

Improving Written Communication Through Minimal Feedback

Matthew J. Traxler and Morton Ann Gernsbacher

*Department of Psychology, University of Oregon, Eugene, Oregon,
U.S.A.*

We propose that writers must form accurate representations of how their readers will interpret their texts to convey their ideas successfully. In two experiments, we investigated whether getting feedback from their readers helps writers form better representations of how their texts are interpreted. In our first experiment, one group of subjects (writers) wrote descriptions of a set of geometric figures; another group of subjects (readers) read those descriptions and used them to select the figures from sets of similar looking distractor figures. Half the writers received feedback on how well their readers selected the figures, and half the writers did not receive this feedback. Those writers who received feedback improved their descriptions more than those writers who did not receive feedback. In our second experiment, half the writers received two treatments of feedback on their descriptions of one set of figures, whereas the other half of the writers did not receive feedback. Then, all the writers described a new set of figures. Those writers who had previously received feedback wrote better new descriptions than did those writers who had never received feedback. We concluded that feedback – even this minimal form of feedback – helps writers to envision how readers interpret their texts.

INTRODUCTION

Humans communicate through various media: linguistic codes (both spoken and written language), non-linguistic codes (such as mathematical equations), non-verbal gestural codes (such as facial expressions) and

Requests for reprints should be addressed to Matthew J. Traxler, Department of Psychology, University of Oregon, Eugene, Oregon 97403-1227, U.S.A.

This research was supported by NIH Research Career Development Award KO4 NS-01376 and Air Force Office of Sponsored Research Grants 89-0258 and 89-0305 (all awarded to the second author). We thank Caroline Bolliger, Sara Glover, Maureen Marron, Jeff McBride, Rachel R. W. Robertson and Jon Shaw for help in testing our subjects; Kevin Kono, Cindy Ly and Mike Stickler for help in transcribing the written descriptions; and Daniel Kimble, Jennifer Freyd, Michael Posner, Peter Wason and an anonymous reviewer for their feedback.

more. Communication is successful when the receiver creates a mental representation of the message that matches the sender's mental representation of what he or she intended to convey. Communication is less successful whenever the sender and receiver fail to establish common mental representations.

Communication may be unsuccessful for many reasons: the sender might not encode the message properly; the channel might not accommodate the type of information the sender is trying to convey; or the receiver might not decode the message properly. Written communication seems particularly prone to failure – especially when compared with spoken communication. Empirical data overwhelmingly document that primary, secondary and university students have difficulty expressing themselves in writing (Bartlett, 1981; Bridwell, 1980; Hayes et al., 1987; Stallard, 1974). Even professional writers, with years of experience, often fail to produce texts that their intended audience can easily understand (Duffy, Curran & Sass, 1983; Swaney, Janik, Bond & Hayes, 1981).

Why do educated writers fail to produce understandable texts? If the messages are written legibly, the channel is not at fault. And if we assume that college-aged receivers are just as good at listening as they are at reading (as Gernsbacher, Varner & Faust, 1990, have demonstrated), receivers' ability to decode the message is not at fault. Instead, we propose that a major source of failure is the ability of senders to encode written messages.

We propose that educated writers are often unable to encode their messages properly because they fail to assess how their texts will be decoded (interpreted) by their readers. This is not to say that writers do not appraise their texts; skilled writers continually evaluate whether their words convey their intentions (Hayes, 1988; Scardamalia & Bereiter, 1983). Indeed, Sommers (1980) suggests that writers simultaneously refer to two mental representations while writing and revising. One representation is of the message they want to convey; the other representation is of the text as they have written it. If writers perceive a mismatch between the two representations, they revise.

We propose that comparing only these two representations is insufficient. Writers must also form and compare a third mental representation – the representation that their readers are likely to form when comprehending the text. Furthermore, we propose that forming this third representation is difficult, because it requires taking a naive perspective. If writers already know what they want to convey, they have already formed the mental representation that they want their readers to form. Forming it again – from their readers' perspective – is difficult.

Our proposal that writers often have difficulty forming a “naive representation” is supported by the following finding: Writers are worse at

detecting problems in their own texts than they are at detecting problems in other people's texts. For instance, fifth-grade writers can detect approximately half the problems detected by their teachers in texts *written by other students*; however, they detect only 10% of those problems in *their own texts* (Bartlett, 1981). Freshman college writers are also relatively unsuccessful at detecting problems in their own texts; they tend to focus on sentence-level problems while ignoring equally serious problems at other levels, such as a lack of organisation and focus. Even the sentence-level problems they detect are few and minor (Hayes, 1988; Hayes et al., 1987).

Let us return to our comparison with spoken communication to discover why writers are so poor at detecting problems in their own texts. Detecting problems in spoken communication (i.e. conversation) is considerably easier because conversation is collaborative, i.e. speakers and listeners actively interact – they collaborate – to establish mutual representations (Clark & Schaeffer, 1987a; 1987b; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987; Schober & Clark, 1989). The collaboration between speakers and listeners allows listeners to request clarification when they do not understand speakers. Likewise, speakers can solicit responses from listeners to determine when they are not being understood. In this way, speakers realise when their listeners' mental representations do not match what they intended to convey.

Unfortunately, writers and their readers (unlike speakers and their listeners) do not enjoy the luxury of collaborating to establish mutual representations. Readers are typically absent when writers encode their messages, and writers are typically absent when readers decode those messages. However, if writers received feedback from their readers, it might help them better understand how their messages are interpreted. In other words, it might help them envision the mental representation that their readers develop when comprehending their texts. Furthermore, this feedback might help writers improve subsequent messages; it might teach them to envision the mental representation that their readers are likely to develop when comprehending other texts. We tested these two hypotheses in the two experiments we report here.

