely to occur to any student, as necessary to include. For example, virtually every student included the number of participants in the Subjects section. Other propositions may have seemed just as obviously inappropriate to include in a particular section, such as including the fact that students were sent to different parts of the room in the Subjects section.

Because the control group did not see any of the models, their scores represent in some sense the baseline obviousness or salience of the propositions—that is, the likelihood that a student would have decided to include a proposition on the basis of the material we presented, without seeing any models. In the analyses to be presented, we report the average scores for each context category. The control group's scores, therefore, indicate the average salience of the propositions in that context. The scores for students in the various model conditions reflect possible changes in salience due to seeing propositions in different models. Higher scores in these analyses do not represent "better" performance because the context categories include unpredictable mixtures of relevant and irrelevant propositions. Rather, the important consideration is the comparison between the scores for the control group and the Models group—that is, whether seeing the models changed the probability that those propositions would be included relative to the baseline established by the control group. Comparisons among the Models groups are also useful. Comparing the probabilities in the A's Only and the BC Only categories, for example, indicates whether students were influenced more by the higher quality models than by the poorer quality models.

Relevance Judgments. This analysis sought to investigate whether models helped students include more relevant information than the no-models control group, and if so, what kinds of models had the greatest effect. We asked four experienced experimental psychologists to rate each proposition on a 5-point scale (5 = essential, 4 = relevant, 3 = unnecessary, 2 = extraneous, and 1 = misplaced). The level of agreement was high overall and for each section (overall agreement, 0.85; Subjects, 0.69; Materials, 0.97; Design, 0.82; and Procedure, 0.81 using Cronbach's alpha). Summing the four ratings produced a ranking of propositions, ranging from a possible score of 20 to a low of 4. On the basis of these scores, we created three rather conservative relevance categories: essential/relevant (16–20 points), unnecessary (12–15 points), and extraneous/misplaced (4–11 points). Appendix C indicates the resulting relevance category for each proposition.

For each student text, we counted the number of essential, unnecessary, and extraneous propositions. For example, the following excerpt from a student in the ABC group contains many of the essential propositions, but omits exactly what the subjects were told the study would be about and how subjects were debriefed. It also contains some propositions categorized as unnecessary and extraneous, such as the facts that subjects were sent to different parts of the room and that the randomization was achieved by flipping a coin.

> Procedure
>
> To begin, the subjects were asked to read and sign consent forms from the Office for Protection of Human Subjects. To randomly assign the subjects to each presentation condition, 2 subjects were paired together. A coin was then flipped to designate the presentation condition for each. This procedure was repeated for all 25 groups of 2 until 25 subjects were assigned to each presentation condition. The 2 groups were then put on different sides of the room. The materials for both groups were passed out and general instructions about the experiment were given. The testing was ready to begin.
>
> Subjects were instructed to turn over their top sheet and study the passage. Cond. A subjects had a titled passage cond. B sub. did not. After 2 minutes of study, subjects had 3 minutes to write down all of the complete ideas they could remember from the passage on the back of their paper. They then were instructed to turn over the 2nd sheet of paper and again study the passage for a two minute time period. After the time limit expired, subjects were again asked to write down with in 3 minutes all of the complete ideas they could recall. Upon completion, the testing was complete.

The scores for essential, unnecessary, and extraneous propositions do not account for everything the students wrote. The score for essential propositions is a fairly good measure of whether students accurately conveyed the gist. However, the unnecessary and extraneous categories are obviously not exhaustive—the set of all possible intrusions is unbounded. For example, in the excerpt above, the sentences describing the testing as "ready to begin" and then as "complete" are not accounted for in any way. The scores in these categories also do not account for redundancy, wordiness, and misplacement of relevant items from one section in another (except where this explicitly occurred in the models). Therefore, the unnecessary and extraneous scores underestimate the number of intrusions. What they do represent is the students' suscep-

Table 2

Proportion of Available Propositions in Models by Relevance Category and Topic

	MODELS					
	A1	A2	A3	Total in A's	B	C
RELEVANCE						
Essential						
Washing	.67	.58	.67	.88	.47	.44
Hierarchy	.72	.61	.63	.87	.29	.40
Unnecessary						
Washing	.21	.29	.36	.57	.64	.29
Hierarchy	.21	.29	.36	.57	.64	.32
Extraneous						
Washing	.06	.15	.15	.24	.03	.21
Hierarchy	.12	.12	.15	.35	.06	.29
Words						
Washing	270.0	259.0	339.0		283.0	227.0
Hierarchy	388.0	361.0	459.0		336.0	323.0

tibility to the specific "red herrings" included in the models and other materials.

We used the same categories to assess the content of the models themselves, scoring the models for the presence or absence of the propositions in each category. Table 2 presents the results for the 10 models, 3 A models, 1 B, and 1 C with a close variant for each topic. This analysis indicates that efforts to vary the quality of the models appear reasonably successful. The A models are very similar to each other in the number and kinds of propositions they contain. Each A model contains more essential propositions than do the B and C models. Because the A models share some propositions but not others, the table also shows how many propositions appear in at least one A model. This indicates that the students who saw all three A models were actually exposed to nearly 90% of the total essential propositions. The B models contain a high proportion of unnecessary propositions, and the C models, the most extraneous propositions. However, the B and C models do not differ much in the proportion of essential propositions. Like most good student papers, the A models are not ideal in terms of propositional content; no paper contained all of the essential propositions and each contained a fair number of unnecessary and extraneous propositions.

Results

Holistic Ratings

Of the four holistic scales, only the organization scale revealed significant overall differences between the Models Group and No-Models group. Students in the Models group received higher scores than students without models [Models = 4.1, Control = 3.7, $F(1, 80) = 6.4$, $p < .01$]. This result may reflect better organization of sentences within sections as well as inclusion of information in the appropriate section. The holistic ratings were not affected by the variations in the quality and labeling of the models. The holistic rating scales did register differences in the students' proficiency in psychology. Students with higher final grades in Psychology 201 received significantly higher ratings ($p < .05$) on inclusion of relevant information, exclusion of irrelevant information, and organization, and marginally higher ratings on elaboration ($p < .10$) than students with low final grades. This suggests that the scales proved to be fairly sensitive despite relatively low interrater reliabilities.

Propositional Context

This analysis investigated whether students were more likely to include a proposition in their texts if they had seen it in a model. It also investigated whether students treated propositions that had appeared in the A models differently from those in the B and C models. Propositions occurred in four contexts: in the A models as well as the B or C models; only in the A models; only in the B or C models; and in none of the models. The data represent the *average probability* that the propositions in these contexts were included in a student's text. The first question to be addressed is whether the models changed the likelihood that a proposition would be included relative to the control group.

