International Journal of Science Education
Vol. 28, Nos 2–3, 15 February 2006, pp. 203–233

Writing Experiment Manuals in Science Education: The impact of writing, genre, and audience

Gert Rijlaarsdam[a,b]*, Michel Couzijn[a], Tanja Janssen[a], Martine Braaksma[a,c] and Marleen Kieft[a]
[a]*University of Amsterdam, The Netherlands;* [b]*Utrecht University, The Netherlands;* [c]*University of California, USA*

In this study, Grade 9 students wrote experiment manuals for their peers describing a simple physics investigation to explore whether air takes space. Peers executed these manuals and their processes were videotaped. In several experimental conditions, these videotapes were played back for authors. Then they had to rewrite the experiment manual. Three weeks later they wrote a letter-of-advice, explaining to peers how to write an experiment manual. Both measures (rewritten manuals and letter-of-advice) showed clear effects of the condition in which writers saw real-time readers' feedback on their own manual, on understanding of the genre of an experiment manual, as well as on the understanding of physics topics introduced.

Introduction

Among teachers and other educators there is a general pedagogical notion that the speech act of "explaining" is an effective learning activity, mostly tested in a tutoring or peer teaching situation (Ocel, Palmer, Wittich, Carmichael, & Pawlina, 2003; Topping, 1998). Explaining entails a communicative situation in which one explains something (how it works, what it does, what it is, how it relates to something else, why it happens) to somebody else. When a student explains a curricular item to a peer, the act of explaining determines, and even enhances, his/her understanding of a particular subject. The effect of this learning activity has been tested in various studies on the effect of peer tutoring, peer teaching, and peer evaluation, and is one of the explanations why tutors learn so much by tutoring. This idea is also accepted

*Corresponding author. University of Amsterdam, Graduate School of Teaching and Learning, Wibautstraat 4, Amsterdam 1091 GM, The Netherlands. Email: g.c.w.rijlaarsdam@uva.nl

ISSN 0950-0693 (print)/ISSN 1464-5289 (online)/06/020203–31

DOI 10.1080/09500690500336932

in higher science education, as witnessed by the study advice given on a number of academic websites:

> Also, by talking to other students, you will develop a better understanding of the material. You will find that Princeton students are especially eager to work in groups. After all, the true test of your understanding is if you can explain it to someone else. (Princeton University, 2002)

The idea that giving explanations is helpful has invaded "common sense" about learning. The following comment from a former science student (Lee) illustrates this point:

> Well, if "memorising" would work, I mean, I'm sure it would work on a test or something—but in the long term it doesn't do any good … I know I understand it really when I can go over it, or like tell someone, explain it to someone. If I know enough to explain it to someone else, then I know I have learned it. (Linn & His, 2000, p. 223)

"Explaining" is more than just describing (King, 1997). Explanations contain statements of two types: initial conditions and law-like generations (Hempel, 1965). The sequence of conditions (causes) is linked to one another and connected to the target event (effect) using statistical inferences, temporal approximations or theoretical constructs, unobserved ideas, and scientific models to establish physical causality. The explainer adds in his/her own words the why and how and relates new information with the questions a peer-receiver has or could have. Explaining something to someone else often requires the explainer to think about and present the material in new ways; for example, relating it to the other's prior knowledge or experience, translating it into terms familiar to the other, or generating new examples. Thus, explaining, defending and the like improves understanding for the individual *doing the explaining* because it forces the explainer to clarify concepts, elaborate on them, and evaluate material; that is, alter their own knowledge in some manner (Bargh & Schul, 1980). Various other lines of research have supported the beneficial effects of learners generating and verbalizing their own explanations during learning (Chi & Van Lehn, 1991; King, Staffieri, & Adelgais, 1998).

The learning gain in the explainer stems from three features of explaining: The explainer must retrieve his/her topic knowledge in a setting where he/she changes position—from "knowledge knower" to "knowledge sender"; the retrieved knowledge must be translated into verbal units, attempting to connect the units into some coherent text; and the more obvious the recipient of the message is present and the more access the sender has to receive feedback signals, the stronger emphasis on coherence and accuracy of the text and, therefore, the more the sender will monitor his/her output for relevance and accuracy.

In science education, we may create learning environments that stimulate the use of verbal explanation, aiming at the acquisition of explaining skills (communication about science as an educational aim) and at improvement of learning (communication as a tool for learning). We need to engage students in a classroom community undertaking scientific activities. The classroom can be considered an inquiry community as well as a discourse community. In this study we focus on three key elements of written communication: genre, audience, and the act of writing itself in a context where students explain something to each other.

Theoretical Framework

An important component of science education is *doing experiments*. In earlier days experiments predominantly had an illustrative function: "laws" in physics were illustrated by the teacher demonstrating experiments in front of the class in a presentational format. Later on, students in primary and secondary education were expected to deduce or induce the same kind of laws through hands-on experiences. It turned out that hands-on activities without some form of mind-on supplemental activities were not as effective as promoted (Yore, Bisanz, & Hand, 2003). The function of enculturation was added with an emphasis on viewing active inquiry as the way knowledge is produced and controlled within the discipline of science (Brillhart & Debs, 1981; Gratz, 1990). Writing began to play a role in the curriculum particularly as writing laboratory reports is part of the process of enculturation. A laboratory report may show the extent to which the student masters the experiment and the disciplinary knowledge incorporated within the experiment. At the same time, writing is advocated as a means to communicate about science and as a means to construct knowledge; that is, a tool for learning (Prain, 2002).

In research on writing, a similar distinction is made between "learning-to-write" and "writing-to-learn" (Rijlaarsdam & Van den Bergh, 2004, p. 3). Most of the research on writing has focused on learning-to-write: How do children acquire writing skills and how can education support and enhance this process? The main trend in the past two decades has been on a cognitive approach, guided by the heuristic cognitive process models (Bereiter & Scardamalia, 1987; Flower & Hayes, 1980; Hayes, 1996; Hayes & Flower, 1980; Scardamalia & Bereiter, 1986). In these models, several cognitive activities that could play a role in writing processes were identified. The models were based on think-aloud protocols (Hayes & Flower) or experiments (Bereiter & Scardamalia). A common feature of these models is that writing is seen as an intentional act of problem-solving, as "juggling with constraints" (Flower & Hayes). These models claimed to distinguish between novice and expert writing behaviours, resulting in Bereiter and Scardamalia proposing two models: the knowledge-transforming model and the knowledge-telling model. The knowledge-transforming model aims to describe the writing process driven by a rhetorical situation, whereas the knowledge-telling model describes a process more reflective of a mere memory dump of an associative chain of content elements. What distinguishes the expert from the novice is a sense of purpose (Galbraith & Rijlaarsdam, 1999). For a thorough analysis of writing process models, refer to Alamargot and Chanquoy's (2001) monograph.

From the beginning, the cognitive approach was critiqued by those who stressed the social and motivational context of the writing process (Bizzell, 1982; Brandt, 1992; Nystrand, 1989). These critics believe that differences between writers not only reflect varying cognitive skills, but also the cultural, academic, and personal background of individuals (Galbraith & Rijlaarsdam, 1999). Learning to write is viewed as a process of enculturation in the discourse community. The social–cognitive approach shares the same emphasis on process as the original cognitive studies of writing, but does not treat writing as an individual process. Instead, writing is seen as interaction between writer and reader and not just a neutral, cognitive activity. Given that discourse communities provide individuals with the means to construct their identities, individual allegiances to different discourse conventions are negotiated in a process filled with conflict. One important conflict, for example, is the confrontation between the need for self-expression and the necessity to comply with a set of external constraints; the conflict between the writer's private, unarticulated conception of a topic and the constraints of articulating within a particular set of public conventions, including anxieties about the response of readers to one's private thoughts.

In this socio-cognitive approach, the learning-to-write and writing-to-learn paradigms are allied. Klein (1999) identified four general hypotheses about writing-to-learn that have been proposed in the literature regarding how writing might result in learning. Three of the hypotheses refer to components of writing processes while one is on genre. The hypotheses about the processes involved are less well developed and have been the subject of much less empirical research.

