

# Writing: The Process of Discovery

**Veerle Baaijen (V.M.Baaijen@rug.nl)**

Center for Language and Cognition Groningen, University of Groningen  
Oude Kijk in 't Jatstraat 26, 9700 AS, Groningen, The Netherlands

**David Galbraith (D.Galbraith@staffs.ac.uk)**

Centre for Educational Psychology Research, Staffordshire University  
College Road, Stoke-on-Trent, ST4 2DE, United Kingdom

**Kees de Groot (C.M.de.Groot@rug.nl)**

Center for Language and Cognition Groningen, University of Groningen  
Oude Kijk in 't Jatstraat 26, 9700 AS, Groningen, The Netherlands

## Abstract

This paper describes the results of a study investigating the process by which writers develop their understanding through writing. It argues that, contrary to problem-solving models of writing, the crucial ingredient is implicitly guided text production. Two groups of writers, varying in the extent to which their writing is assumed to be directed towards rhetorical goals, were asked to write either planned or non-planned texts. Key-stroke logs were collected, and changes in subjective understanding about the topic were measured. The results show that developments of understanding are strongly related to the extent to which writers modify their texts during writing, and this is highest in the conditions expected to promote implicitly guided text production. We conclude that these findings support a dual-process model of writing.

**Keywords:** Planning; knowledge change; writing processes; keystroke logging; text production.

## Introduction

Writing is an ideal area in which to study the ebb and flow of thought. Although the end product is a fixed knowledge object which has to be comprehensible in the absence of the writer, the process by which it is produced is an extremely dynamic one, in which writers both have to work out what they think about a topic and how best to communicate this to their readers. For this reason, writing is typically characterized as a process of discovery. Bereiter and Scardamalia (1987), for example, characterize expert writing as a knowledge-transforming process, during which writers actively transform their thought in response to their evolving goals, and contrast this with the knowledge-telling process employed by novice writers, in which a fixed store of ideas in long term memory is translated directly into text. They claim that the knowledge-transforming model accounts for the “the peculiar value that many have claimed for writing as a way of developing one's understanding” (Bereiter & Scardamalia, 1987, p. 302). In this paper, we describe the results of an experiment investigating the conditions under which writers develop their understanding and how this is related to a simple indicator of one of the processes involved in writing.

According to Bereiter and Scardamalia's (1987) knowledge-transforming model, discovery through writing is a consequence of rhetorical problem solving. This claim has three important features. The first is an emphasis on the explicit thinking processes involved in the generation and evaluation of content rather than on the processes involved in translating thought into language. Second, and following from this emphasis, the crucial contrast between the knowledge telling and knowledge transforming processes is the goals toward which writing is directed. Thus, in the knowledge-telling model, the goal is to retrieve ideas stored in memory and translate these into text. By contrast, in the knowledge-transforming model, content retrieval and evaluation is mediated by the writer's communicative goals: expert writers develop an elaborate representation of their audience and the rhetorical situation and use this to guide the generation of content. This leads to the re-evaluation of existing content in long term memory and to the formulation of new content. Third, the extent to which writers are able to engage in this reflective evaluation of content depends on how they manage the interaction between high-level thinking processes and the formulation of content in text. Translating processes and higher level thinking processes are assumed to compete for limited cognitive resources, and hence it is assumed that writers will be less able to engage in rhetorical problem-solving the more they try to carry out text production at the same time as generating content. It is this conflict which is assumed to be responsible for the beneficial effects of outlining prior to writing. Kellogg (1988) has provided convincing evidence that outlining is associated with the production of better quality text, and that this is because it enables writers to clearly separate the reflective processes involved in generating, organizing and evaluating ideas from the processes involved in formulating these ideas in well-formed text.

Overall, the knowledge-transforming model and associated research on the benefits of outlining suggests that discovery through writing is a consequence of the strategic modification of content in order to satisfy rhetorical goals, and that this will be enhanced when the writer is able to focus on higher level thinking processes free from the demands of simultaneously formulating full text.

