Designing Thinking-Aloud Studies in ESL Reading

J. Mark Rankin
University of California, Los Angeles

This paper presents and discusses advantages of options available in the design of thinking-aloud studies in ESL reading. It briefly introduces the method, contrasts its use in reading with its use in writing, and suggests possible variables around which to center such research. It presents suggestions as to choosing subjects and reading materials, planning, preparing, and training subjects for thinking-aloud sessions. Also presented are options for recording, transcribing, and analyzing data.

INTRODUCTION

Thinking-aloud has a relatively short history as a research method in the study of second language learning and use. Adapted from cognitive psychology, this method has been referred to as a “stream-of-consciousness disclosure of thought processes while information is being attended to” (Cohen 1983).

The subjects read (usually silently, but cp. p.124) and at intervals report their thoughts on the text and on their reading processes to the investigator. Thinking-aloud can be distinguished from other similar procedures in that, while performing, subjects of language research are not remembering past language use (introspection, retrospection) nor are they asked to make generalisations on their patterns of language behaviour (self-reports). Although this type of disclosure of thought processes is admittedly incomplete, it has been shown to be an effective research tool, especially when compared with other methods which rely on research subjects’ observations of their own behaviour (Ericsson and Simon 1980; Steinberg 1986).

Thinking-aloud has recently been used to develop our understanding of the writing processes of native English speakers (Flower and Hayes 1981) and of non-native speakers (Raimus 1985; Zamel 1982, 1983; Gaskill 1986). However, up to now, little research has been done using thinking-aloud to examine the reading strategies of second-language learners. A few studies have used thinking-aloud as a teaching technique with native speakers (Davey 1984) and to examine native speaker strategies (Bird 1980; Scardamalia and Bereiter 1984; Olshavsky 1977; Church and Bereiter 1983). Hosenfeld (1977, 1984) broke ground with studies of native English speakers thinking aloud in English while reading in another language, but the use of thinking-aloud to study non-native speaker reading processes has been relatively rare (Block 1986).

Perhaps one of the reasons for reading researchers’ reluctance to use this method has been the lack of a readily available guide to its use. Although useful suggestions

J. Mark Rankin did the research for this paper while working for his MA at UCLA. He can be contacted at 278 E, Alegria Avenue, Sierra Madre, CA 91024, USA 1.
of a general nature concerning the design of thinking-aloud projects have been made in Hosenfeld (1984), Block (1986), and Bird (1980), these research reports were not intended to provide in-depth discussion of the method itself. Cohen (1984) does discuss factors such as text-selection, setting up interview sessions; and in making further specific proposals as to method, this paper should be seen as supplementing Cohen's suggestions.

The paper, then, is an attempt to provide a practical manual to the design and implementation of research on thinking-aloud. It has grown out of experience gained while performing thinking-aloud research with university-level ESL students, drawing upon previous research and suggestions from colleagues. The paper describes a general framework for thinking-aloud research design, and explains how options available within that framework can allow research to be tailored to the collection of specific types of data.

READING vs. WRITING

Thinking-aloud has recently enjoyed increased popularity as a research tool among native-English writing researchers, and has been used with non-native speakers writing and thinking-aloud in English (Gaskill, 1986). When asked to think aloud, subjects are generally asked to say aloud the thoughts running through their heads while composing a text on a given topic or while reading a text provided by the researcher. They are told to say whatever they think of, whether related to the task or not. The transcripts of these spoken records of mental processes, called protocols, are then analysed for patterns. These patterns can then form the basis for generalisations and further research.

There are similarities between these studies and studies of native-speaker reading (and presumably between L2 reading and writing) in that both try to uncover processes behind communication via written language. However, there are important distinctions between the two types of inquiry. The protocols produced when people write reflect the genesis and production of a text, while protocols of reading reflect the apprehension, comprehension, and interpretation of text written by another person.

If writing can be considered an intertwining set of decisions made with regard to the choice and arrangement of ideas and words in a text, then reading is the other end of this exchange: a set of decisions leading to an interpretation of the meaning expressed by the writer, based on textual cues. In thinking-aloud studies of writing, the goal has been to develop a model of the writing process by examining the types of decisions a writer makes while writing, e.g. the goals and plans of Flower and Hayes (1981). The emphasis seems to be on constructing a hierarchy of decisions rather than on the motivation for the decisions themselves.
Thinking-aloud studies in reading have been less concerned with developing a comprehensive model of the reading process than with filling in the gaps in an evolving paradigm of existing models (Samuels and Kamil 1984).