- *Spontaneous text production (shaping at the point of utterance)*. This hypothesis claims that the basic process of encoding thought in language leads to a better understanding of material. This hypothesis relies on the translating or formulating component of the Hayes and Flower model; that is, learning is in formulating the knowledge. Shaping at the point of utterance is expressive writing and transactional, and is based on discovery through free writing. The cognitive load is minimal with the emphasis being based on encouraging the individual to keep writing. There is some evidence that writers generate most new ideas in the phase of initial drafting and that the most original ideas appear in the pre-writing phase and not in writing the final text. But further evidence for this hypothesis is restricted to anecdotes. One of Klein's points of critique on this hypothesis is that exploring the writer's experience and knowledge by writing can indeed generate ideas, concepts, and experiences, but cannot lead to the revision of student's existing conceptions.
- *Forward search.* This hypothesis proposes that the learning is in the reviewing component of Hayes and Flower's model. The crucial ingredient is that the learner selects and organizes ideas in a previously written text, written explicitly to discover or generate ideas. When reviewing, the writer-learner elaborates a detailed representation of the communicative context and of his/her goals in that context. To satisfy these goals, the writer must evaluate and modify his/her existing content. This is the process through which writers recursively review the initial drafts of their texts to transform their ideas iteratively (Klein, 1999). The writer's ideas are preserved in the text and this allows the learner to reread them and to develop them further. Galbraith (1996) suggested that discovery is a consequence

of the writer's implicit disposition towards a topic and the text emerging in the course of the spontaneously articulating thought. A disposition is semantically complex and the expression of this semantic network is necessarily constrained by the limited amount of information that can be expressed in a sentence. A sentence can only partially express the disposition and, therefore, writing is a dialectic between the text so far written and the writer's disposition. Only when the disposition is coherent will the text be coherent; otherwise, additional drafts must follow. The evidence for this hypothesis is weak. Expert–novice studies suggest that experts review initial drafts deeper, but this does not imply that they learnt more.

- *Backward search.* This hypothesis claims that the learning is a result of the goal-directed planning component of the writing process model. As in forward search, the representation of the communicative context is guiding the generating and organizing processes of content. In contrast to the forward search hypothesis, the object of the process is not text, but thought. The writer's rhetorical goals inform the selection of content goals, which in turn inform the selection of operations to transform content. This may lead to changing other content in the text or to changing the rhetorical goal. Studies show that different goals lead to different texts, but until now effects on learning or knowledge transformation have not been tested.
- *Genre.* Klein's fourth hypothesis does not pertain to a component of the writing process, but to the form/function relationship of text. According to the genre hypothesis, the operations and forms of organization required by different genres lead to equivalent operations upon content. As a result, knowledge is organized and ideas are linked together. The specific operations and forms of organization required may vary across disciplines and across different activities within disciplines.

As science education not only strives to educate students into the "laws of science", but also to teach them to communicate about science, to participate in the societal debate on issues where science plays a role—teaching genre is an aim (Yore et al., 2003). A teacher may choose a specific genre because of its hypothesized effects on learning (cognitive power of the genre), as a learning tool, or as an element from the science curriculum that should be acquired (Prain, 2002).

We see a genre as a functional convention, dynamic in nature. A genre is loosely defined by its pragmalinguistic parameters: what is the most effective way to tell something to somebody in some context? At this level, texts from different discourse communities and from different speech cultures are recognized by language users as belonging to the same type, such as a research report, a letter of complaint, a lecture. But at the same time, one will detect differences between the operationalized text within these genres, differences between cultures, differences between disciplinary communities, and differences between levels of expertise. These differences are rooted in the cultural–historical practices of language users.

Naturally, one acquires a specific genre by participating in the practice of the discourse community. Children build up their genre knowledge from language experiences within this community of practice. Genre is embedded in discourse communities and is a social marker. Those who know the rules of the game can easily detect those who do not. A genre represents the communicative context to a certain degree and sets the socially accepted parameters for appropriate content generation, organization, stylistic choices, and voice. For science educators, the question becomes what relevant genres should be part of the curriculum and how should these genres be taught: how to teach a genre in learning-to-write and writing-to-learn contexts?

Studies in effective instruction of writing have revealed two findings: the use of an interactive teaching format as opposed to a presentational format and the qualities of text within genres via all kinds of inquiry activities is greatly improved with active involvement of learners, guided by teacher and peers. Hillocks' (1986) meta-analysis of intervention studies in teaching writing shows that the presentational teaching format does not work. The interactive format, in which students and teachers share roles and activities, generates far better results. Scales (comparing various texts of different quality), inquiry (active involvement with content), and peer feedback with clear goals are ingredients in teaching scripts that resulted in better writing. Schriver's studies (1991, 1992) with reader protocols revealed that students' learning accelerated when they were confronted by their readers, who were reading aloud their text with interruptions. This kind of experience assisted students moving from producing writer-based prose (knowledge telling) to reader-based prose (knowledge transforming).

Genres are not acquired in a deductive way, from knowledge about rules to application of rules, but via active participation in well-chosen communicative activities. Students must construct the genre themselves and develop awareness of text qualities in certain communicative circumstances, through guided trial and error. This process of construction fits the learning-to-learn paradigm in literacy (Bonset & Rijlaarsdam, 2004). In the learning of the genre, they also develop skills to acquire genres in a broader sense; they learn how to acquire the basics of genre and to specify by experience the various adaptations to specific situations.

An effective writing as communication curriculum requires that students play different roles (Rijlaarsdam & Van den Bergh, 2004). First, they must be in a position to experience communication; that is, to experience the effects of written and spoken communication by participating *in communication*. As readers/listeners, they must undergo what a text does to them; and as writers/speakers, they must get insight into these reader and listener processes. They must represent a text as not merely a neutral collection of words, but an *intentionally arranged selection*, intended to communicate a more or less *subjective* perspective of a state of a world. It is assumed that both the writer and the reader experiences within such a context support the construction of intentional cognition, communicative awareness, and insight into the fundamental rhetoric of genres. This is the experiential level of communication depicted in the first pair of arrows in Figure 1. As soon as the reader not only responds to texts, but also provides verbal comments, a new communicative situation comes into being, in which the reader becomes the sender and the initial writer

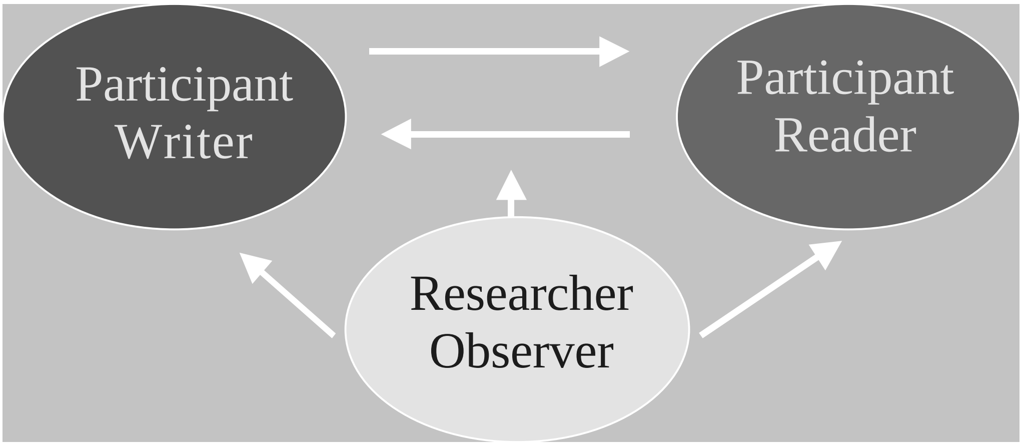


Figure 1. Student-as-learner participation model in the language arts curriculum (adapted from Rijlaarsdam & Van den Bergh, 2004)

becomes the receiver. Now both participants are involved in two roles: they participate in the communication about the text and both are moving from participant in the initial communication to a metaposition (they write and read or speak and listen *about* that initial situation). This second level of communication can strengthen and deepen the initial construction of intentional cognition. A third condition, not implemented in the present study, is the observer role: here the learner is not a participant in the process of writing or reading, but observes the writing or reading of the communication between writer and reader. In studies on acquiring argumentative skills of 15-year-old students, observing other students while they were practicing reading and writing proved to support the learning of writing and reading skills better than practicing these skills (Braaksma, Rijlaarsdam, & Van den Bergh, 2002; Braaksma, Rijlaarsdam, Van den Bergh & Van Hout-Wolters, 2004; Couzijn, 1999).

From writing-to-learn studies, it is less clear what ingredients of instruction scripts are relevant or effective. Bangert-Drowns, Hurley, and Wilkinson's (2004) recent meta-analysis of 46 studies was discouraging for those who support the hypothesis that writing facilitates learning: the average effect size proved to be small (median = .20; range = –.08 to 1.48). Treatment effects were correlated with treatment length (r = .33). Time per writing assignment ranged from 3 to 55 min: longer assignments were related to less positive effects (r = –.70). This might be caused by lower motivation, because writing is tough, especially for weaker writers, and by less time spent on coverage of course content. The number of assignments was not related to effect size. Whether or not the treatment contained personal writing did not influence the effect size, nor did the presence or absence of feedback. The contribution of metacognitive prompts was significant: when treatments included prompts requiring students to evaluate their current understandings, confusions and feelings in relation to the subject matter, the effect size was significantly larger although still small (.09 versus .26).