In both experiments, we investigated whether providing writers with a minimal amount of feedback from their readers would improve written communication. One previous study suggested that an elaborate form of reader-supplied feedback does improve written communication. Swaney et al. (1981) attempted to revise a text that four professional document designers were unable to improve (as measured by readers' performance on comprehensive questions). Swaney et al. (1981) revised the text with the help of verbal "think-aloud" protocols. Think-aloud protocols are the comments readers make when instructed to "think aloud" while reading a text (Ericsson & Simon, 1980; Olson, Duffy & Mack, 1984). The "think

aloud” protocols that Swaney et al. (1981) used to revise the text were produced by an independent group of readers. The results of a comprehension test administered to another group of readers showed that the document revised according to the think-aloud protocols was more understandable than the original version.

Swaney et al.’s (1981) results suggest that feedback in the form of readers’ think-aloud protocols improves written communication, maybe because they give writers a better sense of readers’ mental representations. Note, however, that Swaney et al. (1981) used readers’ think-aloud protocols to revise a text that was originally written *by someone else*. Thus, the researchers had an outsider’s perspective when they revised the text, and this outsider’s perspective, along with the think-aloud protocols, could have been what led to the improvement. Furthermore, feedback in the form of readers’ think-aloud protocols is very elaborate (and “expensive” to collect). In our two experiments, we evaluated the effects of a more minimal form of feedback, and we evaluated the effects of this minimal feedback on the original writers.

EXPERIMENT 1

In our first experiment, we asked one group of university students, whom we called *writers*, to write descriptions of several Tangram figures. The Tangram figures were based on those used by Clark and his colleagues (e.g. Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989). They were solid black, geometric shapes, as shown in Fig. 1. Each writer wrote descriptions of one set of eight figures, either set A or set B. After the writers described the eight figures, another group of university students, whom we called *readers*, read each description and tried to select each “target” figure from three distractor figures. An example target figure and its three distractors are shown in Fig. 2. Each reader read descriptions written by two writers – one who would subsequently receive feedback and one who would not. Thus, each reader contributed data to both the feedback and no-feedback conditions. Furthermore, of the two writers whose descriptions each reader read, one writer had described the figures in set A and the other had described the figures in set B.

Thus, there was a first writing session, during which writers wrote descriptions of eight figures, and a first reading session, during which readers read the eight descriptions and selected each target figure from its distractors. The events of these two sessions are summarized in Fig. 3. Performance during the first reading and writing sessions provided a baseline from which we could compare performance during two more writing and reading sessions.

The following week, a second writing and reading session occurred. At

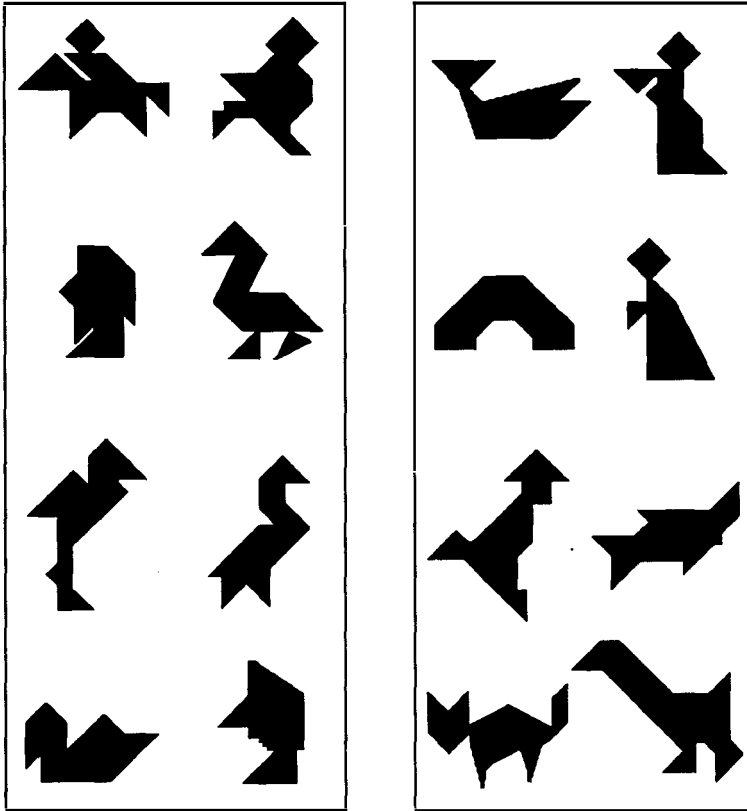


FIG. 1 Experimental stimuli: Eight target figures in sets A (left) and B (right).

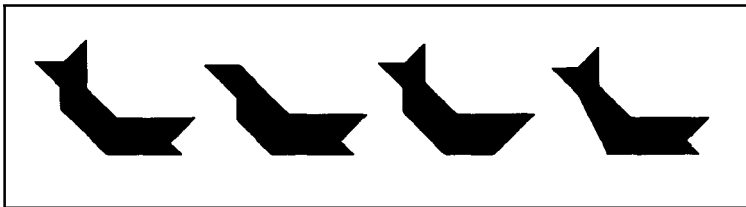


FIG. 2 Example target figure and its three distractors.