Overall Comparisons of the Models and No-Models Groups. Table 3 presents the average probability that propositions in the four context categories appeared in a student's text, comparing the control and Models groups. Overall, seeing models increased the probability that propositions would be included [No Models, .36, Models, .43, $F(1, 91) = 13.4$, $p < .01$].[3]

Seeing propositions in the models increased their likelihood of inclusion (as compared to the control group)—no matter whether the propositions appeared in high or low quality models. The effect is significant for the All-Models context [$F(1, 86) = 15.3$, $p < .01$], for the A's-only context [$F(1, 86) = 10.3$, $p < .01$], and for the BC-only context [$F(1, 86) = 6.4$, $p < .05$]. In contrast, propositions that did not occur in any models (No Models context) had just the same likelihood of appearing in the Models group's texts as in the control group's texts. In other words, seeing a

Table 3

Inclusion Probability (SDs) for Propositions in Different Contexts in Student Papers Produced With and Without Models

	NO MODELS	MODELS	*Difference*
CONTEXT			
All Models	.46 (.20)	.57 (.21)	*.11*
A Models Only	.42 (.23)	.57 (.25)	*.15*
BC Models Only	.31 (.16)	.38 (.11)	*.07*
No Models	.24 (.12)	.21 (.09)	*–.03*
Marginals	*.36*	*.43*	

proposition in any of the models encouraged students to include that proposition in their texts: Inclusion in a model made that proposition more salient or more plausible than it would have been otherwise. Omitting a proposition from the models left its salience unchanged from the baseline.

As the last column in Table 3 suggests, the quality of the context also had an effect. The biggest differences between the Models group and the control group occur in the A's-only and All-Model contexts, with differences of 15 and 11 percentage points respectively. The size of the difference diminishes by half in the BC-only context (.07), and it disappears in the No-Model context (–.03). This decline is reflected in a significant context-by-presentation interaction [$F(3,273) = 6.7$, $p < .01$]; the contrast between the mean difference for the A's-only context (.15) and the BC-only context (.07) was significant [$t(273) = 2.1$, $p < .05$]. This suggests that the Models group preferred to import propositions from the A's-only context as compared to the B and C models. These students may have recognized the quality of the models and discounted the importance of propositions included only in lower quality models.

Finally, the overall tendency for models to increase the inclusion of propositions was affected by the assignment of topic (Washing or Hierarchy), with a significant topic-by-presentation interaction, $F(1, 91) = 9.1$, $p < .01$. For the Models group, the average likelihood that a proposition would be included was .43, and this probability was the same for both topics. For the control group, the overall probability for the Washing topic is the same as these, .42, but for the Hierarchy topic, it drops to .30. This result is consistent with findings to be presented below that the Hierarchy topic was more difficult than the Washing topic and that the control group found less to say on this topic. These results also suggest that seeing the models compensated for the greater difficulty of the Hierarchy topic.

Effects of Quality and Labeling. Table 4 breaks out the inclusion probabilities for each of the four Models groups. Varying the quality of the models (AAA or ABC) in effect varied the students' opportunities to see propositions in the different contexts.

Students in all four Models groups had a good opportunity to see propositions that appeared in the All-Models context—for students in the AAA group, these propositions appeared in at least 2 of the A models; for students in the ABC group, they appeared in 1 A model and in at least 1 of the B or C models. Not surprisingly, the inclusion probabilities for the All-Models context are the same across the four Models groups.

Propositions in the A's-only context were much more likely to be noticed by students in the AAA group than in the ABC group. The AAA group saw all 3 A models, and these propositions each appeared in at least 2 of them. The ABC group saw these propositions in at most 1 model—their A model. As might be expected, then, students in the AAA group were significantly more likely to include propositions from the A's-only context than were students in the ABC group [AAA = .64, ABC = .51, $F(1, 86) = 7.8, p < .01$]. In fact, only the probability for the

Table 4

Inclusion Probability (SDs) for Propositions in Different Contexts in Texts Produced in Four Model Conditions

	MODELS			
	Labeled		Unlabeled	
	AAA	ABC	AAA	ABC
CONTEXT				
All Models	.59 (.22)	.57 (.23)	.60 (.18)	.52 (.20)
A Models Only	.68 (.22)	.50 (.26)	.59 (.23)	.52 (.25)
BC Models Only	.36 (.12)	.41 (.09)	.33 (.09)	.42 (.14)
No Models	.23 (.12)	.24 (.09)	.18 (.05)	.18 (.09)

AAA group was significantly higher than the probability for the control group [AAA = .64, Control = .42, $t(57) = 3.8$, $p < .01$]. For the ABC group, the probabilities are in this direction but the difference did not reach significance [ABC = .51, Control = .42, $p = .12$).

Conversely, students in the ABC group were the only ones who saw the propositions in the B and C models. The ABC group was significantly more likely to include these propositions than the AAA group (ABC = .41, AAA = .35, $F(1, 86) = 7.6$, $p < .01$]. The ABC group was also likelier to include these propositions than the control group [ABC = .41, Control = .32, $t(56) = 2.6$, $p < .05$], while there was no difference in the probabilities for the AAA group (.35) and the control group (.32).

Because by definition the models did not contain any propositions in the No-Models context, we might expect these propositions to remain at baseline salience and to appear equally often in the texts of the AAA group as the ABC group. And as expected, we found no differences between these groups. However, students who saw unlabeled models were *less* likely to include these propositions than students who saw labeled models [Unlabeled = .18, Labeled = .24, $F(1, 86) = 6.4$, $p < .05$] and less likely than the control group [Unlabeled = .18, Control = .24, $t(56) = 2.1$, $p < .05$]. The students who saw unlabeled models may have judged that if a proposition had not appeared in any of the models, it was probably safest to leave it out. It is not clear why students who saw labeled models did not react the same way.

For these analyses we found no other systematic effects of labeling the models. Labels might have been expected to have some effect on the inclusion of propositions in the BC-only context. If students were taking the quality of the models into account, then one might expect labels to accentuate the effect by making their task of choosing from better models easier. However, students in the ABC group were just as likely to include propositions from the B and C models when the models were labeled as when they were unlabeled. In other words, seeing that the models had lower grades did not deter students in the ABC group from importing some of the propositions they contained. As noted previously, however, students preferred to import propositions from the A's Only and All-Models contexts. These results suggest that students were able to judge the relative quality of the models without the labels.

Summary of Context Analysis

The results so far suggest that seeing a proposition in a model increased the chances that students would include it in their texts over the baseline probability set by the control group. In contrast, propositions that were omitted from the models were equally likely to show up in the texts of the Models group as of the control group. This suggests that the effect of models is to heighten the salience or plausibility of the propositions they contain and not simply to encourage an all-out knowledge dump by students of every piece of information that comes to mind.