They have been used recently for a variety of purposes: to find out what strategies readers actually use, as a teaching method, to test the effect of instruction on strategy use, and as a means of investigating reading as problem-solving, among others.

**REASONS FOR USING THINKING-ALOUD**

Although reading is thought of as a form of communication between the reader and the writer of a text, the interaction takes place in the reader’s mind rather than between two parties face to face (Widdowson 1984; Salimbene 1986). Therefore, the process does not lend itself to direct observation. Since readers do not generally leave a trail of thought processes as they proceed through a text, it is difficult to recreate the act of reading afterwards so it can be studied. The problem of how to observe the mind in action has plagued scientists for centuries, but as yet a definitive solution eludes us. For now, we have to rely on the methods we have.

Various techniques have been employed in order to examine the reading process. A brief discussion can be found in Baker and Brown (1984, 23–28). All the methods have their disadvantages. Comprehension questions typically test product, not process (Alderson 1984, Baker and Brown 1984). They may be dependent on the readers’ memories. Moreover, they tend to structure the readers’ comprehension, forcing them to conform to the question setter’s interpretation. So called ‘on-line’ techniques (e.g. examination of eye-movements) do give indications of reading actually in progress. However, as Baker and Brown point out, they cannot by themselves assess comprehension, and need to be supplemented by performance tasks. The use of interactive computer programmes (Mitchell and Green 1978) or of reading recorders (Thomas and Augstein 1973) can also provide information on when readers decide to pause, skip, re-read, etc., but again need to be supplemented. In Harri-Augstein and Thomas’s later work, this is achieved through post-reading and conversational reporting of the reading process (Harri-Augstein and Thomas 1984), which brings this particular investigatory technique close to thinking-aloud.

As with the other methods, there are drawbacks to thinking-aloud (Ericsson and Simon 1980). Subjects’ reports on their mental processes are not complete, and may be influenced by their perception of what the researcher “wants” them to do. Baker and Brown are skeptical about the accuracy of subjects’ reports on their own reading processes. However, thinking-aloud has continued to gain respect as a research method because it is especially well-suited to the task of providing perhaps
the most direct access we have to the mental processes involved in reading while it is going on.

In uncovering these hidden thought processes, researchers hope to be able to contribute to our body of knowledge in both theoretical and practical ways. By knowing what strategies second-language readers actually use when reading (as opposed to what we think they do), we will not only improve our understanding of reading as a communicative act but also our understanding of how it might best be taught. Information is needed on how reading strategies may vary according to the background of the reader and to the reading task itself.

Possible subject variables include native language, amount and type of prior instruction, level of reading skill, overall level of second language proficiency, and a host of individual differences which could be examined. Task variables include text structure, familiarity of content, difficulty/readability, and purpose in reading, among others. The scope of a particular thinking-aloud study will be limited, of course, by the setting in which it takes place, and since the methodology can be considered a type of case study, small studies which are tightly focused can give the most detailed data. Thus a great deal of careful planning should go into any thinking-aloud project.

SETTING UP PROJECTS

Selecting Subjects

Subjects should be chosen according to criteria dictated by the purpose of the study. Since the practical constraints of transcribing and analysing the protocols will limit the number of subjects, they should be representative of the research population, yet exhibit the characteristics under investigation. For example, a study wishing to distinguish between patterns of strategy use by readers at different levels of proficiency would use subjects from each level available. Subjects of the same level of proficiency would be required if the variables were, say, language background or teaching method.

In any event, backup subjects should be chosen in case they prove unable to perform the somewhat demanding task of thinking aloud in a second language. A certain level of linguistic competence must be assumed, since the subjects have to be able to express their thoughts in the second language.

It is possible, of course, to have the subjects read in their second language while thinking aloud in their first, but two complications arise from this procedure. First, researchers or their assistants must be proficient enough in the subjects' first language(s) to be able to understand and interpret the protocols. This may or may not be a problem, depending on the resources available. The second complication involves the nature of the task itself. Requiring subjects to switch back and forth
between languages while reading and verbalising would seem to encourage translation and other strategies where they might otherwise have been less evident. Of course, manipulation of the language modality may be desired in order to examine its effects as a variable more closely.

**Selecting Appropriate Reading Materials**

Just as subject selection should reflect the aim of the study, so should the selection of reading passages. Considerations include text structure, length, difficulty (including vocabulary) and content. All of these are important to any study, and each can be manipulated as a variable in itself.