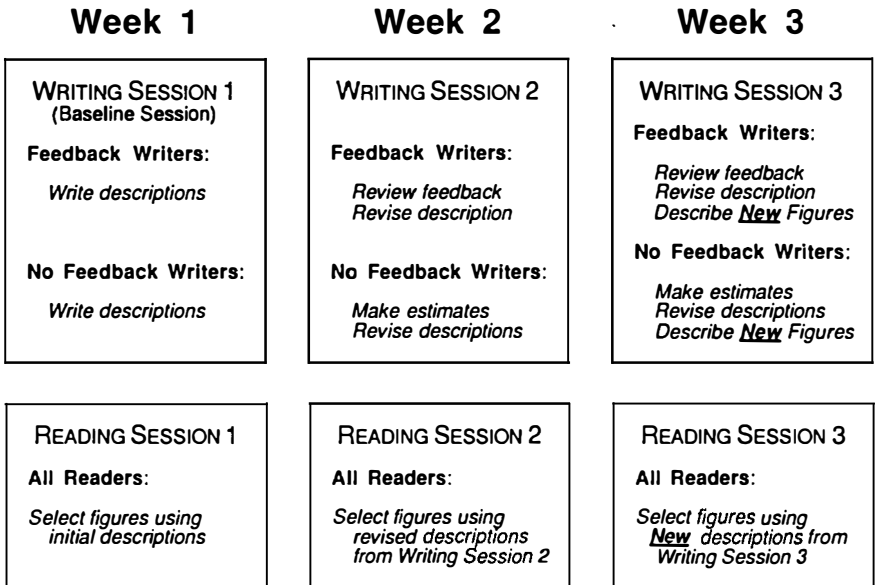


FIG. 3 Summary of events during Experiment 1.

the beginning of the second writing session, half the writers received feedback on how well their readers had used their initial descriptions to select the figures, and half of the writers did not receive this feedback. The feedback was simply a number for each of the eight figures indicating how many readers (none, one, or both) were able to select that figure from its distractors successfully. During the same time that the writers who received feedback were evaluating their feedback, the writers who did not receive feedback estimated how many of their readers selected the correct figure. So, both groups of writers spent the same amount of time reviewing the figures and their descriptions of those figures. Then, both groups of writers revised their descriptions. Later that week, the same readers read the revised descriptions and again tried to select each target figure. The events of these second writing and reading sessions are summarised in Fig. 3.

The next week, a third writing and reading session occurred. At the beginning of the third writing session, the writers who had previously received feedback again received feedback: They were told how well their readers had used their revised descriptions to select the figures during the second reading session, and they were reminded how well their readers had used their initial descriptions to select the figures during the first reading session. The other half of the writers did not receive feedback, but they

again performed the estimation task. Then, both groups of writers again revised their descriptions. In a final reading session, readers read these (re-)revised descriptions.

Thus, there were three writing and reading sessions. Our dependent measure was how many figures the readers correctly selected during each of the three reading sessions. As mentioned before, performance during the first sessions provided a baseline against which we could compare performance during subsequent sessions. The first writing and reading sessions provided a baseline because at this point none of the writers had received any feedback and none of the readers had seen any figures. The change in performance between the baseline and the subsequent two sessions illustrated the effect of feedback on the writers' ability to revise their descriptions.

We predicted that feedback would improve the descriptions because feedback should help writers envision the mental representations that their readers formed. If, while writing and revising their descriptions, writers hold a representation of the information they want to convey, a representation of what they have actually written and a representation of how their text will be interpreted by their readers, then providing writers with this type of feedback should allow them to form better representations of their readers' interpretations. In other words, feedback – even this minimal form of feedback – should help writers envision the mental representations that their readers form. Writers who do not receive feedback should be disadvantaged in this respect.

Methods

Subjects. A total of 32 undergraduates undertaking introductory psychology courses at the University of Oregon participated to fulfil a course requirement. Most of them were in their first or second year at the university. Sixteen subjects were randomly assigned to be writers, and 16 were randomly assigned to be readers.

Materials. The experimental materials comprised two sets of eight target Tangram figures, based on those used by Clark and his colleagues (Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989). As illustrated in Fig. 1, no two target figures within a set were identical, and no target figure appeared in more than one set. Half the writers described set A and half described set B. Pilot data suggested that the target figures in the two sets were equally difficult to describe. Each target figure was mounted on a 4 × 6 inch card, and the eight target figures in each set were placed in a three-ring binder.

The experimental materials also included distractor figures (that were

presented to the readers during the selection task). Three distractor figures were created for each target figure by making slight alterations to each target figure, as illustrated in Fig. 2. Pilot data suggested that across the two sets of target figures (A and B), the distractor figures were roughly equal in discriminability from their target figure. The target figures and their distractors were each mounted on a 4×6 inch card and placed in a three-ring binder. Within each binder, the figures were arranged so that each group of four figures contained a random arrangement of one target figure and its three distractors. The groups of four (a target figure and its three distractors) appeared in a different order than the order in which the writers viewed the target figures (and the order in which the writers' descriptions were presented to the readers). Each binder also included two more groups of four figures; in these two groups, all four figures were very similar looking, and all of them were distractors. These two additional groups of four distractors prevented readers from using the process of elimination to determine which group of four contained the target figure described last. Thus, the binders used for the readers' selection task contained 40 figures, arranged as 10 groups of 4 (8 groups of 4 contained a target figure and its 3 distractors, and 2 groups of 4 contained only distractors).

Design. The experiment involved three sessions separated by 1-week intervals, as summarised in Fig. 3. During the entire experiment, each writer was yoked with two readers. Similarly, each reader was yoked with two writers. As a result, we measured two readers' selection performance for each writer, and each reader read descriptions written by two different writers (one who received feedback and one who did not). We assigned two readers to each writer to increase the reliability of the feedback that the writers received (i.e. the data from the readers' task) and the reliability of our dependent measure.

Procedure: Writing Session 1. The writers read instructions telling them to "describe each of the eight geometric figures so thoroughly that another person reading your descriptions would be able to select each target figure from a group of very similar looking distractor figures". The writers were told that two other subjects would actually read the descriptions they wrote, and that these two "readers" would have the task of selecting the figures from distractors using only the writers' descriptions. To help the writers envision the readers' task, the writers were shown an example target figure and its three matching distractors (e.g. Fig. 2).

