<table>
<thead>
<tr>
<th><strong>ILL Number:</strong></th>
<th>15124178</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Borrower:</strong></td>
<td>ICO</td>
</tr>
<tr>
<td><strong>In Process Date:</strong></td>
<td>20051212</td>
</tr>
<tr>
<td><strong>OCLC Number:</strong></td>
<td>18824112</td>
</tr>
<tr>
<td><strong>Lending String:</strong></td>
<td>*IBS, UPP, MDY, XFF, NDJ</td>
</tr>
<tr>
<td><strong>Patron:</strong></td>
<td>Isabelli-Garcia, Christina</td>
</tr>
<tr>
<td><strong>Journal Title:</strong></td>
<td>Applied language learning.</td>
</tr>
<tr>
<td><strong>Volume:</strong></td>
<td>7</td>
</tr>
<tr>
<td><strong>Issue:</strong></td>
<td>1-2</td>
</tr>
<tr>
<td><strong>Month/Year:</strong></td>
<td>1996</td>
</tr>
<tr>
<td><strong>Pages:</strong></td>
<td>5-24</td>
</tr>
<tr>
<td><strong>Article Author:</strong></td>
<td>Cohen, Andrew</td>
</tr>
<tr>
<td><strong>Article Title:</strong></td>
<td>Verbal Reports as a Source of Insights into Second Language Learner Strategies</td>
</tr>
<tr>
<td><strong>Imprint:</strong></td>
<td>Presidio of Monterey, CA ; Defense Language Institute, Foreign Language Center, 1989-</td>
</tr>
</tbody>
</table>

**Shipping Address:**
AMES LIB
ILL
ILLINOIS WESLEYAN UNIV
PO BOX 2889
BLOOMINGTON IL 61702-2889

**Fax:** (309) 556-3018
**Ariel:** 208.34.150.253
**Odyssey:**

**Ship Via:** Ariel

**Email:** LVIS
**Notes:** LVIS
Notice: Warning concerning Copyright Restrictions

The copyright Law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Any electronic copy or copies, photocopies or any other type of reproduction of this article or other distribution of this copyrighted material may be an infringement of the Copyright Law. This copy is not to be “used for any purpose other than private study, scholarship, or research [section 107]. “If a user makes or later uses any form of reproduction of this copyrighted work for purposes in excess of section 107, Fair use, that user may be liable for copyright infringement.
Verbal Reports as a Source of Insights into Second Language Learner Strategies

Andrew D. Cohen
University of Minnesota

Since a number of different definitions of second-language (L2) strategies have appeared in the literature, first I define the different kinds of behavior that fall under the rubric of L2 learner strategies and, afterwards, classify different types of verbal report. Next, I briefly indicate areas in which verbal report has made a contribution to our understanding of L2 learner strategies. Then, I focus on problematic issues regarding the methodology itself. Finally, I note ways to refine verbal report methods as I list the types of needed information to understand the particular verbal report. I hope that this information should assist researchers in comparing across studies and in replicating studies that have appeared in the literature. Hence, I do not end on a justification of verbal report methods as others have done, but rather on the fine-tuning of such methods.

Within the last fifteen years, verbal reports have been used increasingly as a source of data on the strategies of learning a second or foreign language. A major impetus for this research technique in L2 acquisition has been its successful use in first-language studies, especially in research on cognitive processes in first-language reading and writing. In fact, this impetus continues to manifest itself; at least two new books on verbal reports in first language have just appeared, one dealing with reading (Pressley & Afflerbach, 1995) and the other dealing with writing (Smagorinsky, 1994).

Defining “Second-Language Learner Strategies”

Second language learner strategies encompass both L2 learning and L2 use strategies. Taken together, they constitute the steps or actions selected by learners either to improve the learning of an L2, the use of it, or both. Language-use strategies actually include retrieval strategies, rehearsal strategies, “cover” strategies, and communication strategies. What makes the definition for language-learning and language-use strategies broad is that it encompasses those actions that are clearly aimed at language learning, as well as those that may well lead to
learning but which do not ostensibly have learning as their primary goal. Whereas language-learning strategies have the explicit goal of assisting learners in improving their knowledge in a target language, language-use strategies focus primarily on employing the language that learners have in their current interlanguage.