From work done on the influence of *textual organisation* on recall (Carrell 1984), it may be inferred that this factor may influence strategy use as well. Passages should be analysed for differences which may predispose subjects to use particular strategies and not others. This becomes more important if the protocols produced in the reading of different passages are to be compared. For example, a spatially organised text may predispose subjects to create certain types of pictures in their minds, while definitions of abstract concepts may preclude these entirely. Of course, in cases where subjects' capability to use specific strategies with certain types of text is under investigation, text structure will be a major factor in the choice of reading materials used.

*Length* is also an important factor in passage selection. A passage should be long enough to allow the subjects to become involved in reading, but not so long that they become fatigued by the demands of thinking aloud for extended periods. A length of between 300 and 1000 words seems to be appropriate under most conditions, but the characteristics of the research population will be the deciding factor. (It may be that younger or less proficient readers should be asked to use shorter passages than older or more proficient readers.)

Another important consideration should be *level of difficulty*. The cognitive load imposed by reading the passage should not be so great as to prevent the subjects from being able to think aloud, nor should a passage be so far below the subjects' ability that it is only perceived at a superficial level, thus encouraging little strategy use.

Material with *subject matter* entirely unfamiliar to the subjects is to be avoided, and topics requiring prior cultural knowledge may bias some subjects' responses. Topics having to do with the human condition, such as family relationships or problems of everyday life, provide interesting material as long as they are universal enough to be understood by the subjects of the study.

However, if the research is focused on teaching strategy use through thinking-aloud, it may best serve research goals to use materials from the subjects' current
classroom reading material. Material with a particularly heavy vocabulary load, logical inconsistencies, passages with pictures or diagrams essential to comprehension, or other materials designed to elicit particular strategies may be useful depending on the research hypotheses.

Normally, passages work best if self-contained. If excerpts from larger works are used, care should be taken that the excerpt does not require knowledge of the previous and subsequent portions of the work in order to be understood. If current classroom reading materials are used for the thinking aloud sessions, prior knowledge of this type can be verified in advance.

In any case, care should be taken to pilot the materials beforehand to avoid last-minute setbacks. The researcher should try thinking aloud with the proposed texts as a test of appropriateness. If a passage is hard to make sense of or too boring to be thought-provoking, then the subjects probably will find it hard to work with as well. Another way to gauge the appropriateness of a passage is to observe the subjects’ reactions during practice. Enough passages should be selected in advance to allow an educated choice for the passages to be taped.

**Planning the Sessions**

Most studies of writing which use thinking-aloud have used videotape to record the subjects while they wrote. This provides not only an audio record of verbalised thoughts, but also enables the researcher to easily coordinate them with the text itself. By focusing the camera on the pages as they are written, it is possible to see how the evolution of the text corresponds to the decisions made in writing it. This is not as easy in reading.

Generally, thinking-aloud studies of reading have used audio tape, since the advantages of video do not always outweigh the disadvantages. Since the process of reading takes place within the mind of the reader, interactions with the writer through the text are made without leaving any visible trace. Of course, if the reader marks the text while reading, reads aloud, or traces the words with a finger, videotape could aid in coordinating the protocols produced while reading with the portions of text which cued them. However, video images of the written page tend to be hard to read, and focusing on the page excludes the facial cues to subjects’ thoughts which may be useful in interpretation of their comments. On the other hand, a full-face shot may prove to be distracting to the subject.

There are several solutions to the problem of how to cue responses and coordinate them with the text, each having its own limitations. Hosenfeld (1984) suggests that the subject be allowed to read and verbalise at will, but that the researcher prod the subject when necessary with open-ended questions. This procedure can provide perhaps the most complete record of the subject’s thoughts, since it is the reader
who controls progress through the text. However, this method requires that copious notes be taken so that the protocols can be matched to the text later. The use of assistants may be advisable to lessen the workload of the researcher.

Block's study (1986), among others, involved placing dots at predetermined points on the page (after each sentence, or after each paragraph), as reminders for the subjects to respond. The advantage here lies in the comparative ease with which the text and the protocols can be coordinated. The researcher need not be present during taping, which may be important under certain circumstances. The disadvantage here is that by predetermining the junctures at which the subjects must verbalise, those thoughts that occur between the dots may be lost. Also, the dots themselves are sometimes blocked out by a subject engrossed in reading.