The writers were given a binder containing the eight figures they were to describe, a packet of eight blank 8.5×11 inch ruled pages, and their choice of pen or pencil. Although the eight figures and eight blank pages

were numbered "Figure 1" through "Figure 8", the writers were allowed to describe the figures in any order and return to previously written descriptions. The writers were given neither a time limit nor a minimum or maximum length requirement. They were told, however, that they would have to remain in the experiment room until everyone in the session finished.

The writers' handwritten descriptions were typed into a computer. The typists corrected only spelling errors. Errors of grammar, punctuation, capitalisation and so forth were left uncorrected. Although all the writers were treated identically during writing session 1, we randomly assigned each writer to one of two treatment groups, feedback or no-feedback.

Procedure: Reading Session 1. The readers read instructions that said that their task would be "to read descriptions of geometric figures and to select each geometric figure from a set of very similar looking distractors". The readers were told that each target figure would be accompanied by several very similar looking distractors, and that they should examine all the figures before making their selection.

Each reader read two sets of eight descriptions: one set written by a writer who would subsequently receive feedback and one set written by a writer who would not receive feedback. Thus, each reader contributed data to both the feedback and the no-feedback conditions. Half the readers read descriptions written by a feedback writer first, and the other half read descriptions written by a no-feedback writer first. Furthermore, each reader was assigned to two writers such that one writer had described the eight figures in set A and the other writer had described the eight figures in set B. Thus, no reader read more than one description of each figure. Half the readers read descriptions of the figures in set A first, and half read descriptions of the figures in set B first. The readers read typewritten copies of the descriptions and selected the figures from a binder containing 40 figures. Each figure was labelled with a number.

The readers read each description and searched through the binder for the figure that they judged to fit the description best. They were told that the figures were arranged in the binders in groups of four, and that eight of the four-figure groups contained a target figure and three distractors, but two groups contained four distractors. Thus, the readers were faced with a two-tiered task. On a first pass, they needed to select the group of four figures, all of which could possibly fit the description; on a second (and more time-consuming) pass, they needed to select from the group of four the one figure that best fit the description.

After reading and selecting figures for each of the eight descriptions written by one writer, the readers took a short break. After the break, the readers were given another three-ring binder containing the other set of 40

figures, and the typed descriptions written by the second writer to whom they were assigned. Again, they read each description, selected a figure, and recorded their response by writing down the number of the figure. The readers were told that there was no time limit for selecting each figure, but that they would have 1 hour to read and select the figures for two sets of eight descriptions. They were also told that they would have to remain in the experiment room until everyone in the session finished.

Procedure: Writing Session 2. At the beginning of writing session 2, all the writers received the three-ring binder that contained the eight figures that they had described during writing session 1. All the writers also received packets containing typed versions of the descriptions they wrote during writing session 1. The packets given to the feedback writers contained two sets of numbers (i.e. their feedback). One set of numbers, which appeared on the first page, indicated how many figures (out of eight) each of their two readers had correctly selected. The other set of numbers, which appeared at the top of the page on which each description was written, indicated how many readers (out of two) had correctly selected the figure that the description described. The feedback writers were told to review these numbers and then answer the following questions with one or two sentences:

1. How successful were your readers in selecting the figures?
2. Which description(s) did the readers understand the most?
3. Which description(s) did the readers understand the least?
4. What are the differences between the descriptions that the readers understood the most and the descriptions that they understood the least?

The packets given to the no-feedback writers also contained typed copies of the descriptions they wrote during writing session 1, but they did *not* contain the numbers summarising their readers' selection performance. Nevertheless, we wanted the no-feedback writers to spend the same amount of time evaluating their descriptions and re-examining the figures as the feedback writers did. Therefore, the no-feedback writers were told to review each description and estimate how many readers had correctly selected each figure. Thus, for each description, the no-feedback writers wrote down a number (0, 1 or 2) that indicated how many readers they thought had correctly selected that figure. Then, the no-feedback writers were told to review their estimates and answer the following questions with one or two sentences:

1. How successful do you think your readers were in selecting the figures?
2. Which description(s) do you think the readers understood the most?

3. Which description(s) do you think the readers understood the least?
4. What are the differences between the descriptions that you think the readers understood the most and the descriptions that you think they understood the least?

The no-feedback writers spent slightly more time estimating their readers' performance and answering the questions than the feedback writers spent evaluating their readers' actual performance and answering the questions.

After they had reviewed their descriptions and answered the questions, both the feedback and no-feedback writers revised their descriptions. The feedback writers were told to concentrate on the descriptions of figures that both readers failed to select correctly. They were told to spend less effort revising descriptions of figures that at least one reader had selected correctly, and to spend the least effort revising descriptions of figures that both readers had selected correctly. The no-feedback writers were told to distribute their effort similarly – concentrate on descriptions of figures that they predicted both readers had failed to select correctly, spend less effort revising descriptions of figures that they predicted at least one reader had selected correctly, and spend the least effort revising descriptions of figures that they predicted both readers had selected correctly. All the writers were told to rewrite their revised descriptions completely, even if they made no changes. We required the writers to rewrite all their descriptions so that they would be encouraged to make changes, as opposed to simply resubmitting the typed versions.

Procedure: Reading Session 2. The readers were told that they would again be reading descriptions of figures, and their task was again to select the described figures. They were told that the descriptions were written by the same writers who wrote the descriptions they read during reading session 1. The descriptions were presented in the same order during reading session 2 as they were presented during reading session 1. The readers read and selected figures for the eight descriptions written by one writer; they took a short break, and then they read and selected figures for the eight descriptions written by the other writer.