Thus, strategies for learning the subjunctive in Spanish as a foreign language, for example, could include grouping together and then memorizing the list of verbs that take a subjunctive in constructions like quiero que vengas (“I want you to come”), or noticing the difference in imperfect subjunctive inflections between the -ar conjugation (e.g., cantara) and the -er and -ir conjugations (e.g., comiera, existiera). The specific strategies for memorizing this group might involve writing these verbs inside a box in the notebook and reviewing the contents of the box regularly, as well as noting what these verbs have in common semantically. Language-learning strategies would also include strategies for learning new vocabulary, such as using flash cards, possibly with keyword mnemonics to jog the memory if necessary.

Strategies for using the subjunctive include four subsets of strategies: retrieval strategies, rehearsal strategies, “cover” strategies, and communication strategies. In the above example with the subjunctive, retrieval strategies would be selected for retrieving the subjunctive forms when the occasion arises in or out of class, and for choosing the appropriate forms. For those learners who keep a list of verbs taking the subjunctive, a strategy may involve visualizing the list and cross-checking to make sure that the verb that they wish to use in the subjunctive form actually requires the subjunctive. Likewise, a language-use strategy would entail using the keyword mnemonic to retrieve the meaning of a given vocabulary word. So, say that a learner encounters the verb ubicar (to locate), which she had learned by means of the keyword mnemonic ubicar, and she wants to retrieve the meaning of the word. The language-using strategies would include any efforts by the learner to retrieve the meaning of the word ubicar—involving the linking of the Spanish sounds /ubiks/ with the English /ub k/, and then perhaps seeing an image of someone who keeps turning up everywhere the language learner looks.

Language-use strategies also include strategies for rehearsing target language structures (such as form-focused practice), as well as strategies for covering one’s self in the language classroom (such as participating in classroom tasks to look good in front of other students or the teacher, without intending to learn or communicate any particular aspect of the target language). An example of rehearsal would be form-focused practice, for example, practicing the subjunctive forms for different verb conjugations. An example of a “cover” strategy would be using a memorized and not fully-understood phrase in an utterance in a classroom drill in order to keep the action going. Some cover strategies reflect efforts at simplification (e.g., learners use only that part of a phrase that they can deal with), while other such strategies comp lexify the utterance (e.g., saying something by means of an elaborate and complex circumlocution because the finely-tuned vocabulary is lacking or to avoid using the subjunctive). Both cases represent an attempt to compensate for gaps in target language knowledge.

Communication strategies constitute a fourth subset of language-use strategies, with the focus on conveying meaningful information that is new to the recipient. Such strategies may or may not have an impact on learning. For example, learners may use a vocabulary item encountered for the first time in a given lesson to communicate a thought, without any intention of trying to learn the word. In contrast, they may insert the new vocabulary item into their communication expressly to promote their learning of it.

It is fair to say that verbal report data have enhanced our understanding of all the above types of L2 learner strategies. For example, verbal report data have been collected from learners as they generate mnemonic devices for remembering new vocabulary words and as they attempt to retrieve these words later by means of these mnemonic devices (see Cohen & Aphek, 1980, 1981). Likewise, verbal report data have been used to reveal instances where learners are using material over which they have little control. For example, an earlier study by Cohen and Aphek (1979) revealed an instance of a survival-oriented cover strategy, when an adult learner of Hebrew used a nonexistent form yariya instead of ered (I will get off). Retrospective verbal report provided by the learner in class revealed that although she had been exposed to the future, she had not as yet acquired productive control over the tense. She explained that she lived on a street called Yordei Hasula and knew that the first word of the street name was also derived from the verb laredet ‘to get off.’ So she improvised by making up a word that might convey the future tense. This insight was only made possible through the use of verbal report—in this case, through immediate retrospection.

Since verbal report as it has been applied to the field of learner strategies is not one measure, but rather encompasses a variety of measures intended to provide mentalistic data regarding cognitive processing, the next section provides a brief classification of the report types.

Classifying Verbal Reports

Verbal reports include data that reflect (1) self-report: learners’ descriptions of what they do, characterized by generalized statements about learning behavior—e.g., “I tend to be a speed listener” (2) self-observation: the inspection of specific rather than generalized language behavior, either introspectively, i.e., within 20 seconds of the mental event, or retrospectively—e.g., “What I just did was to skim through the incoming oral text as I listened, picking out key words and phrases” and (3) self-revelation: “think-aloud,” stream-of-consciousness disclosure of thought processes while the information is being attended to—e.g., “Who does the ‘they’ refer to here?” Verbal reports can and usually do comprise some combination of these (Radford, 1974; Cohen & Hosenfeld, 1981; Cohen, 1987). Self-report data tend to appear frequently on questionnaires that ask learners to describe the way they usually learn and use language. Self-observation implies reference to
some actual instance(s) of language learning or use. For example, entries in journals or diaries that retrospectively describe some language-learning or language-use event involving the subjunctive would count as retrospective self-observation. Self-revelation or think-aloud data are only available at the time that the language learning or use events are taking place, and imply that the respondent is describing, for example, the struggle to use the correct form of the subject, and not attempting to analyze this struggle. Thoughts which are immediately analyzed would constitute introspective self-observation—for example, "Now, does this utterance call for the present or imperfect subjunctive? Let me see..."