This meta-analysis stipulates that a metacognitive position of the learner/writers could contribute to learning. Monitoring is a crux for learning. This fits our assumption that, for learning-to-write, monitoring is the facilitator for improvement, for building awareness about what a text should include in certain circumstances or contexts to be effective (Braaksma, Van den Bergh, Rijlaarsdam, & Couzijn, 2001).

What Genre Should be Taught in Science?

Numerous genres have been identified in science text and the list is steadily growing (Gallaghan, Knapp, & Noble, 1993; Goldman & Bisanz, 2002; Unsworth, 2001). One genre that is relevant to include in secondary education is the hybrid genre associated with a manual for an experiment: instruction for the steps in an experiment and theoretical explanations for anticipated observations, measurements, and results. The experiment manual provides logically ordered series of procedures with scientific or safety justifications and anticipated observations and measures with theoretically based explanation that warrants the predicted outcomes. The experiment manual is intended to provide the procedures and rationale for a student to complete and understand a science investigation. It is different from a traditional laboratory report because of the time perspective: in an experiment manual the observations and measurements are anticipated or predicted, while a traditional laboratory report states what was observed or measured. The experiment manual is communicative functional in classrooms: students must inform other students about what steps were to be executed, what phenomenon was to be studied, what might happen, and why these results occurred. It is also a step towards the work of "real" science and, as such, an introduction into the world of science.

The manual can serve both the learning-to-write paradigm and the writing-to-learn paradigm. As a manual is a strongly reader-oriented genre, students must have an awareness of the needs of their readers. At the same time, a science experiment illustrates a scientific insight and this insight is embedded in the manual that students will compose. Note the hybrid genre introduced is more complex than just listing the directions. A physics experiment in a classroom serves to build a better or a new understanding of a certain phenomenon in nature. Embedded in the directions (a linear sequence of procedures) are the critical observations ("if the water does not flow into the Erlenmeyer flask"), the answers to related questions ("the water can't flow into the water, because the air in the bottle takes space"), theoretical thinking ("you may prove that by lifting my finger. What you then see is that the water flows into the Erlenmeyer flask") and explaining relations between two events ("why is it that the water flows when the finger is lifted from the straw?").

The present study compares the effect of various pedagogical variables as choices within the genre hypothesis:

1. The effect of pre-writing versus post-writing activities. The pre-writing activities conditions refer to learning by text scales (Hillocks, 1986) and the post-writing activities conditions refer to experiencing reader feedback (Schriver, 1991,

1992). Here we compare what works best in the initial communicative act: starting the learning sequence in the role of reader and then writer, or starting in the role of writer, then reader and writer again.
2. The effect of revision as part of the forward search hypothesis (Klein, 1999). Here we add a new distinction: the weak hypothesis and the strong hypothesis (Rijlaarsdam, Couzijn, & Van den Bergh, 2004; Rijlaarsdam & Van den Bergh, 2004). Simply revising (working on an earlier version) is called the weak hypothesis. Communicative-driven or goal-oriented and reader-oriented revision (i.e., revisions after some kind of intervention) is called the strong hypothesis.
3. The effect of "second-order" writing in addition to pre-writing activities as compared with addition to post-writing activities. Second-order writing refers to an additional writing task, in which participants communicate about communication—what we call a secondary communication act (see Figure 1).

Method

This study utilized an experimental design in which Grade 9 students were randomly assigned to various conditions of pre-writing/post-writing activities, revision with/without reader feedback, and second-order writing activities while completing an experiment manual writing task. Students crafted a hybrid text for peers containing procedures and explanations required to conduct and understand a simple physics experiment.

The Science Experiment

The experiment manual for a simple physics investigation for peers should serve two goals: it should enable a classmate to *perform* the experiment without problems as well as to *understand* the procedures, anticipate the observations and measurements, and help explain these data as evidence for a claim. These goals should stimulate the writer to take several needs for information into account: the manual should contain explanations of the goal and mechanism of the experiment as well as explicit and clear instructions on how to perform the experiment.

The physics experiment was presented individually to students in some conditions. The research assistant introduced this activity. This experiment will show you that *air takes up space*. This means that boxes or bottles, which are commonly called *empty*, in fact contain air and, as long as the air is inside, nothing else can go in. If you fill an empty box with books, the air has to go out at the same time. Three figures were shown for the experiment (Figure 2). The research assistant read the accompanying text out loud.

Subsequently, the student was asked to perform this physics experiment. Several problems had to be overcome. One problem was choosing the necessary and appropriate equipment. At the student's disposal were the following objects: three bottles (one too small, 150 ml; one too big, 400 ml; and one the appropriate size, 250 ml), four corks (one too small with a hole; one too big with a hole; one the

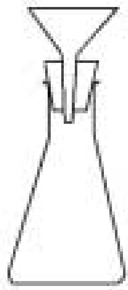
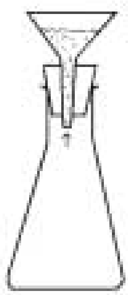
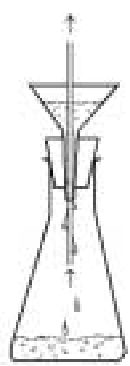
	***Figure 2a*** ***Construction with bottle, cork and funnel*** "To see for yourself that air takes up space, we will do the following experiment. In this figure you see a construction that you can make with some objects on the table (one bottle, one cork and one funnel). You have to choose the objects well so that the construction fits exactly, i.e. so that no air can escape."
	***Figure 2b:*** ***Water poured in the funnel will not run in the bottle*** "If water is poured into the funnel as in this figure, this will not run into the bottle. The reason for this is that the air is still inside and cannot escape. You may say that the air 'obstructs' the water."
	***Figure 2c:*** ***A straw helps the air escape*** "This figure shows how this problem is solved. A straw that is stuck through the funnel creates a passage for the air, just like a chimney. The escaping air will free up space within the bottle, which is immediately occupied by the water. Therefore you will see that the water starts flowing into the bottle. Now you must put your finger on the top of the straw. The air cannot escape any longer and the water will stop flowing. As soon as you lift your finger, air and water will flow: the air out of the bottle and the water into it."

Figure 2. Instruction for the experiment: slides and texts

appropriate size but without a hole; one the appropriate size with a hole), three funnels (one a size too big; one two sizes too big; one the appropriate size), and three straws (thin and red; too big and red; too big and blue). The students spent some time finding the correct combination of equipment; that is, the medium-size bottle, the medium-size cork with a hole, the smallest funnel, and the thin red straw. All other combinations did not work properly. In their manuals, the students had to identify properly which objects they used, or the experiment would fail.

A second problem consisted of making the apparatus airtight. If the apparatus was not hermetic, with the objects pushed firmly together, the experiment would fail. As soon as water was poured into the funnel, the apparatus would show whether it was hermetic: if not, the water would run into the bottle and not stand in the funnel.

A third problem arose when the straw was put into the bottle through the water in the funnel. Placing the open straw through the standing water resulted in a small water plug within the straws, which had to be removed (by carefully blowing or sucking on the straw) before the air could pass through.

The student had to overcome these and other problems to bring the experiment to a successful conclusion. Only when necessary—for example, if the student got stuck and could not resolve the problem himself/herself—did the research assistant give advice.

Subjects

The subjects consisted of 107 students from the ninth grade of three different secondary schools, representing two academic tracks, preparing for higher education, at the polytechnic and the academic level. Their average age was 15 years. The students were randomly selected from nine classes after removing from each class the five lowest achievers and the five highest achievers in writing skill as judged from a recent school assignment. We assumed that the within-group variance on writing skill would decrease as a result. The reason for this homogenization is that the study focuses on *typical* student behaviour, not on differences between high-skilled or low-skilled students. The selected students were randomly assigned to the seven experimental conditions.

Procedures

All writing and reading assignments were administered individually; that is, each student worked privately in a room with a research assistant giving standardized information and answering questions. The writing assignments (experiment manual, letter-of-advice) were given orally and were read from a scripted protocol.

The research design for learning to write an experiment manual is summarized in Table 1. Sample sizes for each condition represent the students completing all components of the condition.