Procedure: Writing Session 3. At the beginning of writing session 3, all the writers received a three-ring binder that contained the eight figures they had previously described. All writers also received packets containing typed versions of the descriptions as they had revised them during writing session 2. The packets given to feedback writers contained four sets of numbers (i.e. their feedback). Two sets of numbers summarised their readers's performance during reading session 2, when the readers had used the writers' revised descriptions to select the target figures. More specifically, one set of numbers, which appeared on the first page, indicated how

many figures (out of eight) each of their two readers had correctly selected using the revised descriptions. A second set of numbers, appearing at the top of each description, indicated how many readers (out of two) had correctly selected the figure which that description described (again, using the revised descriptions). The remaining two sets of numbers reminded the writers of their readers' performance during reading session 1, when they read the *initial* descriptions. After the feedback writers had reviewed these numbers, they answered the following questions in one or two sentences:

1. How successful were your readers in selecting the figures after reading your revised descriptions?
2. How successful were your readers in selecting the figures after reading your revised descriptions compared with how successful they were after reading your initial descriptions?
3. What do you think you could do to help the readers do a better job of selecting the figures?

The packets given to the no-feedback writers also contained typed copies of their revised descriptions, but their packets did not contain the numbers summarising their readers' selection performance. Again we wanted the no-feedback writers to spend the same amount of time evaluating their revised descriptions and re-examining the figures as the feedback writers did. Therefore, the no-feedback writers were again told to estimate how many readers (out of two) had correctly selected each figure using their revised descriptions. Then, the no-feedback writers were told to review their estimates and answer the following questions with one or two sentences:

1. How successful do you think your readers were in selecting the figures?
2. How successful do you think your readers were in selecting the figures after reading your revised descriptions compared to how successful they were after reading your initial descriptions?
3. What do you think you could do to help the readers do a better job of selecting the figures?

Again, the no-feedback writers spent slightly more time estimating their readers' performance and answering the questions than the feedback writers spent evaluating their readers' actual performance (their feedback) and answering the questions. After answering the questions, both feedback and no-feedback writers once again revised their descriptions.

Procedure: Reading Session 3. The procedure followed in reading session 3 was identical to the procedure followed in reading session 2. However, the descriptions given to the readers were the descriptions that the writers had revised during writing session 3.

Results

If feedback helps writers envision the mental representations that their readers form, then feedback should have improved writers' ability to revise their descriptions. Our results supported our prediction. Figure 4 presents the readers' mean percent correct score on the selection task during the baseline (first) reading session, the second reading session and the third reading session. The filled squares represent the readers' performance when they selected figures using descriptions written by writers who received feedback; the unfilled squares represent the readers' performance when they selected figures using descriptions written by writers who did not receive feedback.

An analysis of variance (ANOVA) on these data showed a reliable main effect of session [baseline vs 2 vs 3: $F(2,14) = 13.67, P < 0.001$] and a reliable interaction between session and feedback [$F(2,14) = 3.80, P < 0.05$]. Further ANOVAs explored this interaction. As Fig. 4 illustrates, performance at the baseline session did not depend on whether the initial descriptions were written by writers who would or would not subsequently receive feedback ($F < 1$). This result ensures that our first writing and reading sessions were indeed baseline sessions. However, as Fig. 4 also illustrates, performance at the subsequent sessions did depend on whether the writers received feedback. More specifically, when the readers selected figures using descriptions written by writers who received feedback, the effect of session was reliable [$F(2,14) = 9.99, P < 0.002$]. In contrast, when the readers selected figures using descriptions written by writers who did not receive feedback, the effect of session was not reliable [$F(2,14) = 3.07, P < 0.08$].

Another way to view these results is to describe them in terms of improvement. Figure 5 presents improvement scores, which we computed

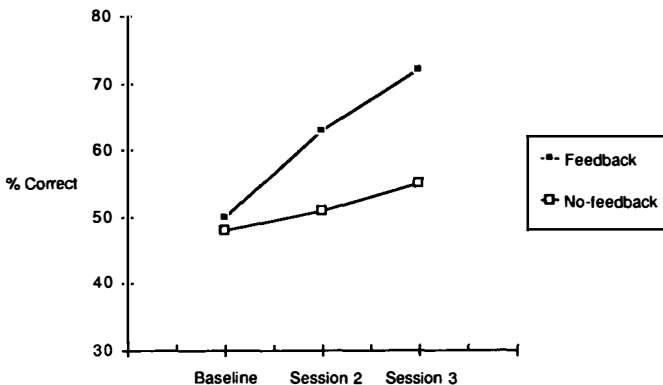


FIG. 4 Subjects' mean percent correct scores in the selection task in Experiment 1.

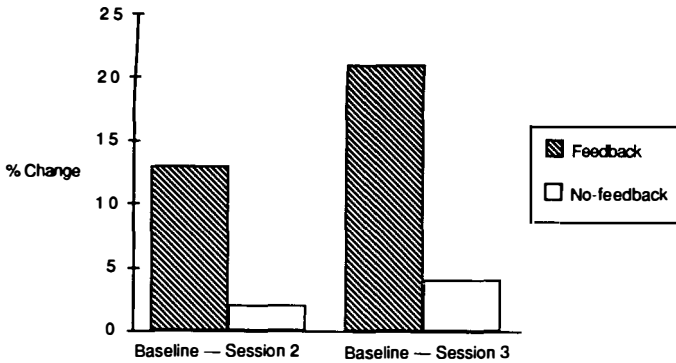


FIG. 5 Subjects' mean improvement in selection during Experiment 1.

by simply subtracting performance at the second and third reading sessions from performance at the first, baseline session. The hatched bars represent the readers' improvement when they selected figures using descriptions revised by writers who received feedback; the unfilled bars represent the readers' improvement when they selected figures using descriptions revised by writers who did not receive feedback.