Recently, Galbraith (2009) has questioned this account on empirical grounds. In a series of experiments examining the conditions under which writers develop their understanding, Galbraith and his colleagues have found different patterns of development of understanding through writing than would be expected on the basis of the rhetorical problem-solving model. In brief summary, the essential pattern of their findings is as follows. First, although writers whose writing is assumed to be directed towards rhetorical goals (high self-monitors) do develop more new content after making notes than when they are required to write full text, as would be expected if discovery depended on the extent to which writing was directed towards rhetorical goals, this new content was not associated with increases in writers' subjective understanding of the topic. Second, there was also evidence that writers whose writing was assumed, not to be directed towards rhetorical goals, but rather to be implicitly organized (low self-monitors), developed new content after writing full text, without pre-planning, and that this was associated with developments of subjective understanding. (See Snyder and Gangestad, 1986, for a review of differences between low and high self-monitors.)

On the basis of these experiments, Galbraith has suggested an alternative dual-process account of discovery through writing. This proposes that the development of understanding in writing depends on an interaction between two different kinds of process. The first of these is an explicit planning process. This involves the retrieval of content from an explicit store of ideas and the goal-directed manipulation of these ideas in working memory designed to create a coherent knowledge object that satisfies rhetorical goals. This is essentially equivalent to the knowledge transforming model of Bereiter and Scardamalia, with the crucial difference that, by itself, this process only involves the reorganization of existing knowledge and is not associated with developments of understanding. The second is an implicit text production process. This operates on an implicit store of conceptual knowledge in semantic memory, which Galbraith defines as the writer's disposition towards the topic, and involves synthesizing content during text production. The key features of this process, for present purposes, are that it is engaged when writers have to formulate their thought in explicit propositions, and that, because the process is guided by the implicit organization of material in semantic memory, the sequence in which content is produced is unpredictable. Content is synthesized in the course of formulation rather than being directly retrieved from memory and translated into text. This process is assumed to lead to developments of understanding when the content it produces is different from the explicit content stored in episodic memory.

The model suggests that the implicit text production process will be at a maximum when writing is (i) dispositionally guided, i.e. for low self-monitors, and (ii) not outline planned, i.e. the order in which content is produced is governed by the implicit organization of content in semantic memory rather than by an explicit, pre-determined

plan in working memory. The implicit text production process will be minimized when writing is (i) directed towards rhetorical goals, i.e. for high self-monitors, and (ii) controlled by an outline, i.e. when the sequence of text production is pre-determined. Furthermore, it suggests that, because changes in content can be induced by both explicit planning and implicit text production, but only implicit text production leads to the development of understanding, there will be no direct relationship between the overall amount of change in content and the development of understanding. Instead, the development of understanding will be directly linked to the extent that new content is produced by the implicit text production process.

This experiment set out to test these claims by using key-stroke logging to provide a direct measure of the extent to which content was modified during the course of text production, and examined how this varied depending on the conditions under which writing took place, and how it was related to developments in the writer's personal understanding of the topic. The present paper will report the results for a simple indicator of content modification during text production, which we will label as the *text modification index*. This corresponds to the total number of words recorded by key-stroke logging divided by the total number of words appearing in the final text. When writers transcribe their thoughts directly into text the index should be 1: all the words that are written down during text production will be retained in the final text. To the extent that the writer changes the way that they express their ideas during text production the index should increase: writers will produce more words during the process of text production than appear in the final text.

The design of the experiment was based on a previous experiment by Galbraith, Torrance and Hallam (2006) and manipulated two variables: self-monitoring and planning. Each group was asked either to make an outline before writing or to sum up their overall opinion of the topic prior to writing (a procedure we call synthetic planning, and which differs from outline planning in that it does not specify the order in which content should be produced during text production.). Our aim was to replicate the conditions of this earlier experiment with a view to assessing how the text modification index varied under these conditions. We expected that, if the dual-process model is correct, the text modification index should be at a maximum when low self-monitors produce synthetically planned texts, and that increases in subjective understanding should be associated with high levels of text modification, rather than with the overall amount of change in content produced in the different conditions.