Table 1. Summary of research design: conditions, sample sizes and manual tasks

Condition		*n*	First version manual	Revised version manual
0	Writing only	14	√	
Post-Writing Activities Condition				
1	Learning by enhanced revision	14	√	√
2	Learning by reader feedback: video own text	19	√	√
3	Learning by reader feedback as 2, plus receiving written comments	7	√	√
4	Learning by reader feedback as 2, but video someone else's text	20	√	√
Pre-Writing Instruction				
5	Learning by text scales: performing peer-written manuals, think aloud	20	√	
6	Learning by text scales: as 5, plus producing written comments as feedback	14	√	

Conditions—Pre-writing activities: Learning by text scales

Participants received two authentic manuals written by unknown peers. The research assistant gave no explanation. No figures like those in Figure 2 were shown. The student-readers were instructed to perform a physics experiment as described in the manual, using the objects on the table in front of them. They were asked to think aloud during the experiment and were prompted to continue talking whenever they fell silent. The readers were also instructed to give real-time comments on the quality of the manual. The readers' performance of the experiment and the oral comments were recorded on videotape. The following instructions were given:

> You will need some of the objects on the table in the following physics experiment. To execute the experiment, you will have to use the manual. In this manual, someone has attempted to describe for you what you should do in order to perform and understand the experiment.
>
> Your task consists of two activities. First, you should *precisely* follow the instructions in the manual. Try to do exactly what is written: no more, no less. Do not rely on your intuition; just follow the instructions literally. Whenever the manual is not clear to you, say so *immediately*. Also point out any missing or incorrect information. Think aloud continuously while doing the experiment, so I can follow your line of thinking.

Some random selected participants then wrote a general comment about the quality of the manual. The instruction was:

> One of your fellow students has tried to write this manual as clearly as possible, so that it would cover all the necessary information for you to do the experiment well and to understand it. Now I ask you to write down your comments on this text. Do you think it is a good manual? What are its weak points, what was unclear or incorrect, what was

> missing? And what are its strong points, which helped you to do the experiment well? Look it over thoroughly and give as many comments as possible.

Writing a short list of comments took about 3–4 min. Second, all participants were asked to write their own version of the manual without using the manual they had experienced. They were instructed:

> Now write your own manual for another student who has to do the same experiment. Think of your experiences with the manual and your comments on it.

The average time for writing their version was 12 min. When they had completed the assignment, the students were prompted to read their text aloud and correct mistakes.

Conditions—Post-writing activities: Reader feedback

During a first session, all participants wrote a first version of the manual after having mastered the execution of the experiment. The differences between the five post-writing conditions were implemented in the second session, 3 weeks later.

No revision. Participants skipped the second session. The conditions served as an internal control group—writing only.

No reader feedback. Here the student-writer was his/her own reader. The writer received his/her manual written 3 weeks previously and was asked to redo the experiment. The same collection of objects was at his/her disposal. The instruction focused the students' attention on revising the manual text:

> Many people discover problems in their texts when they read them after some time. While you are using your own manual, you may get useful ideas for improving it. You will probably find that it contains some good points, but it may also be susceptible to improvement. So your last task will be: think of as many points as possible that may cause problems for another student who has to execute and understand the experiment without problems. Revise your manual so that all shortcomings are corrected.

First, the students were asked to read their text aloud. Second, they had to use the text in redoing the experiment. While doing the experiment and after having finished, they had to generate and write down as many comments as possible on the quality of the text. The final task was to rewrite the manual, using the first version as a model.

Reader feedback. Students were confronted with authentic readers. Three conditions of reader feedback were implemented. In two conditions writers were provided with videotaped readers of their own manual and in one condition writers were provided with videotaped readers of somebody else's manual. After rereading the first version aloud, student-writers were shown video-recordings of student-readers who had used their own or somebody else's manual. The video observations lasted for 10–12 min and contained non-edited recordings of students using and commenting on the experiment manual. For the two "own manual" conditions the instruction was as follows:

> Some fellow students tried to do the experiment using the manual that you wrote. You will see on video what has come of it. What should you look for? While looking at the video, you may get useful ideas for improving your text. You will probably find that it contains some good points, but it may also be susceptible to improvement. Try to concentrate on information that will help you in improving your manual. Does the student you see understand your intentions? Does he/she perform the experiment well? What are his/her comments on your text? Take notes while looking at the video. Whenever you want, the tape can be stopped so you can write or ask something.

After having seen the video recordings, a random set of the participants was given the written comments of the observed reader as a supplement to the notes they had taken themselves. Other participants did not receive written feedback. Then all participants in these two conditions (reader only and reader/written comments) were instructed to rewrite their first version:

> You have seen one or two students who used your manual. You may have gotten ideas for the improvement of your text. So your last task will be to revise your manual so that all shortcomings are corrected. Make good use of the information you received by looking at the video recordings.

The feedback and revision procedures for writers in the condition where somebody else's manual were similar, but the observed reader responded to a manual written by an unknown peer writer. As a consequence, the observer was confronted with communicative failures of a peer and had to determine whether these shortcomings could also be repaired in his/her own manual. These students also had to revise their text after having seen the video recordings. They did not receive written comments from readers. Again, the revision of the manual took 15 min on average.

Conditions—Second-order writing task

As outlined earlier when describing the Pre-Writing and Post-Writing Activities Conditions, in one of the Pre-Writing Conditions and in one of the Post-Writing Conditions students had an extra writing/reading task. Students wrote a comment on manuals (Pre-Writing Condition) or received written comments following the reader feedback (Post-Writing Condition).

Generalization Task for All Conditions

All students were given a writing assignment 3–4 weeks after the sessions that was intended to tap their knowledge about experimental manuals. Declarative knowledge is supposed to be the first stadium of knowledge expansion when learning in a new domain (Singley & Anderson, 1989). If students in one condition demonstrate more declarative knowledge about the genre than in another condition, then we may conclude that they are likely to have learned more about "giving good instructions and explanations".

All students were asked to write a letter-of-advice to a fictitious fellow student, who needs advice on how to write a manual for an unknown physics experiment. We assumed that the number and quality of the pieces of advice that were given were an indication of the student's knowledge about criteria for writing good manuals. The research assistant gave the following instruction:

> Imagine the following situation: one of your classmates comes to ask you for advice. He or she has to write a manual for a physics experiment and has heard that you have some experience in writing such texts. Although you do not know what sort of experiment your classmate has to write about, still you think that you can explain what points he/she should pay attention to when writing a good manual text. The manual will be used by students of your own age, who must be able to understand and perform the experiment. Write a friendly note to your classmate and give *as many pieces of advice as possible*, clearly stated, that would help him/her to write a high-quality text.

The students were given 20 min for writing this letter-of-advice. The research assistant stated explicitly that the students were not to write a manual, but rather *describe* it.

Quality of Experiment Manuals: Instrumentation and scoring

We constructed from a subset of texts an optimal experiment manual that functioned as a scoring rubric for assessing the quality of manuals. This rubric consisted of five parts or episodes: an *Introduction*, three episodes called *Construction*, *Water*, and *Straw*, and a *Conclusion*. For each of these parts, a list of standard information elements was stipulated, containing four types of speech acts: directions, observations, precautions, and theorizing (Appendix A). Each of the 29 elements addresses a need for information a reader may have in order to understand the physics experiment. Some elements concern the goal of the experiment or its mechanism; others concern the necessary steps, the necessary objects, the observations to be made, explanations for particular observations, and hints or tips.

Each text was scored on the occurrence of the 29 information elements. That is, for each element it was determined whether the writer had realized it in the text (the formulation may differ to some extent). If so, the element was scored with a full point. If not, the score was zero. If the element appeared to be present but serious doubt existed, the element scored a half point. Independent, trained scorers did the scoring. Three scorers scored one-quarter of the manuals. The inter-scorer reliability (Cohen's kappa) turned out to be very high (0.86–0.88); therefore, it was decided that two scorers would assess the remaining manuals. Their inter-scorer reliability was 0.86.

Two examples of a comparably weak and a good manual will illustrate the scoring method.

> Manual 1 (7 points: elements 5, 7, 10, 11, 16, 19, 20)
>
> *You take a bottle. You put a cork in it. You put the funnel in the cork's hole. You put water in the funnel and then you see that no water comes into the bottle and that's why you put the straw in and then the air can go away and the water in.*

Readers using this text would definitely run into problems choosing the correct materials. It is very hard for readers to find the correct combination if they are completely unaware of what apparatus is going to be made. The semi-causal links "and then you see that …" and "and that's why you …" are not really informative. A reader probably would not understand what the experiment is about because the aim is not mentioned explicitly.