First, consider the two bars on the left in Fig. 5: they represent the amount of improvement from the baseline session to the second reading session. As these two bars illustrate, the descriptions revised by writers who received feedback led to a reliable amount of improvement [$F(1,15) = 6.00, P < 0.03$]. In contrast, the descriptions revised by writers who did not receive feedback did not lead to a reliable amount of improvement ($F < 1$).

Now, consider the two bars on the right in Fig. 5: they represent the amount of improvement from the baseline to the third reading session. Recall that the descriptions used by the readers at the third reading session had been revised twice. Thus, those writers who had received feedback had received two treatments of feedback, and these two treatments should have led to even more improvement. And indeed, as the hatched bar illustrates, the descriptions re-revised by writers who received feedback did lead to a reliable amount of improvement over baseline performance [$F(1,15) = 21.00, P < 0.001$], even more improvement than we observed at the second reading session [after the writers received only one treatment of feedback: $F(1,15) = 5.00, P < 0.04$]. In contrast, consider the descriptions re-revised by writers who did not receive feedback (the rightmost, unfilled bar). Although the descriptions re-revised by writers who did not receive feedback did lead to a reliable amount of improvement at the third reading session [$F(1,15) = 6.32, P < 0.02$], the amount of improvement was reliably less than the amount of improvement provided by the feedback writers' descriptions [$F(1,15) = 7.90, P < 0.01$]. Furthermore, a linear

trend analysis of the readers' performance with the no-feedback descriptions was not reliable [$F(1,30) = 1.13, P < 0.25$], whereas a linear trend analysis of the readers' performance with the feedback descriptions was reliable [$F(1,30) = 21.88, P < 0.001$].

These results demonstrate that feedback – even minimal feedback provided by numbers representing readers' selection performance – improved university students' ability to revise written texts.

EXPERIMENT 2

Our goal in Experiment 2 was to investigate whether feedback would continue to improve university students' written communication when they faced a new writing task. If feedback enables writers to envision their readers' mental representations, then this improved perspective should continue – even when the writers describe novel stimuli. In contrast, if the beneficial effects of feedback that we observed in our first experiment were due solely to the writers becoming aware of specific problems in their previously written texts, then we should not observe any benefit of feedback when the writers describe novel stimuli.

During the first two writing and reading sessions of Experiment 2, we followed the same procedure that we followed in Experiment 1. During the first writing session, one group of subjects wrote descriptions of the eight Tangram figures in either set A or set B, and during the first reading session, another group of subjects read the descriptions and selected each target figure from its distractors. As in Experiment 1, performance during this first session provided a baseline. Also as in Experiment 1, at the beginning of the second writing session half the writers received feedback and half did not. Both groups revised their descriptions and, during the second reading session, readers read the revised descriptions.

The third writing session also began like our first experiment: Half the writers received feedback on their readers' success using their revised descriptions, whereas the other half only estimated their readers' success. Then, the critical difference between our first and second experiments occurred. To test whether the beneficial effects of feedback would transfer to a new writing task, all the writers were given a new set of figures to describe. If they had previously described the figures in set A, then their task was to describe the figures in set B; similarly, if they had previously described the figures in set B, then their task was to describe the figures in set A. Thus, the feedback writers had received feedback on only their descriptions of one set of figures; now their task was to describe a new set. If feedback enables writers to better envision the mental representations formed by their readers, then feedback should have improved the writers' descriptions of the new set of figures.

Methods

Subjects. A total of 88 undergraduates undertaking introductory psychology courses at the University of Oregon participated to fulfil a course requirement. Forty-four subjects were randomly assigned to be writers, and 44 were assigned to be readers.

Materials and Design. The materials used for Experiment 2 were identical to those used for Experiment 1. The design was identical also. Again, each writer was yoked with two readers, and each reader was yoked with two writers (one who received feedback and one who did not).

Procedure. The procedure for Experiment 2 was the same as the procedure for Experiment 1 during sessions 1 and 2. The only change occurred during writing session 3. During writing session 3, all the writers were given a *new* set of figures to describe. Writers who described set A during writing sessions 1 and 2 described set B during writing session 3; conversely, writers who described set B during writing sessions 1 and 2 described set A during writing session 3.

Results

If the beneficial effects of feedback that we observed in Experiment 1 were due solely to the writers becoming aware of specific problems in their previously written texts, than we should not have observed any benefit of feedback when the writers described novel stimuli. In contrast, if, as we propose, feedback enables writers to envision their readers' mental representations, then this improved perspective should have continued – even when the writers described novel stimuli.

Our results supported our prediction. Like Fig. 5, Fig. 6 presents improvement scores, which we again computed by subtracting performance during the second and third reading sessions from performance during the first, baseline session. The hatched bars represent the readers' improvement when they selected figures using descriptions written by writers who received feedback; the unfilled bars represent the readers' improvement when they selected figures using descriptions written by writers who did not receive feedback.

First, examine the two bars on the left in Fig. 6; they represent improvement from the baseline session to the second reading session. As these two bars illustrate, the descriptions revised by writers who received feedback led to a reliable amount of improvement [$F(1,42) = 9.53, P < 0.005$]. In contrast, the descriptions revised by writers who did not receive feedback

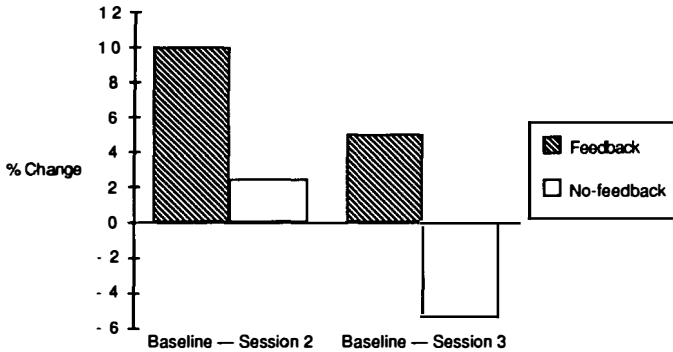


FIG. 6 Subjects' mean improvement in selection during Experiment 2.

did not lead to a reliable amount of improvement ($F < 1$). This pattern replicates Experiment 1.