## Method

### Participants

84 students from the faculty of Arts of the University of Groningen were recruited to participate in the experiment. They were all native Dutch speakers, average age 22.2 years

( $SD = 3.8$ ), and were pre-selected using Snyder's revised 18 item self-monitoring scale (Snyder & Gangestad, 1986). Participants could only take part if they were classified either as a high or a low self-monitor. They were classified as *high self-monitors* (HSM,  $n = 42$ ) if they scored 11-18 on the scale and as *low self-monitors* (LSM,  $n = 42$ ) if they scored 0 - 8 on the scale.

## Design and procedure

High and low self-monitors were randomly allocated to the two planning conditions resulting in the following four experimental groups: (i) HSM outline planning, (ii) HSM synthetic planning, (iii) LSM outline planning and (iv) LSM synthetic planning.

**Writing task** In all four conditions, participants were asked to plan and write an article for the university newspaper discussing whether "our growing dependence on computers and the Internet is a good development or not". The writing task was divided into three phases.

*In phase 1*, participants were first given 10 minutes to list all the ideas they could think of relevant to the topic. It was stressed that each idea should be no longer than a sentence in length. They were then asked to rate how much they felt they knew about the topic on a 7-point scale.

*In phase 2*, participants were given 5 minutes to either write down a single sentence summing up their overall opinion (synthetic planning) or to construct a structured outline (outline planning). They were then given 30 minutes to write a well-structured article for the university newspaper. It was stressed that they had to produce a reasoned argument reflecting their own opinion about the matter. Participants were allowed to consult their written outlines. During writing, keystrokes were logged using Inputlog (Leijten & Van Waes, 2006).

*In phase 3*, immediately after writing, participants were asked again to rate how much they felt they knew about the topic. They were then given 10 minutes to again list all the ideas they could think of relevant to the topic. Finally, they were asked to compare the lists produced before and after writing, and to rate the extent that ideas on list 2 corresponded with ideas on list 1, using a 6-point scale ranging from 1=identical point to 6=no correspondence.

## Measures

**Subjective development of understanding** The ratings of knowledge were used to assess subjective changes in understanding as a consequence of writing.

**Development of ideas** This was assessed using the procedure used in previous research. New ideas were defined as ideas in the second list that received ratings from 4 to 6 for their correspondence with ideas in the first list. Preserved ideas were defined as ideas in the second list that received ratings from 1 to 3 for their correspondence with ideas in list 1. The average length of these ideas was also calculated. These were assessed against baseline measures

of the number and average length of ideas in the list produced before writing.

**Text modification index** In order to assess the process by which writing is carried out a text modification index was calculated. For the text modification index the total number of words recorded by Inputlog are divided by the number of words appearing in the final text.

**Data screening** Preliminary analysis of the data revealed 6 outliers (i.e. scores more than 3 SD's above or below the mean). Three participants had extremely low scores on the initial or post knowledge rating. One had an extremely high score on the mean length of ideas. Two had extremely high scores on the text modification index. These participants were removed from all analyses.

## Results

### Development of subjective understanding

A two-way (2\*2) between subjects ANCOVA with self-monitoring and planning as factors and with prior knowledge as a covariate revealed a significant main effect of type of planning on subjective understanding after writing ( $F(1,73) = 4.61, p = .035, \eta^2 = .033$ ). Figure 1 shows the mean ratings of knowledge before and after writing in each condition (with error bars showing standard errors).