Manual 2 (21 points: elements 2–4, 6–11, 13–17, 19–20, 22–26)

- *You see 3 bottles and 3 corks on the table.*
- *Take the middle bottle and the cork without marks or spots, with an opening.*
- *Put the smallest funnel in the cork's opening.*
- *Push the funnel in really well (no air may escape).*
- *Pour water into the funnel.*
- *You will see that the water stays in the funnel (if you have pushed everything really well together, so no air can escape).*
- *In order to let the water run through, it must be made possible for air to escape.*
- *Put a straw (the thin one) through the funnel until halfway the bottle.*
- *You may not leave the straw in the funnel itself, there's water in there that stops the air from flowing out.*
- *If the water still doesn't run through, there must be water inside the straw, so the air can't go through. (Suck if necessary.)*
- *The water will run through now.*
- *Put your finger on top of the straw, then the water will stop, because no air can escape.*

This text exceeds the first in length as well as in quality: it contains more detailed descriptions, explanations, and hints for readers.

It was possible to interpret almost every textual element written by the students as one of the 29 elements in the standard manual. The object descriptions were scored positively if the description referred *unambiguously* to one of the objects on the table, because unambiguity is the essence of referential communication (Sonnenschein & Whitehurst, 1983, 1984; Holliway & McCutchen, 2004). "Take a bottle" is therefore scored as inadequate (score: 0) while "take the medium size bottle" is considered adequate because only one object answers this description (score: 1).

Generalization Task: Instrumentation and scoring

Students demonstrated their discourse knowledge about criteria for good experiment manuals in a letter-of-advice. They were asked to give as many pieces of advice as possible; this open task served to evoke a non-selective memory search and unimpeded writing. We assessed both the quantity and quality of the advice in each letter. Assessing the quality was done by means of a scoring rubric that contained a categorization of possible pieces of advice. In the categorization, it was assumed that the pieces of advice were either *process oriented* or *product oriented*. Product-oriented advice concerned *style* or *content*. Style and content were further subdivided, resulting in a system of quite specific and recognizable categories (Appendix B). Two scorers categorized the advice according to this scoring rubric. The inter-rater reliability on levels A.1–A.3 and on levels B.2.1–B.2.5 was 0.82 (Cohen's kappa). On levels B.1.1–B.1.11 (subcategories that are less easy to discriminate), the inter-rater reliability was 0.70. These reliabilities were judged to be high enough for making between-group comparisons.

To illustrate our scoring rubric, we present two letters-of-advice and their scores. The first letter-of-advice contains some very general product advice regarding style, which applies to many types of texts. There are two pieces of advice from category B.1.3. If the second seems to be just a paraphrase of the first, cases like this were counted as one piece of advice. The last piece of advice is process oriented: a suggestion to brainstorm before starting to write.

A comparatively weak letter-of-advice:

Score:	Dear someone,
	I am happy that you asked for my advice.
	If you want to write a good manual, you should mind these points:
B.1.3	—Write neatly and precisely, or the person who has to do
B.1.8	the experiment won't be able to read it.
B.1.5	—Don't write too much nonsense.
(B.1.3)	—Write everything in detail.
A.1	—First write down all the important things you can think of.
	I hope this helps. If you still have problems, do call me.

The second letter-of-advice is longer and shows more variation in pieces of advice. The writer seems to have followed the course of a writing process: advice concerning orientation, writing, and revision is present in a natural order (A.1, A.2, and A.3). The letter-of-advice contains three content-oriented pieces of advice concerning the use of instructions, objects, and precautions, albeit not very precise. Also three style-oriented pieces of advice are given.

A comparatively good letter-of-advice:

Score:	Hello, here is my promised letter with advice for your manual for the physics experiment:
A.1	It is very handy if you start by thinking really well
B.2.2	about the things that you need and that you don't and
B.2.1	about what has to be done (If you don't make such a plan, it is better to not write at all.)
B.1.6	Next: keep the order of the activities in mind and also the moments when some tool has to be used.
A.2	Then it is time to start writing:
B.1.7	Emphas the most important things, such as:
(B.2.1, B.2.2)	How to do it! And: which object to use.
B.1.8	You must not forget that children your age must be able to understand your manual.
A.3	So when you are finished, you check it yourself and
(A.3)	correct mistakes. Check if you would be able to do the experiment with your own manual (faultlessly!)

B.2.4 Only if you are certain that you've done everything to keep your classmate from running into problems, you can hand the manual in to the teacher.

All the advice was categorized by using the scoring rubric. If phrases contained more than one advice (as in "Describe chronologically all the things that need to be done"), the advice would receive a double score; that is, a positive score on B.2.1 (instructions) and B.1.6 (organization). To increase the sensitivity of our assessment, the product-oriented pieces of advice concerning content were rated on quality. In this way, justice can be done to differences between advice like:

> Write down what to do with the tools.

and

> Describe very accurately which of the available objects must be used, so that the reader doesn't have to guess and try for a long time; also tell him which precise acts must be performed with the objects.

Two raters scored each content-advice as 0.5 (content advice given, but in a very general wording, or implicit) or 1 (content advice given, precisely formulated, with motive or example). Cohen's kappa for inter-rater reliability was 0.80.

Data Analysis and Results

Descriptive statistics for each condition were calculated for the experiment manual and the letter-of-advice. These data were tested for significant differences using a series of one-way analysis of variance (ANOVA) and pair-wise comparisons. In some cases we applied between-condition analyses, when we compare the effect of conditions on the first version of the manual (pre-writing activity factors). For assessing the effect of post-writing factors, we applied analyses of covariance with the quality of the first manual as covariate and parallel, repeated measurements with the two manual scores as within subject scores. In explorations we tried to assess the effect of the various instructional factors via regression analyses, with instructional factors and first manual quality as possible explanations for the variance in the quality of the second manual quality.

Quality of Manuals

The descriptive statistics for the first and revised versions of the experiment manual are presented in Table 2. Sample sizes, mean scores, and standard deviation are reported for each condition.

Pre-writing activity factors: Responding as a reader and composing comments. Only in the learning by text scale conditions were pre-writing activities performed as a real reader; reading before writing the manual where participants were invited to think aloud and respond. A comparison between the pre-writing activities/no pre-writing activities conditions (condition 0 versus conditions 5 and 6 combined) on the first

Table 2. Descriptive statistics for all conditions (sample size, mean scores, standard deviations)

		First version		Revised version	
Condition		*M*	*SD*	*M*	*SD*
0	Writing only (n = 14)	9.79	2.93		
Post-Writing Instruction					
1	Learning by enhanced revision (n = 14)	9.68	3.48	10.11	3.43
2	Learning by observing readers: video own text (n = 19)	10.95	2.19	14.31	1.67
3	Learning by observing readers: video own text, plus receiving written comments (n = 7)	10.36	2.38	17.43	3.14
4	Learning by observing readers: video other text (n = 20)	11.35	3.27	12.85	3.11
Pre-Writing Instruction					
5	Learning by text scales: performing the manual, think aloud (n = 20)	11.97	3.16		
6	Learning by text scales: as 5, plus producing written comments as feedback (n = 14)	13.54	2.55		

version quality revealed a significant effect for pre-writing (F = 11.57, df = 1, 106, p = .001), due to condition 6 since the scores from condition 5 compared with the conditions without pre-writing activities are not significantly different when tested with an ANOVA (F = 3.65, df = 1, 92, p = .059).

Second-order task as pre-writing activity. The added value of the secondary communicative act, composing comments, was explored using a one-way ANOVA. The analysis revealed non-significant main effect when conditions 5 and 6 are compared (F = 2.34, df = 1, 32, p = .136).

To explore the effects of both pre-writing instruction factors—responding as a reader and composing comments to manual authors—we ran regression analyses for each of the four manual qualities on each of the instructional factors. A significant contribution of responding as a reader is observed for instructions and for composing comments for objects (regression weights respectively β [standardized] = .362, t = 3.38, p = .001; β [standardized] = .270, t = 2.34, p = .021). For the other factor, composing comments, no effects were observed.

Effect of post-writing instruction factors: Revision only, readers' feedback with/without written comments revision. No effect of enhanced revision was observed for condition 1. A within-subjects analysis with first versions and revised versions of the experiment manual as repeated measurements for condition 1 subjects revealed non-significant differences (repeated measures, F = .85, df = 1, 26, p = .37). A between-groups analysis, comparing the enhanced revision condition with the writing only condition, revealed no significant difference (F = .01, df = 1, 26, p = .93).

Revision after observing readers. An effect of observing readers is found for an analysis of covariance between conditions with the initial quality as covariate ($F = 17.54$, $df = 1, 60$, $p < .001$). The effect size of the three observing conditions is large (1.2). Comparison of the effects for the type of reader feedback (own manual, somebody else's manual) reveals that the condition of own manual outperforms the condition of somebody else's manual (repeated measurements: interaction between time of measurement and condition $F = 9.40$, $df = 1, 37$, $p = .004$, effect of time $F = 63.80$, $df = 1, 37$, $p < .001$).