The increased likelihood of inclusion (as compared to the control group) depended in part on which models contained the propositions. Students more often included propositions that had appeared in A-only models or that had appeared in A's as well as lower quality models. Students were less likely to include propositions that appeared only in the B and C models. One interpretation of these results is that students selectively chose propositions that they saw in good models, even if these also showed up in the lower quality models. They may have decided that information that is really critical is likely to show up everywhere, across variations in A models and across quality levels.

Another possibility is that the models may have given students greater exposure to the propositions. Repeated appearances of a proposition in the A models, or across the A, B, and C models, may simply have raised the chances that students noticed the proposition or considered it plausible. In this interpretation, even propositions in the B and C models are stronger candidates for inclusion: The effect is smaller for these propositions simply because they are repeated less often and because fewer students had the opportunity to see them. To pursue the possibility that the source of these effects is simply exposure (how often they were repeated in the models), we recategorized the propositions by how many models they appeared in (from 1 to 5 models) and repeated the analysis. The results are presented in Table 5.

These data reproduce the general finding reported above of significantly higher inclusion of propositions in the Models group than the control group [$F(1, 91) = 7.1, p < .01$]. Seeing a proposition in the models encouraged students to include that proposition in their texts. The likelihood of inclusion significantly increased as the number of repetitions increased [$F(4, 364) = 454.6, p < .01$]. More importantly, however, the increases at each repetition were the same for the control group as for the Models group; no repetition-by-presentation interaction was found. This means that the increase in inclusion cannot be explained by simple exposure. Regardless of how often the propositions were repeated across the models, the control group was exposed to them only in the videotape and the factsheet. Despite the fact that the control group had such low exposure to the propositions, they were still likely to include the propositions that had been repeated more often than others in the models.

The effect of repetition shown in Table 5 probably emerged because the most obvious facts were likely to show up in several models, but stray

Table 5

Inclusion Probability (SDs) for Propositions Repeated in Different Numbers of Models

	NO MODELS	MODELS	*Difference*
REPETITION			
One Model	.14 (.10)	.17 (.11)	*.03*
Two Models	.28 (.12)	.32 (.13)	*.04*
Three Models	.52 (.16)	.59 (.15)	*.07*
Four Models	.52 (.17)	.61 (.20)	*.09*
Five Models	.78 (.14)	.88 (.16)	*.10*
Marginals	*.44*	*.51*	

facts appeared in only one or two models. Each writer may include a few odd extraneous facts, but these inclusions generally vary across populations. Similarly, each writer may forget a random, important fact. Because we did not systematically manipulate the placement of propositions of different types in the models, repetition in our materials is confounded with plausibility. Propositions appearing in different contexts had different baseline saliences. For example, as shown in Table 3, the average baseline probability in the No-Models context (.24) was much lower than in the all-models context (.46). While all categories included a mixture of relevant and irrelevant propositions, some categories clearly ended up with more relevant information. For these reasons, the differences between the control and the Models groups are more meaningful than the absolute sizes of the inclusion probabilities. These differences support the hypothesis that students were sensitive to the context in which a proposition appeared—that they included certain propositions because they had appeared in higher quality models. Differences in baseline salience and repetition may yet play some role in the effects reported here. More research is needed to shed light on these alternatives.

Inclusion of Relevant Propositions

The context analysis suggests that students who see models are more likely to include propositions from those models in their texts. The purpose of the next analysis is to investigate what kinds of propositions students include. Are students who see models more successful at including relevant, and omitting irrelevant, information than the control group? In contrast to the previous analyses, the absolute scores in these analyses are meaningful. Ideally, all students will include a high proportion of the essential information and low proportions of the unnecessary and extraneous information. We will again begin by comparing the control group to the Models group overall and will then discuss the effects of quality and labeling within the models group.

Overall Comparisons of the Models and No-Models Groups. Table 6 presents the average proportion of essential, unnecessary, and extraneous propositions included in the texts of students in the No-Models and Models groups. The results for the Washing and Hierarchy topics are presented separately because topic interacted with most of these measures.[4] We therefore conducted separate ANOVAs for each topic.

Table 6

Proportion of Propositions (SDs) by Relevance Category and Topic in Student Papers Produced With and Without Models

	NO MODELS	MODELS
RELEVANCE		
Essential		
Washing	.67	.60
	(.08)	(.14)
Hierarchy	.45	.56
	(.14)	(.11)
Unnecessary		
Washing	.28	.31
	(.12)	(.14)
Hierarchy	.19	.36
	(.13)	(.15)
Extraneous		
Washing	.11	.11
	(.05)	(.06)
Hierarchy	.11	.15
	(.06)	(.10)
Total Words		
Washing	334.6	306.4
	(52.3)	(79.2)
Hierarchy	304.5	350.2
	(68.4)	(90.3)

For the Washing topic, the control and Models groups included the same proportions of all three kinds of propositions. For the Hierarchy topic, however, the Models group included significantly more essential and more unnecessary propositions than the control group, [essential propositions, $F(1, 80) = 10.3$, $p < .01$; unnecessary propositions, $F(1, 80) = 10.9$, $p < .01$]. The apparent increases for extraneous propositions and total words for the models group did not reach significance ($p < .20$ for both analyses). Overall, these results suggest that models encouraged students to include more topical information, at least for the Hierarchy topic. Propositions in the unnecessary category were likely to be at the wrong level of detail but were at least judged as topical for the section. But the models apparently did not help students discriminate between the essential and unnecessary details.

Why did the effect of models emerge only for the Hierarchy topic? As suggested previously, the Washing topic appears to have been an easier topic than the Hierarchy topic. Table 6 indicates that, even without models, students were able to include almost two-thirds of the essential propositions for the Washing topic. But students apparently had more difficulty describing the Hierarchy experiment. The control group included less than half of the essential propositions, a significant drop from their counterparts writing on the Washing topic [$t(20) = 4.2$, $p < .01$]. Having models available seems to have compensated for the difficulty of the Hierarchy topic: The Models group's performance did not decline. The apparent decrease in the Models group's essential propositions from .60 on the Washing topic to .56 on the Hierarchy topic is not statistically significant. In short, students in the Models group were at least as fluent on the more difficult Hierarchy topic as on the Washing topic, including the same proportions of essential propositions and unnecessary propositions. In contrast, the amount of information that the control group included dropped off for the Hierarchy topic.

Effects of Quality and Labeling. The quality of the models, whether students saw 3 A models or an A, a B, and a C, did not influence how much relevant or irrelevant information they included. The labeling of the models also failed to produce systematic effects. Students who saw labeled models used more total words than those who saw unlabeled models [Labeled, 344.3, Unlabeled, 312.2, $F(1, 80) = 4.4$, $p < .05$].

Proficiency in Psychology. The students' proficiency in psychology had significant effects on their ability to select essential information. Across both topics and all conditions, students with higher course grades included significantly more essential propositions than did students with lower grades [high = .61; low = .54, $F(1, 80) = 9.2$, $p < .01$]. However, students with high grades included about the same proportions of unnecessary and extraneous propositions as students with low grades. Students with high and low grades also used about the same number of words.