The novel results of Experiment 2 are illustrated by the two rightmost bars in Fig. 6. Those two bars represent improvement from the baseline to the third reading session. Recall that the descriptions read during the third reading session were about novel stimuli; neither the feedback nor the no-feedback writers had ever received feedback on those particular descriptions. However, as Fig. 6 illustrates, the descriptions of the new stimuli that were produced by writers who had previously received feedback led to above baseline performance (49% correct); indeed, the amount of improvement was not reliably less than that observed at the second reading session [$F(1,42) = 1.40, P > 0.23$]. In contrast, the descriptions of the new stimuli produced by writers who had never received feedback led to below baseline performance (42% correct), and the amount of improvement was reliably lower than the amount of improvement provided by descriptions written by writers who had previously received feedback [$F(1,42) = 5.14, P < 0.03$].

These results demonstrate that the benefits of feedback – even minimal feedback provided by numbers representing readers' selection performance – transfer to a novel writing task.

CONCLUSIONS

Our two experiments demonstrate that providing university student writers with minimal feedback from their readers helps them improve their texts. In our first experiment, those writers who received feedback improved the texts on which they received feedback. In our second experiment, those writers who received feedback on one set of texts wrote better new texts.

We suggest that feedback gave the writers a better sense of how their texts were interpreted by their readers. Because our feedback only identified those texts which were problematic – it did not identify what the problems were or how they could be solved – the writers had to rely on internal information to improve their texts. We suggest that the internal information on which the writers relied was a mental representation of how their readers interpreted their texts.

Why did our minimal feedback encourage the writers to consult their mental representations? We know that writers frequently overestimate how clear their texts are, and they have difficulty pinpointing where their texts are unclear (Bartlett, 1981; Hayes, 1988; Hayes et al., 1987). In our experiments, the writers who did not receive feedback also overestimated how clear their texts were, and they unsuccessfully predicted which of their texts were less clear. More specifically, in Experiment 1, the writers who did not receive feedback estimated (on the average) that their readers had selected the correct figure 63% of the time using their initial descriptions, and 74% of the time using their revised descriptions. In reality, the readers were successful only 48% and 51% of the time respectively. Furthermore, in Experiment 1, the correlation between the no-feedback writers' predictions of how many readers would select each figure correctly and how many readers actually selected each figure correctly was almost zero.

Thus, the writers who did not receive feedback mis-estimated their readers' success. Therefore, our minimal feedback probably informed the writers (who received feedback) that they were not communicating as well as they thought they were. In other words, our minimal feedback gave the writers a better sense of how well their texts were interpreted by their readers.

Our minimal feedback also identified which texts were less clear. However, if the improvement engendered by our feedback was caused only by writers remediating their less successful texts, then the benefits of feedback would not have transferred to the new writing task (as it did in our second experiment). Instead, we suggest that the writers who received feedback were able to assess more accurately how well their texts communicated their intended message; they were able to compare which texts were more vs less successful, and then they took steps to improve their communication.

What modifications did those writers who received feedback make? One modification was to alter their writing style. Across all writers, we were able to identify four strategies that the writers used to describe the Tangram figures. The most popular strategy was a "looks like" strategy, which involved describing a figure with a visual analogy, such as "This figure looks like a duck". Another popular strategy was a "geometry" strategy, which involved describing a figure by identifying some of the

geometric shapes that composed it and some of the relationships among those shapes. For instance, one writer wrote: "This figure has a square on top, connected at one point to a larger parallelogram." Another strategy was "side-counting", which involved listing the number of sides or points contained in the figure (e.g. "This figure has fourteen sides"). And a fourth strategy was "etch-a-sketch", which involved telling the reader how to draw the figure (e.g. "Start at the top. Draw a line one inch long horizontally to the right"). Although through *post-hoc* analyses we were unable to identify which strategy was the most successful, we did observe that writers who received feedback were more likely to change their strategies on subsequent sessions.¹

A second way that writers who received feedback modified their descriptions was to make them denser. Surprisingly, those writers who received feedback did not consistently write longer descriptions than those writers who did not receive feedback [session 1: $F(1,14) = 1.05, P > 0.25$; session 2: $F(1,14) = 5.40, P < 0.04$; session 3: $F(1,14) = 2.69, P > 0.10$]. Rather, we suspect that those writers who received feedback wrote more detailed descriptions during subsequent sessions.

This more-detailed quality is suggested by the following phenomenon: Writers who received feedback and writers who did not receive feedback were equally successful at conveying the information the readers needed to at least select some figure from the correct group of four (i.e. the target figure and its three distractors). For example, in Experiment 1, when the readers read descriptions written by writers who received feedback, they chose a figure from the correct group of four 82% of the time; when the readers read descriptions written by writers who did not receive feedback, they chose a figure from the correct group of four 80% of the time. The difference between the writers who received feedback and the writers who did not receive feedback was their success at conveying the information readers needed to select the correct figure from the group of four (i.e.