Second-order task as post-writing activity. An effect of written comments is observed when the two conditions of reader feedback on one's own manual are explored using a repeated-measurement analysis (effect of time $F = 124.90$, $df = 1, 24$, $p < .001$); interaction time and condition $F = 15.71$, $df = 1, 24$, $p = .001$). The condition that had access to written peer feedback outperforms the similar condition without written feedback.

When the three feedback factors manipulated in this study are included in a regression analysis to predict the revised manual score, given the first manual score, all three factors and the quality of the first version of the manual contribute significant to the prediction ($F = 43.24$, df = 4, 59, $p < .001$). The written comment factor contributed significantly in predicting the general quality of the manual (β [standardized] = .327, $t = 4.48$, $p < .001$) and three out of four of the component scores (theory, objects, and precautions, β [standardized] = .362, .260, and .244, respectively).

Effects of pre-writing and post-writing activities. An ANOVA of the pre-writing and post-writing activities revealed a significant effect of type of activity ($F = 5.15$, $df = 1, 80$, $p = .026$). Further consideration indicated that the three reader feedback conditions compared with the two pre-writing conditions produced a significant effect (effect size = .51).

Generalization Task

The delayed, transfer writing task (letter-of-advice) was designed to access and assess students' knowledge about experiment manuals. The mean scores and standard deviations for all advice processes, style, and content categories are presented in Table 3. A series of ANOVAs was performed on these data. The ANOVAs yielding significant effects were followed by pair-wise post-hoc comparisons.

Students from the post-writing condition 3 (writers who observed their readers and received written comments for the purpose of revision) produced more pieces of advice than the other students. This holds for the total number of pieces of advice, process-oriented pieces of advice, and content-oriented pieces of advice. The differences with respect to content-oriented advice are most significant,

Table 3. Letter-of-advice: mean scores, standard deviations, and post-hoc comparisons (Duncan, $p < .05$) for the total number of pieces of advice and for the number of pieces of advice for each subcategory

Advice	Condition	*M*	*SD*	Comparison
All categories				
0	Writing	4.93	1.73	
1	Generating self-comments	5.14	2.11	
2	Observing readers: video own text	5.47	2.22	
3	Observing readers: as 2, and receiving written comments	10.86	2.73	3 > all
4	Observing readers: video other text	5.50	2.33	
5	Learning by text scales: performing the manual, thinking aloud	4.45	2.11	
6	Learning by text scales: as 5, plus producing written comments as feedback	7.50	1.70	6 > all but 3
Process				
0	Writing	0.57	0.85	
1	Generating self-comments	0.71	1/07	
2	Observing readers: video own text	1.15	1.30	2 > 4, 5, 6
3	Observing readers: 2, and receiving written comments	1.71	1.38	3 > all but 2
4	Observing readers: video other text	0.50	0.76	
5	Learning by text scales: performing the manual, thinking aloud	0.30	0.57	
6	Learning by text scales: as 5, plus producing written comments as feedback	0.29	0.47	
Style				
0	Writing	1.43	1.16	
1	Generating self-comments	1.50	1.16	
2	Observing readers: video own text	1.79	1.47	
3	Observing readers: 2, and receiving written comments	3.43	1.71	3 > all but 6
4	Observing readers: video other text	2.00	0.79	
5	Learning by text scales: performing the manual, thinking aloud	1.25	1.16	
6	Learning by text scales: as 5, plus producing written comments as feedback	2.71	1.06	6 > 2, 4, 5
Content				
0	Writing	2.93	1.77	
1	Generating self-comments	2.92	1.64	
2	Observing readers: video own text	2.53	2.17	2 > 0, 1
3	Observing readers: 2, and receiving written comments	5.71	1.50	3 > all, but 6
4	Observing readers: video other text	3.00	2.49	
5	Learning by text scales: performing the manual, thinking aloud	2.90	1.68	
6	Learning by text scales: as 5, plus producing written comments as feedback	4.50	2.11	6 > 2, 4, 5

because this category represents knowledge specific for the type of text that was written.

Students from pre-writing condition 6, who not only used the text but also commented on it, appear to have learned more than students from the comparable condition without written comments. The effect size is large (1.5).

Executing a second-order writing task has merit. We cannot determine whether the feedback and revision activities have allowed these students to acquire *more* knowledge about instructional texts or whether these activities have resulted in knowledge that is *more readily retrievable* from memory, which in itself is an important quality of learning as well.

Examples of content-oriented advice. We stated that the content-oriented advice category is particularly relevant to learning by explaining in the context of a physics experiment. Stylistic and process-oriented criteria for good explanations are suitable for many different types of explanations, but the content criteria focus on doing and understanding experiments. Therefore, we offer a sample of advice from this category given by the students.

1. Advice regarding instructions
   - "Describe what that person has to do."
   - "Tell how that person must put everything together and then to execute the experiment."
   - "Write down all the things you want that student to do."
   - "You must include all the actions in your explanations."
   - "Explain how they can do the experiment successfully."
2. Advice regarding objects
   - "Tell about all the things he needs."
   - "Take notes of everything on the table, describe the objects."
   - "… with which things he must make the construction, the amount and so."
   - "Describe precisely the things he needs."
3. Advice regarding precautions
   - "Give hints for people who cannot solve it themselves."
   - "Make sure that no misunderstanding can arise."
   - "What is going on if the experiment does not work, how to improve that?"
   - "If they don't know what they have to do, it helps to tell about the meaning of the experiment."
   - "Tell about the things that can go wrong, what to do."
4. Advice regarding theory-based explanations
   - "Write down what has to come of it, what it is necessary for."
   - "Write conclusions, he must know what he is doing."
   - "Tell about the things you see happening during the experiment."
   - "You must include the answers to their questions."
   - "Tell what the experiment is about and what the intention is, give explanations after every step, explain the final result."

- "The 'why' of the actions, what he is doing, the reason for doing this."
- "The reason for this construction, connection with certain laws or observations, show them why."

Discussion

With the limitations of the design, conditions, and procedures of the study about experiment manuals with Grade 9 students, five conclusions can be drawn from these results:

1. Pre-writing activities, the use of peer written texts as a starting point, was effective.
2. Post-writing activities were more effective than pre-writing activities.
3. We found no evidence for the weak forward search hypothesis. Even when students are encouraged to revise their text by the research assistant, they do not improve their original manuals under their own initiation.
4. The strong forward search hypothesis was supported. When the need to revise is stimulated by observing peer-readers reading a first version of the manual, students improve their second version considerably.
5. Adding a second-order writing task creates an extra effect on generalizing knowledge in the pre-writing condition, an even stronger effect in the post-writing condition, and an effect on the revision in the post-writing condition.

Many of the improvements in the revised manuals are signs of acquiring the features of the genre. They resulted from simple instructional devices or tasks, without teacher intervention, building upon the naïve knowledge of genre already available in students. A teaching strategy that uses the available prior genre knowledge and that stimulates a natural acceleration of learning by confronting writers with the effects of their writing is effective. The principles of an interactive learning environment and experience of and inquiry about communication proved to work well (Hillocks, 1986). Note that we did not contrast an inductive instructional pattern with a deductive pattern; therefore, we did not explore whether an inductive acquisition of a genre is more effective than a deductive approach. What we showed are factors in the inductive learning environment that facilitate learning-to-write and writing-to-learn (more theoretical accounts included in the manual). Given the state of the art in writing research, a choice for deduction (theory first–then application) is not an evidence-based choice. The act of writing a text does impose a high cognitive load for learners. When this task is loaded with another constraint—applying new theory about the genre—the task requires even more cognitive effort, leaving the learner with the problem of how to manage this problem: minimize the effort put into the quality of the text or the effort to be put into the acquisition of the newly introduced text theory?

Authentic feedback appeared to be a strong impulse for goal-oriented and audience-oriented revision. The additional communicative act, where feedback givers and feedback receivers communicate about the quality of a text, had additional effects. We found that students were able to both address the quality of the texts they received and to acquire more knowledge about necessary elements in manuals, but only if they did the extra task of writing evaluative comments about the manual they had to work with. This finding matches Bangert-Drowns et al.'s (2004) result that metacognitive prompting facilitates learning in writing-to-learn settings.

Why would the writing or processing of comments be so crucial to the effectiveness of "learning by explaining something to somebody else"? One plausible explanation for this phenomenon is that formulating the advice requires investing time and effort in the topic (which the other group did not do) and that writing the comments is a means to conceptualize and generalize, so that the comments could be more easily transferred to the text they had to write and could more easily be remembered later: from episodic to semantic memory (Tulving, 1972). The writing activities may well affect the construction of knowledge or at least the condensation of experiences into more general cognitions.