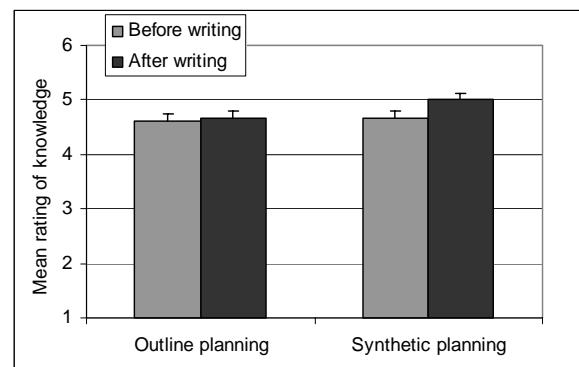


Figure 1: Development of subjective understanding as a function of type of planning.

Planned comparisons comparing mean knowledge ratings before and after writing in the synthetic and outline planned conditions showed that there was a significant increase in knowledge in the synthetic planned condition ( $t(39) = 3.34, p = .002$ ) but no significant difference in the outline planned condition ( $t(37) = 0.47, p = .64$ ).

### Effects on idea change and relationships with developments of subjective understanding

To assess the relationship between changes in the content of the lists produced before and after writing and changes in subjective understanding, we converted the knowledge

ratings to a category variable representing the extent to which knowledge increased, decreased or remained the same. We then carried out a 3-way between subjects MANCOVA, with self-monitoring, type of planning and change in knowledge as independent variables; the number of new and preserved ideas, and the average length of these ideas, as dependent variables; and the number of ideas produced before writing and their average length as covariates. Using Pillai's trace, this showed a significant main effect of type of planning ( $V = .14$ ,  $F (4, 60) = 2.51$ ,  $p = .05$ ) and a significant interaction between type of planning and knowledge change ( $V = .32$ ,  $F (8, 122) = 2.85$ ,  $p = .006$ ). To describe these effects, we will consider them in two stages, starting with the main effect of type of planning and then considering the interaction between type of planning and change in knowledge.

**Main effect of type of planning** There were two important findings here. First, as can be seen in figure 2, the preserved ideas were significantly reduced in length in the outline planning condition but not in the synthetic planning condition ( $F (1, 66) = 5.80$ ,  $p = .019$ ,  $\eta^2 = .05$ ). There was no equivalent effect for the new ideas.

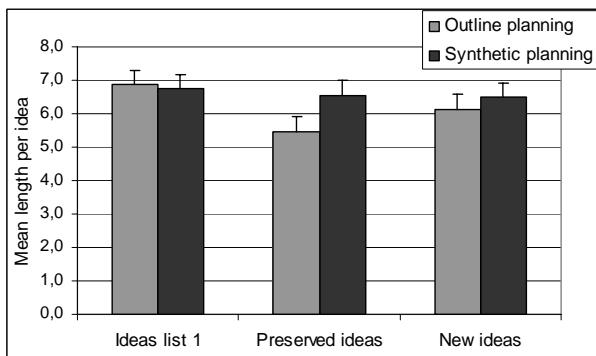


Figure 2: Words per idea for ideas in list 1, preserved ideas in list 2 and new ideas in list 2.

A possible explanation for the effect is that when an outline is constructed it is held in working memory to guide text production. In consequence, when writers refer to ideas in the outline, they label the idea held in memory in an abbreviated form. Although this effect may prove useful as a marker of the extent to which individuals within different conditions construct a mental outline during writing, it does not suggest a substantive effect of type of planning on the content of the lists produced after writing.

The second important finding here is a negative one. The follow-up analysis of the multivariate analysis revealed no apparent effect of either self-monitoring or type of planning on the number of new or preserved ideas produced after writing. Possible reasons for this will be considered in the discussion.