Effects of Topic and Writing Situation. Several aspects of the results so far have indicated that the usefulness of a model depends on the difficulty of the topic and the circumstances of the writing situation. These implications emerge again if we consider the general level of performance of students in this study in relation to the standard set by the models. Comparing Tables 2 and 6 reveals that the average student's text was roughly as good as one of the A models, even in the No-Models control group. Each A model includes about 65% of the essential propositions. For the Washing topic, the average student in our study included about the same proportion—a surprisingly high level of performance. But students writing on the Hierarchy topic included far fewer essentials than any one of the A models. The control group included an average of 45% of the essentials, about the same proportion as the B and C models. The models group included 56% of the essentials, placing them between the A and B models. Note also that for both topics, the control group successfully avoided unnecessary and extraneous propositions—their average proportions for both topics are comparable to an A model. Although the models group included more unnecessary propositions than the control group, their average proportions also remain in the A range. These comparisons provide more evidence that the Hierarchy topic was harder than the Washing topic and that the Washing topic was too easy.

It is possible the A models themselves did not set a sufficiently high standard—including too few essentials and too many intrusions. It is certainly conceivable that the A models legitimized the inclusion of the unnecessary and extraneous propositions they contained, as well as the omission of some essential facts. An interesting question for future research is whether we would have achieved greater benefits from a less naturalistic "super-model" that included all the essentials and no unnecessary or extraneous propositions. However, it is important to keep in mind that the models we used represented the range of performance of students writing Method sections in previous semesters, where the average grade was certainly not an A. As a set, the three A models contained nearly .90 of the essential propositions—a ceiling well above the performance of any students, even those who saw all three A models labeled with their grades.

It is more likely that the conditions under which students wrote their texts made the task easier than normal. In particular, the videotape, experimental materials, and fact sheets may have given students everything they needed to perform well. In fact, the opportunity to select items from a fact sheet that contained blatantly irrelevant details may have alerted students to be more vigilant in excluding irrelevancies. (Students in both groups frequently checked off or crossed out items on their fact sheets.) The Washing topic was sufficiently straightforward that, under these conditions, the average student could describe it adequately. In contrast, the Hierarchy topic was more complex, particularly in the description of the materials and what was done with them. The fact sheet and materials were not sufficient to help students describe the study fully and accurately. Many students were unsure what "taxonomic hierarchies" meant and attempted to describe the tree-structures in convoluted ways. Further, they frequently confused the two tree structures (plants and minerals) with the two recall trials, assuming incorrectly that only one tree structure was used for each trial. Many students who found these points confusing glossed over them and left out details; others who attempted to include them made mistakes. Students in the Models group were more successful at including essential propositions on this topic—perhaps because seeing the essential details expressed in the models reinforced their salience or perhaps because the models helped students find ways to express the details accurately.

Conclusions

This study investigated whether and how seeing textual models affects students' efforts to write in a genre. The study sought to determine whether students who saw models produced texts that systematically differed—for better or for worse—from those written by students who worked without models. It also sought to investigate whether the quality of the models and students' knowledge of their judged quality influenced how much students would rely on the models. Would students benefit more from seeing several high quality models or from seeing a range of quality levels? Would labeling the models lead students to follow "good" models uncritically? Taken as a whole, the results indicate that models do not have automatic benefits for the writing process. Likewise, they begin to reveal why models have produced mixed results in previous studies and in classroom practice. But the results also point to some hypotheses about what aspects of a writer's performance may be influenced by models.

The results presented here suggest that models do influence the content and organization of students' texts. Reading models seems to have reminded writers of concepts that they otherwise would not have included in their texts. Seeing a related or analogous concept in a model may increase the salience or "activation level" of associated concepts in the writer's memory. When writers use models only as a source of memory retrieval cues, however, they may also retrieve items that are associatively related to concepts in the model but that are irrelevant to the task at hand. We saw evidence of this in the increased number of unnecessary propositions included in the models; we might also expect to see idiosyncratic intrusions for individual students. Similarly, reading models may also have reminded students of genre-specific structural patterns at the sentence level, paragraph level, or higher discourse levels, thereby increasing the likelihood that the writer used these patterns while writing. Kucer (1986) reviews research that reading influences a writer's selection of syntax and other text structures, even when the goal is not explicitly to imitate the text that was read. Our finding that the Models groups' texts rated higher for organization is consistent with this hypothesis, but more fine-grained analysis of students' texts is needed.

While models influenced students' selection and arrangement of information, varying the quality and labeling the models did not produce systematic effects. The context analysis indicated that students would rather import propositions that they had seen in the A models than use those that appeared in the B and C models only. If students had adopted a deliberate strategy of choosing propositions from the best models, then we might expect that labeling the models with grades would help them include more information from the good models than from the poor ones. However, we found no effects for labeling. In particular, students in the Models group were just as likely to include propositions from the B and C models when the models were labeled as when they were unlabeled. Labeling the models also did not help students distinguish relevant from irrelevant information. That is, we found no systematic evidence that students who saw labeled models included more essential information or less extraneous information than students with unlabeled models. This may simply mean that students did not need the labels—that they were able to judge for themselves the relative quality of the models.

We also found no overall advantage to providing three good models as compared to models representing good, moderate, and poor quality levels. We had hypothesized that seeing three good models might help students derive a fuller range of acceptable variation than they could from any one good model. In contrast, the opportunity to compare good, moderate, and poor models might help students identify the effective aspects of the models and avoid the mistakes. Our results may mean, in essence, that both combinations of models produce the same effects. Students in the AAA group may have inferred approximately what it took to perform at the A level, generalizing from their three models. And students in the ABC group may have done equally well by following their

A model and comparing it to the B and C models. However, the possible advantages of one combination of models over the other may have been obscured in this study by the nature of the task. Students did not have much difficulty excluding most of the unnecessary and extraneous information presented to them, even in the control group.

The lack of systematic differences between the AAA group and the ABC group does not mean that students ignored the quality of the models. As we have seen, students were less likely to include propositions that appeared only in the B and C models than to include those that appeared in the A models. A simple exposure hypothesis is inconsistent with these results. An exposure hypothesis would assume that students simply imported propositions that they saw repeated most frequently in the models. Repeated appearances of a proposition in the models might provide extra retrieval cues or strengthen the salience or plausibility of the proposition. This interpretation is inconsistent with the results in Table 5, showing that the increased likelihood to include propositions *over the baseline established by the control group* was no greater for propositions repeated in 5 models than those appearing in only 1 model. An exposure hypothesis would also lead us to expect the AAA group to include more essential propositions than the ABC group, because the three A models included a higher proportion of the essentials. However, these groups included exactly the same proportions of essential propositions [Essential, AAA = .58 (.11); ABC = .58 (.14)]. Similarly, the ABC group should have included more unnecessary and extraneous propositions than the AAA group because the B and C models contained much higher proportions of these than did the A's. Again we found no such effects [Unnecessary, AAA = .33 (.15); ABC = .34 (.13). Extraneous, AAA = .12 (.07); ABC = .13 (.09)]. Students did not appear to follow these models slavishly: They seemed to take into account both the quality of the model and the nature of the propositions.