A compromise between these two systems might be to use dots or some other signalling device in practice texts, while allowing subjects the freedom to respond at will to the research passages. They can be instructed to say what is going on in their heads at any time; the dots are there only as reminders. Some subjects have indicated that such signals interfered with their reading, and they tend to perform well enough in their absence once the nature of the thinking-aloud task is grasped.

The problem of coordinating "free-form" verbalisations with the text can be solved by having the researcher sit beside the subject during the task, watching the subject read. With an identical text in hand, the researcher can follow along, making note of the tape counter's position at the appropriate points in the text. In order to follow along with the subject, it is advisable to have him indicate his own progress through the text, perhaps by following the text with a finger along the margin (which some subjects tend to do anyway). Although this activity can be considered a strategy in itself and may in fact alter the process of reading, these limitations can be noted and allowed for in the research.

As an additional safeguard, it may be advisable to have the subject do a retrospective analysis of the research passages after the thinking-aloud session.

Though retrospective and thinking-aloud data are different, having the former in reserve to back up the latter makes good research sense. Besides offering potential aid in transcription and analysis of the protocols, retrospection can provide interesting sorts of data in itself, especially if the subject had trouble performing the think-aloud task.

PROCEDURES

Practice

The importance of practice cannot be overemphasised. The subjects need to be introduced to the thinking-aloud task before they can be expected to perform it.
Church and Bereiter (1983) suggest a three-part series of explanation, modelling, and supervised practice. A session should begin with an explanation of the purpose of the study, stated in general terms so as to avoid biasing the subjects' responses. They might be told that the study is a way for them to become aware of and possibly improve upon their reading behaviour, or as a way for the researchers to gain useful information about the reading process, or both. A brief explanation of the task itself should be immediately followed by a demonstration.

When modelling the think-aloud task, researchers should simply state that they are going to read the passage in their normal way, but will say out loud every thought that comes to mind. The subjects should be told to follow along as the researchers read, noting the stops at various points to make comments. After finishing the passage, researchers should indicate that these comments reflect their own thoughts while reading. What someone else would think about while reading might be very different, but no better and no worse. Subjects can then be given a passage to try out the technique.

If subjects fail to verbalise enough, researchers should interrupt in a non-threatening manner, clarifying the task as needed. The subjects can then continue reading and thinking aloud until the end of the passage. They can then be questioned about trouble spots noted by the researchers during the reading. This kind of retrospective look at the task can bring it into better focus. A typical exchange might go as follows:

R: What were you thinking when you stopped here? (pointing to a word)
S: I was thinking how it looks like a word in Dutch.
R: It does?
S: Yes. I think they mean the same.
R: Good. Now, the next time you do something like that, say so!
S: Oh! Even things like that. I see what you mean.

(Rankin 1986)

Subjects differ in the amount of practice they need, but it is better to allow extra time than not enough. Generally, if subjects are ever going to catch on they will do so within four or five trial passages. During practice, it is better to begin with shorter texts and build to longer ones. Practice sessions are best followed immediately by the taping session, or there may be a loss of familiarity with the task. Too much practice, however, may be counterproductive if the research is designed to examine what the subjects do at a particular point in time. Given the opportunity, more motivated subjects may alter their strategy use as they see what
works better for them. This is commendable if the research goal is improved reading, but extended practice may be better if provided after the data are collected. If two or more separate sessions are necessitated by external constraints, then time should be spent "warming up" before the passages to be used in protocol analysis are presented.

Recording

In order to make the subjects feel more comfortable with the tape recorder they should be familiarised with the equipment, especially if the researcher will not be present for the entire session. Taping both the practice passages and those to be used in protocol analysis is one way to do this. If the tape recorder has been going throughout the entire session, the subjects tend to forget about it sooner. More passages should be taped than necessary, to allow for data loss through unforeseen problems such as a power failure or stage fright.

If the subjects perform adequately on their own, no interference may be necessary during the taping. If prompting is needed to get the subjects going, it should be as non-directive as possible. This is important, since to suggest the use of certain strategies, even indirectly, may encourage the subjects to produce them where they normally may not have used them. Questions like "What are you thinking about?" should be used, rather than "What do you think will happen next?" If research design calls for the measurement of pauses between comments, then prompting should be avoided entirely, as in Block's 1986 study.

Of course, care should be taken to zero the counter at the beginning of each tape, and only one session should be recorded on one side. If the cueing system outlined above is used, the same tape recorder must be used after transcription to match the protocols with the text, since each counter is slightly different and behaves differently with tapes of different thicknesses and lengths.