**Interaction between type of planning and knowledge change** To determine the source of this effect, we carried

out simple effects analysis within the synthetic planning and outline planning conditions, using 1-way MANCOVAs, with change in knowledge as the independent variable, the four measures of the lists produced after writing as dependent variables, and the number of ideas in the initial list and their average length as covariates. This confirmed that there was no significant relationship between idea change and changes in subjective knowledge within the outline planning condition ( $V = .191$ ,  $F (8, 54) = .71$ ,  $p = .68$ ). However, there was a highly significant effect of within the synthetic planning condition ( $V = .645$ ,  $F (8, 60) = 3.57$ ,  $p = .002$ ). Univariate ANOVAs, followed by planned comparisons, on each of the dependent variables showed that there were significant effects for 3 of the variables.

First, there was a significant effect on new ideas ( $F (2, 34) = 6.25$ ,  $p = .005$ ,  $\eta^2 = .25$ ), with planned comparisons showing that participants whose knowledge remained the same produced more new ideas ( $M = 7.8$ ,  $se = 0.68$ ) than both participants whose knowledge decreased ( $M = 1.69$ ,  $se = 2.34$ ,  $p = .05$ ) and participants whose knowledge increased ( $M = 4.85$ ,  $se = 0.91$ ,  $p = .045$ ). Although increased knowledge was associated with more new ideas than decreased knowledge, this difference was not significant ( $p = .65$ ).

There was also a significant effect on the average length of new ideas ( $F (2, 33) = 5.94$ ,  $p = .006$ ,  $\eta^2 = .14$ ) with participants whose knowledge decreased producing longer new ideas than those whose knowledge remained the same ( $p = .008$ ) and those whose knowledge increased ( $p = .04$ ). Finally, there was a marginally significant effect on the average length of preserved ideas ( $F (2, 34) = 2.46$ ,  $p = .10$ ,  $\eta^2 = .04$ ), with a tendency for participants whose knowledge remained the same to produce preserved ideas shorter in length than those produced by participants whose knowledge either increased or decreased.

Taken together, these findings suggest, first, that decreases in knowledge occurred in this condition when writers were relatively unable to think of new ideas, and to express what ideas they could think of concisely. This implies that thinking of new content is generally necessary in order to produce satisfactory text. Second, increases in knowledge were associated with the production of fewer new ideas than when knowledge stayed the same. This contradicts previous research. A possible explanation for this is that new ideas were produced by different processes when knowledge remained the same than when it increased. On the assumption that the length of preserved ideas reflects the extent to which writing has been controlled by an outline (see above), then the marginally significant effect on the length of preserved ideas could indicate that texts where knowledge remained the same were relatively more outline planned than the texts where knowledge increased.

### Relationship with processes

The preceding analysis revealed that, despite the significant difference between synthetic and outline planning in the extent to which writers reported increases in understanding,

there were no differences in idea change within the two planning conditions, and generally that there were no relationships between the amount of change in ideas and increased knowledge. According to the dual-process model, this is because new content is produced by two different processes -explicit rhetorical planning and implicitly guided text production- and only implicit text production leads to the development of understanding. To test these claims, we carried out a 3-way between subjects ANOVA on the text modification index, with self-monitoring, type of planning and knowledge change as dependent variables. This produced clear evidence to support these claims.

First, both self-monitoring and type of planning had a clear effect on the extent to which ideas were modified during text production. There was a significant main effect of type of planning ( $F(1, 66) = 5.55, p = .02, \eta^2 = .06$ ), a close to significant main effect of self-monitoring ( $F(1, 66) = 4.53, p = .06, \eta^2 = .03$ ) and a significant interaction between self-monitoring and type of planning ( $F(1, 66) = 4.45, p = .04, \eta^2 = .04$ ). Figure 3 shows that low self-monitors produced higher levels on the text modification index than high self-monitors and this was reduced when text production was preceded by outline planning.