This study provided no specific instruction to students on how to use a model. We hoped to discover how much students could achieve on their own using various kinds of models. We hoped as well to identify the kinds of problems that might require additional intervention. Choosing models to provide to students becomes an easier task if students can indeed judge models without external cues to their quality and can avoid extraneous information in the models. Instructors who wish to use student texts may feel more comfortable using texts that illustrate some aspects of the genre adequately, but that are flawed in other respects. Further research is needed to explore how well students can judge model texts, including research varying the range of quality levels more systematically and varying the closeness of the models to the task situation.

This study does not provide any direct evidence about how the students reacted to the models while reading them or how they used them while writing. A particularly interesting question is whether students who had the unlabeled models invested greater effort during reading to working out their relative quality. Studies of the role of examples in learning other cognitive skills has shown that the major benefit of examples accrues to students who invest more time in analyzing them (Chi, et al., 1989; Recker & Pirolli, 1990). So it is possible that how students reason about and apply the models is more important than the quality of the models themselves. We speculate that active analysis of models may help students in two ways.

First, active analysis of a model (even before taking on a specific writing task) may help students construct new textual patterns or enrich the patterns they know. While style guides and publication manuals can convey the bare-bones outline of a genre, such tools cannot capture all the complexities of selection, expression, and arrangement—in part because these are less fixed, varying subtly in different sub-specialties or even among journals. Writers may be able to infer such features from models, but this task is difficult because, as Hillocks (1986) has pointed out, the models are so fully elaborated that the schemata underlying them may be obscured. Presumably, writers who actively look for and contemplate such features in the models they read are more likely to construct reliable new structures.

Second, consulting models actively during the writing process may provide the writer with a database for testing whether a candidate idea should be included. The writer may infer that the practice of the writer who produced the model is typical and may include or exclude information on the basis of whether or not it shows up in the models. Such active relevance testing may be the only way to prevent writers from including spuriously salient concepts—such as when a student is determined to give prominent space to an aspect of a project that happened to cost excessive time and energy. Similarly, the writer may test the expression of an idea by looking for a similar idea in the model and imitating the phrasing of the model. In this case, the student may understand the concept but may not know how to convey it as succinctly as more experienced researchers. For example, parts of the Method section have stylized phrasing for complex concepts, such as the description of the experimental design. So imitating a model may help students use more conventional language for talking about variables and levels and measures.

To explore these possibilities, the next step in developing a pedagogical theory of modeling should be to use process-tracing methodologies (such as think-aloud protocols) to investigate how students read models and how they use them during writing. Stolarek's (1994) study of how faculty and students use models is a step in this direction. In this study, participants were asked to write in a simple invented genre that combined description and narration. They were provided with various combinations of materials for learning the genre: a set of guidelines for content and style, a model, and an explication of how the model fit the requirements. Students wrote more successful products and reported using more active and evaluative writing strategies when given models in conjunction with some other instruction than when given models alone or guidelines alone. Stolarek's results seem generally consistent with those reported here. Because students in our study had all previously received instruction in the components of a Method section, none of our conditions duplicates her Model Only condition. Our Models group is somewhat like her combined model with guidelines group.

Additional fine-grained study of students' activities while reading and writing is needed, however. It is possible that students who adopt active strategies for analyzing and consulting models on their own may be successful even without interventions. For students who lack such skills, the appropriate intervention may be instruction in such skills, using pedagogies that demonstrate on-line reading activities and provide practice in using them (Greene, 1993; Haas, 1993). An alternative pedagogical strategy would be to annotate models with critical commentary or to provide explications of a model. Stolarek's study did not report distinct advantages to providing an explication of the model as compared to the more general guidelines; students in her study needed at least one of these to make sense of the model, but having both did not seem to improve performance. One drawback of explications or critical annotations is that they can take on unwarranted prescriptive authority when a wider range of options may be acceptable. Because our students clearly did not take full advantage of the models (on average leaving out about one-third of the propositions considered essential), some additional intervention is warranted.

By their nature, genres create tension between the new and the old, between the original and the conventional, between what needs to be explained and what can be taken for granted. But new and old can only be distinguished in relation to a community and a time period. Determining what is really essential or relevant to include in an experimental article depends on close understanding of conventional practice in the discipline, a sub-specialty, or even a particular journal. At any given time, some methods and concepts will have the status of standard practice; in a text, they may require mention but not much elaboration. Other practices may remain somewhat exotic, requiring more discussion. And any new research study introduces some novelty that requires explanation and justification. For students, or anyone new to a discourse community, all the concepts and practices may be equally new and unfamiliar. Mistaken inclusions of unnecessary detail, omissions, and use of foreign phrasing are probably inevitable in a writer's initial attempts to fit the genre. Model texts are a rich resource that may prove useful to writers in different ways at different stages of their development. For student writers, models may be effective tools for learning the more enduring conventional forms or for understanding those that apply most broadly across the discipline. At a later stage, models (especially those in professional journals) may provide valuable clues to the status of knowledge in the field. It seems likely that early experience in evaluating and drawing from models will be of lasting value.

Authors' Note: Support for this research comes from a Pennsylvania State University Cognitive Studies grant and from a grant from the Society for Technical Communication. Earlier versions of this paper were presented at the Annual Meeting of the American Educational Research Association and the Conference on College Composition and Communication. We are very grateful to Lori Forlizzi and Karen Bartsch for their substantial conceptual and practical contributions and to Ian Nimmo-Smith for timely and helpful statistical advice.

Notes

1. Accounting for each sentence or clause in the students' texts proved an unproductive approach due to the wide variation in writing styles among students. Irrelevant details were sometimes stated in independent clauses ("The paper products were purchased at K-Mart") and sometimes embedded in the same clause with essential details ("The plant and mineral hierarchies were printed on four separate white sheets of paper from K-Mart"). If clauses or sentences were to be scored as relevant or irrelevant units, students with looser syntactic styles might have been judged as conveying more information, while expressing the same basic ideas as students with tighter styles.

2. We used a relatively strict principle for creating these categories. A proposition was counted as present in the A's only if its average score over the three A models was at least 0.50. For example, propositions that were scored at full credit (1.0) in at least two A models were counted as present in the A's, as were propositions with partial credit (.50) in all three models. Similarly, we only counted propositions as present in the B and C models if the average score over those 2 models was .50 or more.