WORKING WITH DATA

Transcription

It is advisable to make a copy of each tape before beginning to transcribe. The copy can be used for the actual transcription, with the original being saved for a final fine-tuning of the correspondence between text and protocols. A two-column format is recommended for transcription, since putting the chunks of text read in one column and the subject's comments in the other allows the researcher to follow the subject's progress through the text more easily. This also allows open space for making notes on the transcript as to type of strategy used, time, or other relevant factors.

The use of a computer makes transcription much easier. The text read by the
Noise Pollution

(Title)

Oh, There must be a loud – great deal of noise

(Prediction)

American cities, with their screeching cars, thundering trucks, roaring subways, wailing sirens, blaring horns, and bellowing factories, are tremendously noisy places, and they are getting noisier.

S-sounds--oh--I haven't heard a noise; I-I feel headache already.

(Personal reaction)

(Rankin 1986)

Portions of the text read loud, skipped, or re-read can be marked, and information gained through retrospection can be added later.

Analysis

The analysis of the protocols thus transcribed will vary according to the focus of the particular study, but certain underlying principles should be observed. Subjects'
comments should be taken in the context of the situation, and care should be taken not to ascribe meaning to them except at face value. Some “reconstruction” of missing words or syllables may be necessary in order to get the sense of a comment, but unless subjects are there to help interpret their own comments, the only “reconstruction” done should be for literal meaning. Each comment will indicate the use of at least one strategy, since even a simple “O.K.” or “right” can be seen as a type of comprehension check. Some comments will be evidence of more than one strategy, since the reader may pose a question and then answer it before moving on in the text, for example.

The researcher’s next task, then, is to assign each instance of strategy use to a category. These categories may be of the researcher’s own design (Hosenfeld 1977, 1984), or may be borrowed and/or adapted from other research. Depending on the focus of the study, categorisation systems such as those found in research on learning strategies (O’Malley et al 1985; Dansereau 1985; Weinstein and Mayer 1986; Oxford-Carpenter 1985), systems developed out of particular theories of reading (Samuels and Kamil 1984), or systems from other disciplines such as cognitive psychology may be most appropriate.

Existing paradigms serve as a good starting point, but one of the reasons for performing thinking-aloud research is to enlarge the scope of our understanding of how these strategies fit together. For this reason, inadequacies of existing categorisation systems present a challenge rather than a problem to the researcher.

Once the reading protocols are analysed for strategy use, and these strategies categorised, comparisons may be made to determine the possible influence of other factors on different subjects’ manner of approaching the text. These comparisons can be between different groups of subjects who read the same text, different texts read by the same subjects, or any combination of these.

However the factors or variables are combined, consideration should be made to base the comparisons not only on frequency or relative frequency of strategy use in a particular situation, but also on the appropriateness of such strategy use.

As a case in point, think of two subjects who read the same text, but whose vocabulary knowledge in the second language varies widely. Subjects who stop often to figure out unfamiliar words will obviously use word identification strategies more frequently than subjects who do not stop for individual words because they already know them. Contrast these subjects with others who have a limited vocabulary, but prefer to skip the words they do not know rather than stopping to figure each one out as they read, and one develops a feeling for the complexities possible in trying to decide which is the more effective reader.

Our task, then, is not simply that of differentiating between patterns of strategy use
in our own subjects, but also involves the search for similarities between our findings and those of our colleagues.

CONCLUSION

Much remains to be done in order to determine what reading strategies second-language learners use, and still more before we can adequately judge which ones are more effective in different situations. Since thinking-aloud allows us the most direct access we have to the reading process as it happens, it offers us great opportunities as researchers. It will take a great number of studies to yield the amount of data needed to make generalizations, so it may be better if each study is limited in scope but done with the thoroughness and control that yield depth to an intensive study.

Planning thinking-aloud studies may seem intimidating at first, especially the first time. The hardest part is taking the first step. Once the researcher begins to pilot some materials and starts to get a glimpse at what is going on in the subjects’ minds, the possibilities begin to open up. Much as a good chef starts out with a general idea for a new dish, gets his ingredients, and makes adjustments as he goes along and sees what works best, a researcher will tailor his design according to his research questions and resources. It is my hope that this paper can serve as a reference for researchers, not as a cut-and-dried recipe, but as a framework to use in formulating thinking-aloud research projects. By considering the alternatives and trying them out, researchers will not only be contributing to our knowledge of second-language reading, but also to our understanding of how best to perform and interpret thinking-aloud research.

REFERENCES