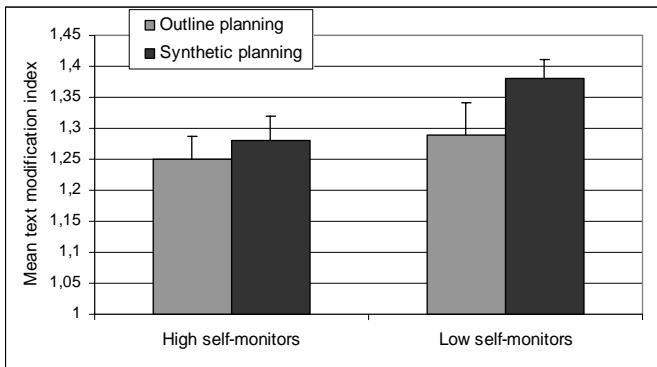


Figure 3: The text modification index as a function of type of planning and self-monitoring.

Second, there was a significant interaction between knowledge change and type of planning ( $F(2, 66) = 3.67, p = .03, \eta^2 = .07$ ). Analysis of simple effects, followed by planned comparisons of the differences between different types of knowledge change, revealed that there was a highly significant main effect of knowledge change within the synthetic planning condition ( $F(2, 34) = 5.59, p = .008, \eta^2 = .22$ ). As can be seen in figure 4, this was a consequence of the fact that increases in knowledge were associated with significantly higher levels of text modification than when knowledge remained the same ( $t(34) = 3.29, p = .007$ ). Although decreased knowledge was also associated with lightly elevated levels of text modification, this was not significantly different from the other conditions.

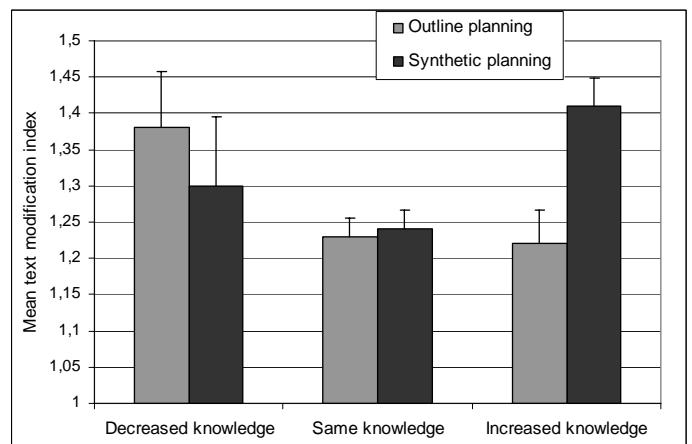


Figure 4: The text modification index as a function of type of planning and knowledge change.

## Discussion

The dual-process model claims that new content is produced during writing by two different kinds of process: explicitly controlled planning to satisfy rhetorical goals and implicitly guided text production articulating the writer's developing understanding. This contrasts with the knowledge-transforming model in two key respects. First, it claims that, although explicitly controlled planning does lead to changes in content after writing, this is essentially a matter of retrieving already existing knowledge which is more appropriate to the rhetorical context than the ideas initially considered relevant to the topic, and hence that changes in content produced by explicit planning will not lead to developments in understanding. Second, it claims that implicitly guided text production is not simply a matter of translating the output of planning into words, but is an active knowledge-constituting process in its own right. Our results provide strong support for both claims.

First, there was clear evidence that content was produced by different processes in the outline planned and synthetically planned conditions. The outline planned condition involved significantly lower levels of text modification during writing than the synthetically planned condition. This is compatible with the claim that changes in content in this condition are a consequence of higher level thinking processes rather than of the modification of content in the formulation of the text itself. By contrast, text modification was at its highest in the condition – the low-self-monitors' synthetically planned texts – where the dual process-model assumes that text production is most implicitly guided, and where new content is assumed to be formulated in the text itself rather than through planning prior to text production.

Second, although both conditions led to a similar amount of change in ideas, as would be expected if both processes play an active role in developing content during writing, only the synthetic planning condition was associated with significant increases in subjective ratings of understanding.

This clearly supports the claim that explicit planning is less strongly associated with the development of understanding than implicitly guided text production is.