3. For this analysis, we used a mixed design. We treated context category as a within-subjects repeated-measure factor with four levels (All, A's-Only, BC-Only, and None). The between-subjects presentation factor was the same as that described above. After finding that context interacted with presentation and with topic, we ran individual ANOVAs on each context category.

4. Topic-by-presentation interactions were significant for essential propositions [$F(1, 80) = 8.8, p < .01$]; unnecessary propositions $F(1, 80) = 3.7, p = .06$]; and total words, [$F(1, 80) = 3.7, p = .06$]. The pattern for extraneous propositions was in the same direction, $p < .20$.

REFERENCES

Bazerman, C. (1989). *Shaping written knowledge: The genre and activity of the experimental article in science.* Madison, WI: University of Wisconsin Press.

Bohrnstedt, G. W. (1983). Measurement. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 69–121). San Diego, CA: Academic Press.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior, 8,* 323–343.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726.

Chi, M., Bassok, M., Lewis. M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Faigley, L., & Hansen, K. (1985). Learning to write in the social sciences. *College Composition and Communication, 36,* 140–149.

Greene, S. (1993). Exploring the relationship between authorship and reading. In A. Penrose & B. Sitko (Eds.), *Hearing ourselves think: Cognitive research in the college writing classroom* (pp. 33–51). New York: Oxford University Press.

Haas, C. (1993). Beyond "Just the facts:" Reading as rhetorical action. In A. Penrose & B. Sitko (Eds.), *Hearing Ourselves Think: Cognitive Research in the College Writing Classroom* (pp. 19–32). New York: Oxford University Press.

Herrington, A. (1992). Composing one's self in a discipline: Students' and teachers' negotiations. In M. Secor & D. Charney (Eds.), *Constructing rhetorical education* (pp. 91–115). Carbondale, IL: Southern Illinois University Press.

Hillocks, G. (1986). *Research on written composition: New directions for teaching.* Urbana, IL: National Council of Teachers of English.

Kucer, S. (1987), The cognitive base of reading and writing. In J. Squire (Ed.), *Dynamics of Language and Learning* (pp. 27–51). Urbana, IL: ERIC Clearinghouse on Communication Skills and National Conference on Research in English.

Law, J., & Williams, R. J. (1982). Putting facts together: A study of scientific persuasion. *Social Studies of Science, 12,* 535–58.

Miller C. (1984). Genre as social action. *Quarterly Journal of Speech, 70,* 151–167.

Nitsch, K. E., (1977). Structuring decontextualized forms of knowledge. *Dissertation Abstracts International, 38,* 08B. (University Microfilms No. 77-30, 369)

Pirolli, P. (1991). Effects of examples and their explanations in a lesson on recursion: A production system analysis. *Cognition and Instruction, 8,* 207–259.

Recker, M., & Pirolli, P. (1990, April) *A model of self-explanation strategies of instructional texts and examples in the acquisition of programming skills.* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 318 797)

Rymer, J. (1988). Scientific composing processes: How eminent scientists write journal articles. In D. Jolliffe (Ed.), *Advances in writing research: Vol. 2* (pp. 211-250). Norwood, NJ: Ablex.

Smagorinsky, P. (1992). How reading model essays affects writers. In J. W. Irwin & M. Doyle (Eds.), *Reading/writing connections: Learning from research.* Newark, DE: International Reading Association.

Smagorinsky, P. (In press.) Models. In A. Purves, (Ed.), *Encyclopedia of the english studies and language arts.* Urbana, IL: National Council of Teachers of English.
Stolarek, E. (1994). Prose modeling and metacognition: The effect of modeling on developing a metacognitive stance toward writing. *Research in the Teaching of English, 28,* 154–174.
Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2,* 59–89.
Tennyson, R. (1973). Effect of negative instances in concept acquisition using a verbal learning task. *Journal of Experimental Psychology, 64,* 247–260.
Tennyson, R., Woolley, F. & Merrill, M. (1972). Exemplar and nonexemplar variables which produce correct concept classification behavior and specified classification errors. *Journal of Experimental Psychology, 63,* 144–152.
Thompson D. (1993). Arguing for experimental facts in science: A study of research article results sections in biochemistry, *Written Communication, 10,* 106–130.
Werner, W. (1989). Models and the teaching of technical writing. *Journal of Technical Writing and Communication, 19*(1), 69–81.
Winsor, D. (1990a). How companies affect the writing of young engineers: Two case studies. *IEEE Transactions on Professional Communication, 33,* 124–129.
Winsor, D. (1990b). Joining the engineering community: How do novices learn to write like engineers? *Technical Communication, 37,* 171–172.

APPENDIX A: Fact Sheet for Hierarchy Topic

Hypothesis 1: Organized condition recall > randomized condition recall.
Subjects = 50 native English-speaking students
Subject pool: Introductory Psychology classes, PSU
Hypothesis 2: 2nd trial recall > 1st trail recall.
Subjects' Psychology professors: Dr. Smith, Dr. Schwartz.
Hypothesis 3: No interaction between factors.
Materials: 54 nouns in 2 taxonomic hierarchies
Source of paper materials: K-Mart, State College, PA.
2 hierarchies: minerals (26 words) and plants (28 words)
Consent forms: from Office for Protection of Human Subjects
Experiment day and time: Wednesday, 7–9 p.m.
Mixed factorial design, 2 × 2
  Factor 1: organized vs. random word presentation, between subjects
  Factor 2: trial 1 vs. trial 2, within subjects
Randomized assignment to condition
Experimental design: devised by experimenters
25 subjects assigned to each presentation condition
Means: Trial 1/Organized: 36 words
  Trial 1/Randomized: 20 words
  Trial 2/Organized: 49 words
  Trial 2/Randomized: 31 words
Results: Main effects of both factors
ANOVA interaction not significant: $F(1, 46) = 0.94$, p. > 05.
Subjects' residences: Pennsylvania, Ohio, New York, New Jersey

Recall period = 3 minutes per trial
Recall score = # of words recalled for each trial
Study period = 60 seconds per sheet
Bartlett's (1932) experiment
Ebbinghaus materials = nonsense syllables

APPENDIX B: Good, Intermediate, and Poor Models

Model A1, Hierarchy topic.

Method

Subjects

The subjects were fifty student volunteers from an undergraduate introductory psychology class.

Materials

The materials included fifty-four common, concrete nouns. The words were organized into two taxonomic hierarchies, minerals and plants. Each hierarchy had four levels. Four stimulus sheets were made up. Two of these sheets contained the heierarchiees in an organized tree structure. The other two sheets presented the words in the same format, but the words were randomly assigned to places in the two hierarchies. Blank paper and pencils were distributed to the subjects.