Examples of learner-strategy studies containing verbal reports in the form of self-report interviews and questionnaires include Naiman, Fröhlich, Stern, and Todesco (1978); O'Malley, Chamot, Stewner-Manzanares, Kupper, and Russo (1985); Wenden (1985); Ramírez (1986); and Oxford, Nyikos, and Crookall (1987). In such studies, the respondents answered interview questions or completed written questionnaires about their language strategies. Since self-report has been shown to be somewhat removed from the cognitive events being described, this approach may produce data of questionable validity. Questionnaire items are more likely to elicit learners' beliefs about what they do, rather than what they actually do. Efforts are often made by investigators to increase the extent of self-observational and self-revelational data and to decrease the amount of self-report. The purpose is to obtain data that describe the learning event at or near the moment it occurs. Such data might be expected to reflect accurately what learners actually do than might the response to a questionnaire item calling for a description of generalized behavior.

Contributions of Verbal Report to Understanding Learner Strategies

Despite frequent criticism (Seliger, 1983; Aflerbach & Johnston, 1984; Olson, Duffy, & Mack, 1984; Lyons, 1986), verbal report has gained popularity in the last several decades because it provides data on cognitive processes and learner responses that otherwise would have to be investigated only indirectly. Furthermore, verbal report has at times provided access to the reasoning processes underlying cognition, response, and decision making. We note that the use of verbal report protocols in L2 learning-strategy investigations has benefited greatly from the extensive use of this research methodology in the native language. Such work, especially in reading and writing (e.g., Garner, 1982; Flower & Hayes, 1984), has paved the way for much of the L2 work.

A recent book by Pressley and Aflerbach (1995) focuses on the use of verbal reports of first-language reading, once again constituting an excellent compendium of ideas for L2 researchers. The authors refer to verbal reports as "a maturing methodology with much interesting work already accomplished and considerable work to be done" (p.1). They demonstrate how the use of verbal report (whether as an exploratory methodology or as a means for testing hypotheses about reading) has yielded an elegant description of reading. They provide a detailed description of what they refer to as before reading, during reading, after reading, monitoring and evaluating strategies, based on a review of 38 primary-data studies. As the authors put it, "The think-alouds were extremely revealing about the dynamics of comprehension difficulties and how understandings of text shift in reaction to comprehension difficulties and surprises in text" (p. 38).

With regard to L2 learning and use, verbal report methods—primarily reflecting self-revelation and self-observation—have been employed as a means of describing strategies in the learning and use of L2 vocabulary (e.g., Cohen & Aphek, 1978, 1981; Neubach & Cohen, 1988; Chern, 1993, Huckin & Bloch, 1993), in L2 listening (e.g., Murphy, 1987), in L2 speaking (e.g., Robinson, 1991, Cohen & Olshtain, 1993; Cohen, Weaver, & Li, 1995), in L2 reading (e.g., Hosenfeld, 1984; Block, 1986; Cavalcanti, 1987; Kern, 1994), and in L2 writing (e.g., Zamel, 1983; Raimes, 1987; Cohen & Cavalcanti, 1987, 1990; Skibniewski, 1990). Verbal report is also used for investigating the subset of L2 communication strategies, especially those used in compensating for gaps in communicative ability (e.g., Pouisse, Bongaerts, & Kellerman, 1986; Pouisse, 1989). In addition, verbal report is used with tasks that combine most or all of the strategy areas, such as in investigating the strategies used in translation of texts (Faerch & Kasper, 1986; Borch, 1986; Gerloff, 1987; Krings, 1987) and those used in taking L2 tests (Cohen, 1984, 1994a, 1994b; Stemmer, 1991; Gordon, 1987; Anderson, 1991; Nevo, 1989).

Despite the extensive use of verbal report methods in numerous recent studies, readers are still sometimes uncertain as to the inferences that they can legitimately make on the basis of these reports. At the same time that Pressley and Aflerbach (1995) refer to verbal reports as a maturing method, they also rightly refer to it as an "underdeveloped" one (p.119). For this reason, I will now consider a series of problematic areas regarding the methodology, with an eye to where development needs to take place.