There are a few factors that may have influenced the internal validity of the results. One might argue that unfair comparisons were made, since the time-on-task differs between the conditions. This is true. The difference is largest *between* the writers' and the readers' conditions: the readers worked in two sessions whereas the writers worked in three sessions. So reader/writer conditions comparisons are confounded by time-on-task. But note that *within* the various writers' conditions, the differences in time-on-task were small. The written comments added to condition 3, for instance, did not prolong the session by more than 5 min. Generating self-comments lasted, due to the method of brainstorming and elaboration, approximately the same as observing the readers on videotape. The effects we observed are unlikely due to differences in time-on-task.

One could critic our limited operationalization or assessment of learning. We have chosen an indirect assessment, by trying to measure the students' declarative knowledge about instructional manuals. Of course, having such knowledge at one's disposal does not guarantee a successful implementation in new tasks. Adding a transfer task (writing a manual for another experiment that illustrates a physics law) could contribute to a more solid outcome.

Some concerns could be expressed about the external validity of the study's generalizability of the results. First, the question arises to which population the results can be generald. Very good and very weak writers were excluded from the experiment. Whether the pedagogy would be equally effective for these groups of students remains unclear and this question can only be addressed by new research. Second, one may wonder to which types of texts the results may be generalized. We specifically chose experiment manuals in the context of a simple physics inquiry, because of the appellative function of the text and the visibility of potential comprehension problems. With more intellectual types of explanations, such as explaining the solar system or the Keppler laws, it could be hard or even impossible for the observer to detect communicative weaknesses. This concern depends on the reader's task. In the present experiment, the reader was supposed to read and think aloud and was prompted to keep talking whenever he/she fell silent. Once readers were familiar with this activity, they disclosed many of their thoughts, concrete or abstract, certain or intuitive, to the observers. It is also important that the reader's task is a well-defined task; not just to read the text and provide comments, but a more focused assignment by which they can demonstrate their understanding. The reader could be asked to draw a scheme based on the explanation or to solve a problem by utilizing the explanation. Feedback by readers, in our opinion, is best served by well-defined tasks and criteria to comment on. If these are absent, the observing is not instructive and may even become boring.

Writers in the present study were informed about the quality of their text by observing readers thinking aloud. One may raise the question that requiring readers to think aloud may result in thinking and reading processes that are artificial, to some extent. Did writers in this study get an accurate or artificial insight in reading processes and reader and text interactions? Ericsson and Simon (1993) argued that concurrent verbal reports need not affect the processes being studied and can be collected in ways that avoid reconstructions or interpretations on the part of the participants. When no specific information is requested different from normal situations, non-directive encouragement by research assistants, and tasks that elicit thinking processes in verbal form, concurrent protocols provided the best insight into what is going on with writers, readers, and thinkers when solving a problem.

The ultimate confrontation with readers is when writers see readers' processes "live", in a face-to-face confrontation. However, a live meeting has disadvantages. First, the explaining student cannot "look back" as he can do when reading written comments, protocols, or tapes. During the mentally demanding task of giving explanations, it is very hard for the student to be self-critical; the generating and the criticizing would easily interfere. Second, it is difficult for the explaining student to not interfere with his listening partner. Very short interruptions in the understanding task (e.g., if the listening student asks for a simple clarification) could lead to a change in the nature of the listener's task: he/she will probably step back from the naturalistic understanding task and start fulfilling the role of evaluator or even helper.

The mode of communication we have chosen in this experiment is a very direct, non-interactive, interruptible, and multi-mode (sound and vision) method. It seems to be a good compromise between the advantages and disadvantages. It would be interesting to experiment with one writing/reading task and several modes of reader–writer communication.

Future research should clarify some issues raised by the present study regarding other combinations of conditions. One issue is the separate effect of observation activities and written comments. In the present study, no condition was implemented where writers only received written feedback. Another issue is the quality of pre-writing activities. In this study, pre-writing activities were individual activities, while one may expect that building genre-awareness is a socially stimulated process. Implementing classroom discussions on quality of texts, stimulating a research attitude to communication, re-creating the rules of a genre in a classroom environment, stimulated by a creative teacher, is a step we have undertaken recently (Rijlaarsdam, 2003; Rijlaarsdam & Braaksma, 2004).

We have already continued this line of research with a study of the effects of peer observation by writers of argumentative texts (Couzijn, 1999; Couzijn & Rijlaarsdam, 2004). With regards to the writing of argumentative texts in the context of science education, Driver, Newton, and Osborne (2000) found that observation of writers had a large effect on students' writing and reading skills. Relatively weak writers seem to profit from observing instead of writing when new tasks are introduced (Braaksma et al., 2002). Observation of writing processes not only affects writing skill (in terms of text quality), but also writing processes (Braaksma et al., 2004). Apparently observation tasks affect metacognitive awareness (Braaksma et al., 2001; Zimmerman, 2000). Introducing writing tasks in science classrooms, as part of the enculturation in the world of science, could inspire teachers to create a double inquiry agenda for their students: to learn about science and to learn about genres and writing. Observation of real readers and writers at work supports both disciplines.

Acknowledgements

This article is based on a paper presented by the first two authors at the international conference on Ontological, Epistemological, Linguistic and Pedagogical Considerations of Language and Science Literacy: Empowering Research and Informing Instruction, Victoria, B.C., Canada, 12–15 September. The authors thank the organizers for the opportunity to leave their research discipline—effective teaching of language and literature—and to mingle in all sorts of literacy and hard sciences. They learnt that terms and concepts ("explaining") in different sciences have different connotations. Happily, Brian Hand (USA) and Larry Yore (Canada) guided us with their comments on previous drafts on our way to embed the literacy jargon into the world of the science educators. Shari Yore's meticulous editing practice taught us much about how to formulate crisp and clear text. The authors also wish to thank David Galbraith (Staffordshire, UK) for his inspiring discussions on writing and learning. Parts of this paper rely on the authors' collaborative work.

References

Alamargot, D., & Chanquoy, L. (2001). *Studies in writing: Vol. 9. Through the models of writing.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing to learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*(1), 29–58.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72,* 593–604.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bizzell, P. (1982). Composing processes: An overview. In A. R. Petrosky & D. Bartholoae (Eds.), *The teaching of writing: The eighty-fifth yearbook of the national society for the study of education* (pp. 49–70). Chicago: University of Chicago Press.

Bonset, H., & Rijlaarsdam, G. (2004). Mother tongue education (L1) in the learning to learn paradigm: Creative redevelopment of learning materials. *L1-Educational Studies in Language and Literature, 4,* 35–62.

Braaksma, M. A. H., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94*(2), 405–415.

Braaksma, M. A. H., Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B. H. A. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction, 22*(1), 1–36.

Braaksma, M. A. H., Van den Bergh, H., Rijlaarsdam, G., & Couzijn, M. (2001). Effective learning activities in observation tasks when learning to write and read argumentative texts. *European Journal of Psychology of Education, 1,* 33–48.

Brandt, D. (1992). The cognitive as the social: An ethnomethodological approach to writing process research. *Written Communication, 9,* 315–355.

Brillhart, L. V., & Debs, M. B. (1981). Teaching writing—A scientist's responsibility. *Journal of College Science Teaching, 10*(5), 303–304.

Chi, M. T. H., & Van Lehn, K. A. (1991). The content of physics self-explanations. *Journal of the Learning Sciences, 1,* 69–105

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on learning and transfer. *Learning and Instruction, 9,* 109–142.

Couzijn, M., & Rijlaarsdam, G. (2004). Learning to read and write argumentative text by observation. In G. Rijlaarsdam (Series Ed.), G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (Vol. Eds.), *Studies in writing. Vol. 14. Effective learning and teaching of writing* (2nd ed.), *Part 1, Studies in learning to write* (pp. 241–258). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84,* 287–312.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis. Verbal reports as data* (rev. ed.). Cambridge, MA: The MIT Press.

Flower, L. S., & Hayes, J. R. (1980). The dynamic of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds), *Cognitive processes in writing* (pp. 31–50). Hillsdale, NJ: Lawrence Erlbaum Associates.

Galbraith, D. (1996). Self-monitoring, discovery through writing and individual differences in drafting strategy. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 121–141). Amsterdam: University Press.

Galbraith, D., & Rijlaarsdam, G. (1999). Effective strategies for the teaching and learning of writing. *Learning and Instruction, 2,* 93–108.

Gallaghan, M., Knapp, P., & Noble, G. (1993). Genre in practice. In B. Cope & M. Kalantzis (Eds.), *The powers of literacy: A genre approach to teaching writing* (pp. 179–202). Pittsburgh, PA: University of Pittsburgh Press.

Goldman, S. R., & Bisanz, G. L. (2002). Toward a functional analysis of scientific genres: Implications for understanding and learning processes. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 19–50). Mahwah, NJ: Erlbaum.