Design

A 2 × 2 mixed factorial design was used with a within-subjects variable being the relationship between the 2nd trial recall and the 1st trial recall. The other independent variable was a between-subjects variable being the organized/randomized word lists. Half of the subjects were assigned to each condition. The first group received the organized hierarchies (N = 25), while the second group (N = 25) received the randomized hierarchies. The dependent variable was the number of correctly recalled words. Only the results from correct responses were used. Synonyms were not counted. Misspellings were counted as correct but repeated words were only counted one time.

Procedure

The subjects were randomly assigned to one of the two conditions of the experiment and they were assigned to different parts of the room. The subjects were verbally instructed to study the two lists of words. The subjects were told that they would be tested on their ability to recall the words from both lists. At the start of the first session, the subjects were told to "study" the first list of words. The subjects were given sixty seconds to study the first list of words. After the sixty seconds were up, the instructor told the subjects to "stop" studying the first list. Then after a short wait, the subjects were instructed to study the second list of words for another sixty seconds. At the end of the second sixty second period, the instructor told the subjects to "stop" and to turn the sheet over. The subjects were then instructed by the experimenter to "recall" as many of the words on the two lists. The subjects were given three minutes to write down as many words as the could remember. After the recall period, the subjects were instructed to repeat the same task.

Model A1, Washing Topic

Method

Subjects

The subjects were fifty student volunteers from an undergraduate introductory psychology class.

Materials

The materials include two paragraphs that contained 250 words. The paragraphs were identical except for a title. Blank paper and pencils were also distributed to the subjects.

Design

A 2 × 2 mixed factorial design was used with a within-subjects variable being the relationship between the 2nd trial recall and the 1st trial recall. The other independent variable was a between subjects variable being the titled/untitled paragraphs. Half of the subjects were assigned to each condition. The first group received the titled paragraph (N=25) while the second group (N=25) received the untitled paragraph. The dependent variable was the number of correct idea units recalled. Only the results from correct responses were used.

Procedure

The subjects were randomly assigned to one of the two conditions of the experiment and they were assigned to different parts of the room. The subjects were verbally instructed to study the passage. They would then be tested on their ability to recall the passage. At the start of the first session, the subjects were told to "study" the passage. The subjects were given three minutes to study the passage. At the end of three minutes, the experimenter told the subjects to "stop" and to turn the paragraph over. At the end of the study period, the subjects were instructed to "recall" as much of the paragraph as they could. The subjects were given three minutes to write down as much information as they could remember. After the recall period, the subjects were instructed to repeat the same task.

Model A2, Hierarchy Topic

Method

Subjects

Fifty Pennsylvania State University undergraduates who enrolled in an introductory psychology course participated in this experiment. They were randomly assigned to either the organized condition or the randomized condition, with 25 subjects in the organized condition and 25 subjects in the randomized condition.

Materials

Subjects were given two pieces of blank paper, and one out of the two sets of fifty-four nouns. The words were divided into two taxonomic hierarchies. The first taxonomic group contained 26 words that referred to minerals and the second group contained 28 words that included plants. Four sheets were prepared for the experiment. Two of the sheets were in an organized fashion. The other two sheets had the words arranged in randomized structure. For the experiment, half of the subjects studied the organized hierarchies, the other half studied the randomized hierarchies.

Design

The experiment called for a 2 × 2 mixed factorial design. The independent variables were organized vs randomized trial. The organized vs. randomized was tested between subjects and trial was tested within subjects. The dependent variable was the number of correct works recalled for each trial. Only correct responses were included in the score. The number of trials and time for study and recall were kept constant from subject to subject.

Procedure

Subjects were divided into two equal groups. Half of the subjects were assigned to the organized conditions, the other half were assigned to the randomized condition. To begin, the subjects were instructed that they would be tested on their ability to recall the words on the two sheets of paper. The subjects were instructed to study the first list of words for sixty seconds. The instructor ended the first study session by saying "stop." The subjects were then given a short break before they were asked to study the second study session by saying "stop." The subjects were then given three minutes to recall as many of the words from both lists as possible. After the first recall period, the instructor repeated the experiment for a second time.

Model A3, Hierarchy Topic

Method

Subjects

Fifty volunteers from an introductory psychology class participated as part of the experiment. Subjects were randomly assigned to one of two conditions, twenty-five in the organized condition and twenty-five in the randomized condition.

Materials

Blank paper, pencils, and two sets of typewritten words were all used in the experiment. The stimulus sets consisted of fifty-four common nouns divided into two taxonomic hierarchies, minerals, and plants. The two sets of fifty-four nouns were either organized or randomized.

Design

A 2 × 2 mixed factorial design was employed. There were two independent variables, Organized/Randomized and Trial. Organized, Randomized was a between subject manipulation and consisted of two conditions, organized (where the words were placed in an organized tree structure) and randomized (where the same words were presented in a tree structure but the words were randomly assigned to positions in the two hierarchies). The trial was a within subjects manipulations and consisted of a second recall of the list of words. The dependent variable was the correct number of words recalled. Several variable were controlled. These included the length of time subjects had to study and recall the two sets of words and each subject was not allowed to see the list of words until the experimenter told them to do so.

Procedure

Subjects were told that they were participating in an experiment in exchange for extra credit. They were given instructions on the experiment and the various steps to follow during the experiment. The subjects were randomly assigned to one of two groups, the organized condition or the randomized condition. Each subject was given his/her own blank paper, pencil, and either the organized or randomized list of words. After receiving the list of words face down, the subjects were instructed to turn the paper over and begin to study the first list of words. The subjects were given sixty seconds to study the first list of words. After the sixty seconds were up the subjects were told to stop studying the list. After a short time, the subjects were then instructed to study the second list of words.

They were given another sixty seconds to study the second list of words. After the sixty seconds were up the subjects were to stop studying the words and to turn the sheets face down. The subjects were then told to begin the recall phase of the experiment. The subjects had three minutes to recall the words on both lists. They were instructed to recall as many words on both lists as they could remember. After the first recall period was over, the subjects repeated the procedure again. Data was collected for all subjects. Finally, all subjects were debriefed on the intent of the experiment.

Model B, Hierarchy Topic

Method

The subjects of this experiment were 50 students attending The Pennsylvania State University. They participated in the experiment in return for extra credit. They were split into two groups by random selection. One, designated the organized group, consisted of 25 students. They received two lists of words, the first had a list of 26 minerals, the second had a list of 28 plants. The words were organized into a hierarchy that had four levels. The randomized group contained 25 subjects and the list of words were the same except that they were randomly assigned to positions in the two hierarchies.

For this experiment, three materials were used: several pieces of blank paper, pencils, and two sets of word lists. The word lists were either organized or randomized.