Verbal Report Methodology

Immediacy of Verbal Report

A distinction has been made in the literature between self-revelational data in the form of immediate, on-line think-aloud protocols (which involve no editing or analysis), on the one hand, and self-observational data in the form of introspective or retrospective self-observation on the other. Ericsson and Simon (1993) have advocated the collection of self-revelational data over other approaches to verbal report because asking questions only about what was needed in short-term memory was seen as a means of making such reports more reliable in that there is no strain on the memory to reconstruct past thoughts. In sharp contrast to this methodological position, the Pressley and Aflerbach (1995) survey of studies in L1 reading found considerable variation as to the immediacy of the reporting and
the amount of interpretation respondents were asked to provide (p. 22).

The researchers found not only self-revelational protocols but also self-observational reports that were collected after each sentence, after each episode, at signaled spots in the text (usually two or more sentences), after every two minutes, at the end of the text, or whenever the readers wanted. Thus, there was a fluctuation both within and across studies as to whether subjects were asked to provide think-aloud, introspective (i.e., within 20 seconds of the event), or retrospective reports (separated somewhat in time from the actual reading). Pressley and Afflerbach (1995) give one explanation for this departure from exclusive use of the think-aloud approach—namely, that to obtain verbal report of otherwise automatized cognition, there is a need to slow down the process by using, for example, the interruptive methods listed above (p. 9).

Not only did Pressley and Afflerbach (1995) have difficulty in determining if verbal reports in the studies that they reviewed reflected traces remaining in short-term memory or rather the subjects' reconstructions of what happened as they read. They were also unable to determine whether there was substantive difference in quality between think-aloud data produced when subjects performed no analysis and the self-observational data when they analyzed what they were thinking (p. 128). The reasons they gave for their inability to make a comparison were (1) there was too little systematic study of this issue in the given research reports, and (2) the verbal reporting itself was influenced differentially by the nature of the training, coaching, or prompting that the respondents received before and during the reporting phase.

Greene and Higgins (1994) considered the issue of delay in the case of retrospective verbal report after the completion of a writing task. The investigators offered four suggestions for improving the reliability and validity of such data: (1) minimizing the time between the process and report by obtaining a report immediately after a writer completes a task, (2) designing prompts that can help writers better access detailed information from their short- and long-term memory (e.g., through the use of concrete examples and contextual cues), (3) making clear to the respondents the purpose of the retrospective accounts, and (4) reporting one's findings in ways that enable readers to see how the conclusions have been derived from the data (e.g., by including enough data in a report so that readers can make their own assessments about the value of research based on retrospection).

**Respondents' Role in Interpreting the Data**

There are researchers who are wary about having subjects interpret why they are doing something. Their rationale is that a request to provide interpretation is more likely to influence how the respondents perform continuing phases of the same task. In addition, they see the asking of a "why" question as likely to produce unreliable answers if at the time the respondent is not thinking about why he/she is doing the action (Ericsson & Simon, 1993, p. 7). Hence, they recommended that interpretation of verbal report be left to researchers, rather than, asking the respondents to categorize their cognitions. Despite these recommendations, Pressley and Afflerbach's (1995) review of 38 primary data studies of L1 reading found that many studies went beyond having readers simply report their thoughts, and requested them to interpret their processes as well (p. 21). Presumably, the insights from self-observation offer a rich enough source of information not available through think-aloud protocols alone that researchers are willing to risk threats to the reliability of the verbal report tasks in order to obtain the data.

**Prompting for Specifics in Verbal Report**

Early descriptions of verbal report methods usually included the stipulation that respondents not be given instructions as to what to report on. They were to be left to their own devices since any instructions might lead to biased processing. But anyone who has been faced with analyzing pages of transcribed undirected verbal report protocols has seen that such data are likely to be general and incomplete. So, even methodological hard-liners like Ericsson and Simon (1993) favor instructions to the respondents to make the verbal reports complete (p. 11).

Thus many studies now do include instructions to elicit particular cognitive behaviors. For example, reading researchers have used different processes in the different studies. Pressley and Afflerbach (1995) found one study that requested that subjects create a summary of what they read, and in which the respondents were informed about the importance of summarization, a second that asked respondents to attend to content and style when reading; and others that required subjects to draw inferences. The authors conclude that prompting respondents to use particular processes may be necessary: "it is reasonable to prompt [processes] in order to assure that a sample of the target processes will, in fact, be observed" (p. 133). With regard to post-experimental assessment, Cantor, Andreason, and Waters (1985) have found that more valid information is produced if the cues involve specific items from the experiment (in their case, animal episodes and geometric form episodes).