Gratz, R. K. (1990). Improving lab report quality by model analysis, peer review and revision. *Journal of College Science Teaching, 19*(5), 292–295.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Mahwah, NJ: Lawrence Erlbaum Associates.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3–30). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science.* New York: Free Press.

Hillocks, G. (1986). *Research in written composition: New directions for teaching.* Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills and National Conference on Research in English.
Holliway, D. R., & McCutchen, D. (2004). Audience perspective in young writers composing and revising: Reading as the reader. In G. Rijlaarsdam (Series Ed.), L. Allal, L. Chanquoy, & P. Largy (Vol. Eds.), *Studies in writing. Vol. 13. Revision: Cognitive and instructional processes* (pp. 87–101). Dordrecht, The Netherlands: Kluwer Academic Publishers.
King, A. (1997). ASK to THINK-TEL WHY®©: A model of transactive peer tutoring for scaffolding higher level complex learning. *Educational Psychologist, 32*(4), 221–235.
King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90*(1), 134–152.
Klein, P. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11*(3), 203–270.
Linn, M. C., & His, S. (2000). *Computers, teachers, peers: Science learning partners.* Mahwah, NJ: Erlbaum.
Nystrand, M. (1989). A social-interactive model of writing. *Written Communication, 6,* 66–85.
Ocel, J. J., Palmer, B. A., Wittich, C. M., Carmichael, S. W., & Pawlina, W. (2003). Outcomes of the gross and developmental anatomy teaching assistant experience. *Clinical* Anatomy, *16*(6), 526–530.
Prain, V. (2002, September). *Learning from writing in secondary science: Some theoretical implications.* Paper presented at the international conference on Ontological, Epistemological, Linguistic and Pedagogical Considerations of Language and Science Literacy: Empowering Research and Informing Instruction, Victoria, B.C., Canada.
Princeton University. (2002). *Undergraduate engineering council's "survival tips".* Retrieved July 31, 2002, from www.princeton.edu.
Rijlaarsdam, G. (2003, August). *Studying processes in writing and reading literary fiction: Outcomes for researchers and learners.* Keynote address at the 10th European conference for Research on Learning and Instruction. Padua, Italy.
Rijlaarsdam, G., & Braaksma, M. (2004). Schrijven en leren schrijven. Niet zelf doen, maar zien hoe anderen het doen [Writing and learning-to-write. Learning from observing others]. *Levende Talen Magazine, 3,* 17–22.
Rijlaarsdam, G., Couzijn, M., & Van den Bergh, H. (2004). The study of revision as a writing process and as a learning-to-write process. Revision in the context of different drafting strategies. In G. Rijlaarsdam (Series Ed.), L. Allal, L. Chanquoy, & P. Largy (Vol. Eds.), *Studies in writing. Vol. 13. Revision: Cognitive and instructional processes* (pp. 189–207). Dordrecht, The Netherlands: Kluwer Academic Publishers.
Rijlaarsdam, G., & Van den Bergh, H. (2004). Effective learning and teaching of writing: Student involvement in the teaching of writing. In G. Rijlaarsdam (Series Ed.), G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (Vol. Eds.), *Studies in writing. Vol. 14. Effective learning and teaching in writing* (2nd ed., pp. 1–16). Dordrecht, The Netherlands: Kluwer Academic Publishers.
Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 778–803). New York: McMillan.
Schriver, K. A. (1991). Plain language through protocol-aided revision. In E. R. Steinberg (Ed.), *Plain language: Principles and practice* (pp. 148–172). Detroit, MI: Wayne State University Press.
Schriver, K. A. (1992). Teaching writing to anticipate readers' needs: A classroom-evaluated pedagogy. *Written Communication, 9*(2), 179–208.
Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill. Cognitive science series.* Cambridge, MA: Harvard University Press.
Sonnenschein, S., & Whitehurst, G. J. (1983). Training referential communication skills: The limits of success. *Journal of Experimental Child Psychology, 35,* 426–436.

Sonnenschein, S., & Whitehurst, G. J. (1984). Developing referential communication: A hierarchy of skills. *Child Development, 55,* 1936–1945.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.

Tulving, E. (1972). Semantic and episodic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press.

Unsworth, L. (2001). *Teaching multi-literacies across the curriculum: Changing contexts of text and image in classroom practice.* New York: Open University Press.

Yore, L. D., Bisanz, G. L., & Hand, B. M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education, 25*(4), 689–725.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

Appendix A. Communicative utterances in an "optimal manual"

T = Theory; I = Instruction; O = Object to use (selection); P = Precaution

<u>Introduction</u>

1.	(T)	The experiment is about air taking up space.
2.	(I)	Notice the initial situation on the table.

<u>Construction episode</u>

3.	(O)	Which bottle should be used?
4.	(O)	Which cork?
5.	(I)	Put the cork on the bottle.
6.	(O)	Which funnel?
7.	(I)	Put the funnel into the cork's hole.
8.	(P)	Push everything tightly together.
9.	(T)	because no air may escape.

<u>Water episode</u>

10.	(I)	Put water into the funnel.
11.	(I)	Check if the water stays in.
12.	(T)	that is because the air in the bottle stops it.
13.	(P)	If it does not stay in, press everything together more tightly.
14.	(T)	that is because air is leaking somewhere.

<u>Straw episode</u>

15.	(O)	Which straw?
16.	(I)	Put the straw into the funnel's hole.
17.	(P)	Make sure the straw's one end is pushed into the bottle.

18. (P) The straw's end mustn't be put into the water.
19. (T) The straw is needed to remove air from the bottle.
20. (I) Check if the water is running into the bottle now.
21. (T) that is because air can go out now.
22. (P) If the water doesn't run, blow, suck or move the straw;
23. (T) that is because the water in the straw must be removed.
24. (I) Hold your fingertip on top of the straw.
25. (I) Check if the water stops flowing,
26. (T) that is because air can't escape any longer.
27. (I) It will go on flowing if you lift your finger again.

<u>Conclusion</u>

28. (T) This experiment proves that air takes up space,
29. (T) because the water will not go into the bottle as long as the air cannot go out.

First, readers want to know what they can expect. Therefore, in the Introduction, information should be given related to the *subject* or *aim* of the experiment; and the reader should be prompted to *familiar* himself/herself with the initial situation with the many objects on the table before him/her.

In the second episode, a construction must be made with only some of the objects. Unambiguous descriptions are of great value to the reader, who would otherwise get lost in construction problems (not knowing what the intended construction looks like). The reader should be warned to make a firm assembly; if not, the construction will leak.

A first observation is done in the Water episode, when the water, counter-intuitively, does not run into the bottle. A reader could not be alarmed by the water running in the bottle due to a leak in the construction, so a preventive warning is required. Also, readers probably want to understand this unusual phenomenon, so an account must be given. Further, a hint is helpful here on how to correct an undesirable situation—and an account for that situation is given.

In the Straw episode the relation is observed between the escaping air and flowing water. Readers are likely to run into several problems here, which can be overcome by correct descriptions, precautions and explanations.

Finally, a concluding part should supply the reader with some information on what inference to make from the observations, because we do not expect 15-year-old students to make such an inference themselves.

Appendix B. Scoring rubric used for categorization of pieces of advice

A. Process-oriented advice	A.1 Orientation on explaining		*`start with doing the experiment yourself'* *`first examine all the objects on the table'*
	A.2 Generating explanations		*`write down all you can think of'* *`while you write, repeat the experiment in your mind'*
	A.3 Checking your explanation		*`re-read the text when you are done'* *`finally, check if your little sister understands the text'*
B. Product-oriented advice	*B.1 Style-oriented*	B.1.1 - Clarity	*`what you write must be very clear'*
		B.1.2 - Length	*`keep the text as short as possible'*
		B.1.3 - Accuracy	*`don't forget small details'*
		B.1.4 - Completeness	*`make sure you mention all the objects'*
		B.1.5 - Correctness	*`beware of mistakes'*
		B.1.6 - Organisation	*'give the instructions step by step'*
		B.1.7 - Accent	*`pay special attention to the theory'*
		B.1.8 - Audience	*`someone your age must understand it'*
		B.1.9 - Goal-directed	*`don't over-emphas the details'*
		B.1.10 - Spelling	*`check for spelling errors'*
		B.1.11 - Other	*`make a drawing if you want'*
	B.2 Content	B.2.1 - Instructions	*`write down everything the reader must do'*
		B.2.2 - Objects	*`also which tools you should use'*
		B.2.3 - Theory	*`and what the whole thing is about'*
		B.2.4 - Precautions	*`if things go wrong, what to do'*
		B.2.5 - Other	*`tell them to clean up afterwards'*