Third, there was no relationship between the amount of change in content in the different conditions and increases in subjective understanding. The dual-process model provides a straightforward explanation for this: increases in understanding depend on the extent to which new content is produced by implicitly guided text production. This explanation is strongly supported by the fact that synthetically planned writing involved significantly higher levels of text modification, and that it was precisely those writers within this condition whose understanding increased who produced the highest levels of text modification. The only exception to this extremely clear pattern was that the few writers who experienced decreases in knowledge in any of the conditions also appeared to engage in relatively high levels of text modification. The important feature of these writers, however, is that they also produced few new ideas. This leads to the general conclusion that increases in understanding occur when writers develop new ideas in the course of formulating the text itself. Understanding will remain the same when text production is either controlled to conform to a higher level plan (as in outline planning) or when the writer's knowledge prior to writing is sufficiently clear for text to be fluently produced. Understanding will decrease when text production does not lead to the formulation of coherent new content.

There is one aspect of these results which does not match previous research. Previous studies (see Galbraith, 2009) have found that low self-monitors typically produce more new ideas than high self-monitors under synthetic planning conditions, and that, under these conditions, the number of new ideas is positively correlated with increases in subjective understanding. The dual process model assumes that this is because high self-monitors typically impose more control on text production than low self-monitors, so reducing the extent to which ideas are formulated during text production. This was partially supported in the present experiment in that high self-monitors did engage in significantly less text modification than low self-monitors in the synthetic planning condition. However, there was no difference in the extent to which low and high self-monitors produced new ideas in this condition, and there was a negative rather than a positive relationship between the number of new ideas and increases in understanding.

We believe that this is a consequence of a difference in the constraints under which synthetic planning was carried out in this experiment. In previous research, the external constraints for the text under synthetic planning conditions have either been left unspecified or writers have been actively instructed to write down their thought free from rhetorical constraints. By contrast in this experiment, writers were instructed to produce a finished article for the university newspaper in the time available. According to the dual-process model, this should lead to an increase in the

extent of explicit planning processes, and since these are prioritized by high self-monitors, a greater increase in the number of new ideas produced by high self-monitors compared to low self-monitors. Furthermore, since these new ideas are produced by explicit planning, which according to the dual-process model is not associated with changes in understanding, there will no longer be a straightforward relationship between the amount of new ideas and increases in subjective understanding, just as we found in this experiment. This explanation could be tested by comparing low and high self-monitors writing synthetically planned text, either with clear rhetorical constraints, as in the present experiment, or free from rhetorical constraints as in previous research.

Our general conclusion is that in order to explain how writers develop their understanding it is necessary to examine the processes by which their ideas are created rather than just assess the extent to which they have modified their beliefs. The simple index of text modification that we have used in this paper has shown clear distinctions between different kinds of knowledge change, which strongly supports the broad claim that the development of thought during writing depends on two different kinds of process. Further research is needed, using on-line measures such as key-stroke logging, to examine in detail how ideas are formulated during text production and how this results in developments of the writer's understanding.

## Acknowledgements

We would like to thank Professor Kees de Bot for his helpful comments on earlier drafts of this paper.

## References

- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Galbraith, D. (2009). Writing as discovery. *British Journal of Educational Psychology Monograph Series II 6 - Teaching and Learning Writing*, 5-26.
- Galbraith, D., Torrance, M., & Hallam, J. (2006). Effects of writing on conceptual coherence. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 1340-1345.
- Kellogg, R.T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (2), 355-365.
- Leijten, M., & Van Waes, L. (2006). Inputlog: New perspectives on the logging of online writing. In K.P.H. Sullivan & E. Lindgren (Eds.), *Computer Key-Stroke Logging and Writing: Methods and Applications*. Oxford: Elsevier.
- Snyder, M., & Gangestad, S.W. (1986). On the nature of self-monitoring. *Journal of Personality and Social Psychology*, 51, 125-139.