**Guidance in Providing Verbal Reports**

Not only has it been proven effective to have respondents receive specific prompts as to what to report about, but it has also been seen that instruction in how to provide verbal report for a given task improves the quality of the data. Ericsson and Simon (1993) have found that to assure that the verbal report does not interfere with the task at hand, there must be warm-up trials after the instructions with tasks that yield easy-to-analyze think-aloud, introspective, and retrospective reports. The researchers suggest that—to ensure consistency—subjects be given trials on these warm-up tasks until they are able to make verbal reports without confounding them with explanations and justifications (p. xxxii). "In some studies, more
extensive warm-up procedures are used explicitly to train the subjects to conform to the think-aloud instructions” (Ericsson & Simon, 1993, p. 82). In a study in which subjects were asked not only to think aloud, but also to give a reason for each response they made before keyboarding it into the computer, the respondents who provided verbal report after receiving training improved more on the computerized cognitive task than those who did not receive the training (Berry & Broadbent, 1984). In the review of 38 primary studies of verbal report in L1 reading, Pressley and Afflerbach (1995) found that while in some studies the respondents were given an opportunity to practice, in others they were not (p. 22).

### Reactive Effects of Verbal Report

Verbal report that involves intervening during the performance of a task has been criticized for the inevitable reactive effects that such intervention causes. Stratman and Hamp-Lyons (1994), for example, conducted an exploratory study to determine the extent of reactivity, in which they had writers engage in two revision tasks eight weeks apart, one with think-aloud verbal reports. All subjects were trained in providing think-aloud protocols. The researchers found for the eight subjects in their study that thinking aloud increased the number of new “word-level” errors (morphological, tense, and spelling, p. 103). Contrary to the investigators’ expectations, thinking aloud was found to inhibit word or phrase additions. They also found that while thinking aloud did not have an impact on complex meaning changes at the microstructural level, it stimulated the production of entirely new sentences (p. 107). They concluded that thinking aloud does alter the nature of processing in the revision phase of writing. They posited that think-aloud protocols may systematically influence the correction of organizational-level errors (i.e., reordering of displaced sentences, adjusting faulty paragraph boundaries, detection of faulty pronoun references, detection of redundancies, detection of word-level errors—in morphology, tense, and spelling—and introduction of new word-level errors) and influence the amount and kind of microstructural meaning changes as well.

While the 1994 study by Stratman and Hamp-Lyons on the use of verbal report during the revision phase of writing produced reactive results of a negative nature, a series of other studies would suggest that there may be positive consequences of verbal report. Collecting retrospections (termed intervention protocols) at various points during the writing has also been found to improve the reliability of the data collection task (Swanson-Owens & Newell, 1994). It was found that the interruption of writing for the purpose of reflecting on process served as a supportive measure in helping writers learn about composing, and thus to provide scaffolding for a subject’s learning during data collection. Similarly positive outcomes of verbal report have been reported for studies in the areas of vocabulary learning and reading as well. For example, Crutcher (1990) conducted a study of vocabulary learning with keywords and obtained retrospective reports for half of the items. He found that retention of the words was better for those items.

With regard to verbal reports in L2 reading, Nyhus (1994) looked at the attitudes of NNS of English toward the use of verbal report to elicit their reading comprehension strategies. The respondents were seven third-quarter students in the Commanding English Program in General College at the University of Minnesota—a bridge program for refugee and immigrant non-native speakers of English. Five of the respondents were Vietnamese, one Chinese, and one Russian. Most had been in the U.S. for only two to three years. The study looked at their attitudes toward the effects of think-aloud and retrospective verbal report on their reading. They were also asked to assess verbal report as a research methodology.

The respondents were shown a videotape of the researcher reading aloud and providing a think-aloud verbal report from a sociology text. Three excerpts from a sociology text were chosen for use with the respondents. Two were for practice readings and the third for the data collection. Red dots were placed between sentences to remind the respondents to verbalize their thoughts. Two sets of interview questions were developed, the first twelve questions to be asked following the respondents’ initial think-aloud verbal report and the second eleven questions to be asked following the respondents’ retrospective verbal report. The respondents were asked to read the text as they normally would but to say all of their thoughts aloud, in English. They were told they could read the text silently, but all chose to read it aloud. The respondent and the researcher then listened to the recording of the verbal report and the respondents provided a retrospective verbal report by pausing the tape when they wanted to make additional comments about thoughts that had occurred to them while reading the text. The researcher also had the respondents report on what they had been thinking but not verbalizing. Next, the researcher interviewed the respondents regarding their views about the think-aloud methodology. Then, there was a second interview to elicit attitudes toward the retrospective methodology after the task had been completed.