The experiment used a 2 × 2 factorial design. All subjects were given the list of words to study. Out of the 50 subjects, 25 were given the organized list of words only and 25 were given the randomized list of words only. In order to eliminate confounding variables, the subjects were given the list of words face down. Also subjects were told to study the list of words only when the instructor told them to do so. After randomly assigning the subjects to the two groups, they were then sent to different parts of the room. The subjects in the organized condition were directed to sit on the right side of the room and the subjects in the randomized condition were directed to sit on the left side of the room. The experimenter gave instructions for the experiment. The experimenter told the subjects that they were to study both lists of words for sixty seconds. After the subjects had studied both of the lists for sixty seconds each they were to write down as many words as they could remember. The subjects would have three minutes to recall as many words from both lists as possible. Once this had been completed the experiment would be repeated.

Model C, Hierarchy Topic

Method

Subjects: Fifty introductory psychology students volunteered for the experiment. Subjects were randomly assigned to two groups and sent to different parts of the room.

Materials: The experiment included a couple of pieces of blank paper, some pencils, and two sets of stimulus material that contained fifty-four words.

Design: This experiment was a 2 × 2 mixed factorial design. The organized condition was a between-subject manipulation. Group one received the organized condition in which the words were organized into two hierarchies. Group two received the random condition in which the words were randomly assigned to the two hierarchies. All subjects were given the words to read. The subjects studied the first set of words. The instructor told the subjects to study the words for sixty seconds. After the minute was up, the instructor told the subjects to stop. After a short rest, the instructor told the subjects to study the second set of words. After the minute was up, the instructor told the subjects to stop. Then the instructor told the subjects to recall as many of the words as possible. The subjects would have three minutes to recall as many words as possible. This procedure was repeated after a short break.

Procedure: Subjects were given the words face down. Once the instructor had given directions, the subjects were told to study the words. After sixty seconds, the subjects were told to stop studying the first set of words and turn the sheet over. Then, following a short break, the instructor told the subjects to study the second sheet of words. After a minute was up the instructor told the subjects to stop studying the words and to turn the paper face down. The instructor then told the subjects to recall as many of the words as possible. After three minutes were up, the subjects stopped writing. The whole procedure was repeated after a short break. The number of correct words were scored.

Appendix C: Propositions Available for Hierarchy Topic

Sorted by Relevance (Essential, Unnecessary, Extraneous) and Context (1 = All Models, 2 = A's Only, 3 = BC Only, 4 = No Models)

Subjects

ESSENTIAL	1	Subjects are 50 students.
	1	Subjects are Introductory Psychology students.
	1	Subjects are volunteers.
	3	Subjects earned extra credit toward their course grade.
	4	Subjects are native English speakers.
UNNECESSARY	1	Half the subjects were randomly assigned to the organized condition.
	1	Half the subjects were randomly assigned to the random condition.
	3	Subjects are Penn State students.
	3	Subjects were sent to different parts of the room.
	4	Subjects' residences are in Pennsylvania, Ohio, New York, or New Jersey.
EXTRANEOUS	4	Subjects' psychology professors are Dr. Smith or Dr. Schwartz.

Materials

ESSENTIAL	1	Words are arranged in 2 taxonomic hierarchies.
	2	Materials are 54 nouns.
	2	One hierarchy was for minerals.
	2	One hierarchy was for plants.
	2	Four stimulus sheets were prepared.
	2	Two sheets presented the plant and mineral hierarchies in an organized tree structure.

	2	Two sheets presented the words arranged randomly in the tree structure.
	4	Each hierarchy has four levels.
	4	The mineral hierarchy contained 26 words.
	4	The plant hierarchy contained 28 words.
	4	The same tree structure layout was used for organized and random hierarchies.
	4	Words include concrete common nouns.
	4	Words were typed on standard white paper.
UNNECESSARY	1	Materials included blank paper, pencils.
EXTRANEOUS	4	Paper was purchased at K-Mart, State College, Pennsylvania.
	4	Materials include consent forms from Office for Protection of Human Subjects.
	4	Half the subjects studied the organized hierarchies.
	4	Half the subjects studied the randomized hierarchies.

Design

ESSENTIAL	1	The experimental design was a 2 × 2 mixed factorial.
	1	One independent variable was organized versus random presentation of the word lists.
	1	Presentation was manipulated between subjects.
	2	One independent variable was first versus second recall trial.
	2	Trial varied within subjects.
	2	Dependent measure is number of words correctly recalled for each trial.
UNNECESSARY	3	Subjects were randomly assigned to presentation conditions.
	3	25 subjects (half) were assigned to each presentation condition.
EXTRANEOUS	2	Study and recall times were constant for all subjects.
	3	All subjects were given the words to read.
	3	Organized presentation had the words in organized tree structures.
	3	Randomized presentation had the words randomly assigned to the same structures.
	3	The subjects studied the first set of words.
	3	The subjects were told to study the words for sixty seconds.
	3	The subjects were told when to stop.
	3	The subjects were given three minutes for recall.
	3	The subjects were told to recall as many of the words as possible.
	3	This procedure was repeated after a short break.
	4	The subjects were told to study the second set of words.
	4	Experimental design was devised by experimenters.
	4	Misspellings were counted as correct.
	4	Order of recall did not matter.

	4	Repeated words were only counted once.
	4	Synonyms were not counted.

Procedure

ESSENTIAL	1	The first study period began with the command "study."
	1	Study period was 60 seconds per sheet.
	1	The study period ended with the command "stop."
	1	Same procedure was used for second stimulus sheet.
	1	The recall period began with the command "recall."
	1	Recall period was 3 minutes per trial.
	1	The experimenter repeated the procedure for the second trial.
	2	Subjects were told they would be tested on their ability to recall all the words.
	3	Subjects were given a plant and a mineral hierarchy.
	3	Subjects wrote all the words they recalled on a blank sheet of paper.
	4	Subjects were told they would study the words contained on each sheet for 60 sec.
	4	Order of recall did not matter.
	4	The subjects were told to turn over the recall sheet.
	4	Subjects received a written debriefing.
UNNECESSARY	3	Subjects were sent to different parts of the room.
	3	Subjects in the organized condition (group 1) sat on the right.
	3	Subjects in the random condition (group 2) sat on the left.
	3	Stimulus sheets were distributed face down.
	4	Subjects labeled the sheet "trial 1."
	4	Subjects were given blank paper and pencil.
EXTRANEOUS	1	Subjects were randomly assigned to presentation condition.
	3	Recall score was the number of words recalled for each trial.
	4	Each subject read and signed an informed-consent form.
	4	Experiment took place on Wednesday, 7–9 p.m.
	4	Experimenter flipped a coin.
	4	Heads was group 1.
	4	Tails was group 2.
	4	Half were assigned to each condition.
	4	Misspellings were counted as correct.
	4	Repeated words were only counted once.
	4	Synonyms were not counted.
	4	Subjects were told they would receive extra credit for participating.
	4	Subjects were dismissed.