¹Three independent judges examined the descriptions produced in our first experiment and classified each writer's predominant strategy or strategies for each session. A predominant strategy was one that a writer employed more than four times during a session. The writers could employ more than one strategy in a given session, and they often did. For instance, the writers commonly employed the "looks-like" strategy in conjunction with the "geometric" strategy. Once these predominant strategies were classified, we tallied the number of times that each writer changed his or her predominant strategy or strategies. For instance, if during session 1 a writer employed only the "looks-like" strategy, but during session 2 added the "geometry" strategy, we considered that writer as making one change. Or, if during session 1 a writer employed three strategies, but during session 2 he or she replaced one of those three strategies with another, we considered that writer as making two changes. Those writers who received feedback changed their strategies more often than writers who did not receive feedback [$F(1,14) = 4.87, P < 0.04$].

select the actual target figure from its similar distractors). This difference could emerge only if the writers who received feedback wrote more detailed descriptions.

Finally, perhaps feedback increased motivation. As we have mentioned before, writers who received feedback were told how well they were conveying their message, whereas writers who did not receive feedback had no objective basis on which to evaluate their performance, and they underestimated their readers' failure. Furthermore, those writers who did not receive feedback might have had less invested in the task because they were never confronted with their success or failure. Both a lack of personal investment and an underestimation of the task's difficulty might have led no-feedback writers to apply less effort.

However, differences in motivation (or effort) cannot fully account for our results because previous studies demonstrate that more effort does not guarantee better writing (Beach, 1979; Duffy et al., 1983; Hayes, 1988; Swaney et al., 1981). For instance, high school students who are motivated by between-draft evaluations from their teachers make more revisions on later drafts than do students who do not receive between-draft evaluations. If making more revisions is an indication of greater effort, then students who receive between-draft evaluations apply more effort. However, despite their greater effort, those students do not produce better texts than other students do (Beach, 1979). Although in our own experiments we cannot rule out motivational effects, we do not think differences in motivation (or effort) account fully for our results.

In future work, we shall ask the following questions: Do the beneficial effects of this minimal type of feedback improve university students' ability to write about a completely new domain? What if the writing task is no longer to describe abstract geometric shapes, but rather to describe abstract aromas (novel but describable scents)? Does feedback help practised university student writers more than it helps average university students' written communication? Perhaps practised university student writers (e.g. those majoring in journalism) have more experience developing the mental representations that their readers form; if so, they might better translate feedback into a mental representation.

Finally, we shall more specifically test the hypothesis that writers have difficulty communicating their ideas because they have difficulty forming a "naive representation". One prediction motivated by this hypothesis is that university student writers should detect problems in other writers' descriptions more successfully than they detect problems in their own descriptions, just as Bartlett's (1981) fifth-grade writers detected problems in other fifth-grade writers' texts more successfully than they detected problems in their own texts. Furthermore, if writers have difficulty communicating their ideas because they have difficulty taking their readers'

perspective, then a manipulation in which writers become readers (i.e. if they perform the selection task themselves on a novel set of stimuli using another writer's descriptions), they should cause writers to produce better written descriptions.

Testing these hypotheses should illuminate why written communication, compared with spoken communication, is particularly prone to failure. As we gather more information about why written communication often fails, we can further suggest how it can be improved.

Manuscript received March 1991

Revised manuscript received June 1991

REFERENCES

- Bartlett, E.J. (1981). *Learning to write: Some cognitive and linguistic components*. Washington, D.C.: Center for Applied Linguistics.
- Beach, R. (1979). The effects of between-draft teacher evaluation versus student self-evaluation on high-school students' revising of rough drafts. *Research in the Teaching of English*, 13, 111-119.
- Bridwell, L.S. (1980). Revising strategies in twelfth grade students: Transactional writing. *Research in the Teaching of English*, 14, 107-122.
- Clark, H.H. & Schaeffer, E.F. (1987a). Collaborating on contributions to conversation. *Language and Cognitive Processes*, 2, 19-41.
- Clark, H.H. & Schaeffer, E.F. (1987b). Concealing one's meaning from overhearers. *Journal of Memory and Language*, 26, 209-225.
- Clark, H.H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.
- Duffy, T., Curran, T. & Sass, D. (1983). Document design for technical job tasks: An evaluation. *Human Factors*, 25, 143-160.
- Ericsson, K.A. & Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- Gernsbacher, M.A., Varner, K.R. & Faust, M. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 430-445.
- Hayes, J.R. (1988). Writing research: The analysis of a very complex task. In D. Klahr & K. Kotovsky (Eds), *Complex information processing: The impact of Herbert A. Simon*, pp. 209-234. Hillsdale, N.J.: Lawrence Erlbaum Associates Inc.
- Hayes, J.R., Flower, L.S., Schriver, K.A., Stratman, J. & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in psycholinguistics. Vol. II: Reading, writing, and language processing*, pp. 176-240. Cambridge: Cambridge University Press.
- Isaacs, E.A. & Clark, H.H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116, 26-37.
- Olson, G.M., Duffy, S.A. & Mack, R.L. (1984). Thinking-out-loud as a method for studying real-time comprehension processes. In D.E. Kieras & M.A. Just (Eds), *New methods in reading comprehension research*, pp. 253-286. Hillsdale, N.J.: Lawrence Erlbaum Associates Inc.
- Scardamalia, M. & Bereiter, C. (1983). The development of evaluative, diagnostic, and

- remedial capabilities in children's composing. In M. Martlew (Ed.), *The psychology of written language: Developmental and educational perspectives*, 67–95. New York: John Wiley.
- Schober, M.F. & Clark, H.H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211–232.
- Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication*, 31, 378–387.
- Stallard, C. (1974). An analysis of the writing behavior of good student writers. *Research in the Teaching of English*, 8, 206–218.
- Swaney, J.H., Janik, C.J., Bond, S.J. & Hayes, J.R. (1981). *Editing for comprehension: Improving the process through reading protocols*. Technical Report No. 14, Document Design Project, Carnegie-Mellon University.