For the most part, the respondents viewed the effects they attributed to verbal report as beneficial. Most felt that think-aloud verbal report affected their thinking about reading in a positive way. They reported that it enhanced their awareness and assessment of various aspects of the reading process, including an awareness of themselves as readers and of their interaction with the given text. Only two of the seven had negative comments about verbal report, and these were the students whose English was the most limited. Since all verbal report was conducted in English, performing the verbal report in English was most likely to the detriment of those with more limited English skills. There may, in fact, be a second-language threshold below which attempts to provide verbal report in the target language are counterproductive.

Despite several cases of difficulty in reporting in English, all respondents viewed verbal report as useful in various ways. They saw it as a means for placing students at a given level, as a diagnostic tool for determining their specific reading needs at a given level, and as a solitary or group study technique. The students...
reported that in small groups they discovered alternative ways of thinking about a text. Retrospective verbal report generated by having readers listen to and comment on a playback of their think-aloud verbal report provided still more insights. It was seen as a means of helping readers, instructors, and researchers alike to gain further insight into readers’ thinking and reading processes.

**Towards Robust Verbal Report Methods and Complete Write-Ups**

What has emerged from this discussion of methodological issues in verbal report as applied to language learner strategies is that we are in need of both more refined measures and more details about the verbal report methods of each study. This more detailed information would facilitate cross-study comparisons regarding both learner strategies and the research methodology itself. So, for example, Pressley and Afflerbach (1995) propose a study of reading strategies that would call for a carefully detailed comparison between think-aloud verbal reports and delayed reports. The study would assess the extent to which ongoing verbal report might interfere with the natural reading processes, and the extent to which delayed stopping after every sentence or few sentences might shift the nature of subsequent reading, if at all. Pressley and Afflerbach (1995) would also wish to investigate the question of how long reports can be delayed before they decay (p. 13). In making their plea for greater completeness in the description of verbal report methods, they include a listing of variables for which more complete and systematic information is desirable (pp. 120-123).

**Issues of Method and of Write-Ups for Verbal Reports**

Let us now relate Pressley and Afflerbach’s listing of variables to second-language studies. The following list includes areas for refining verbal report methods and for encouraging write-ups describing the methods in detail sufficient to ensure comparison across studies.

**Subjects’ Characteristics**

For the purpose of comparison across studies, the educational background of the respondents, their knowledge of the task at hand, and their motivation to perform the task should be made clear. In addition, their level of language proficiency (especially in the case of L2 studies) and their age should be indicated. Pressley and Afflerbach (1995) also suggest that their short-term memory capacity and their spatial ability be noted, but this would entail special psychological testing that is usually not conducted in L2 acquisition research. These authors also stress the need for studies with larger numbers of subjects, since most studies are of individual cases or small groups. Their point is that while the accumulation of small-scale studies of verbal report does help to generate a large-scale picture, comparison across them can be somewhat problematic, especially if the approaches to data collection are different. The problem is that most researchers do not have the budget to conduct verbal report work with large groups.

Whereas Pressley and Afflerbach (1995) limit themselves to respondents who were performing a task in their native language and providing verbal report in that language, research into L2 learner strategies is faced with the issue of choice of language for verbal reporting. When dealing with groups of speakers of numerous languages, the verbal report protocols may need to be in the target language. In cases where the respondents share the same native language or speak a limited number of languages, it may be advisable to give them a choice as to language of verbal report, since the less proficient they are in the target language, the more difficulty they may experience trying to perform the task and provide verbal report in the target language at the same time. The study by Nyhus (1994), in fact, found that the two poorer ESL readers were the ones reporting difficulty providing the verbal report, which was in the L2. Regardless of whether bilinguals use one language or the other for their verbal report, it is important that the researcher indicate the extent to which one or the other is used.

**Characteristics of the Materials**

When textual material serves as a stimulus for verbal report data, it would be helpful if the investigator specified the genre of the material, its topic, its length, and its difficulty level for the given respondents. While some or most of these variables may be provided as a matter of course (especially if texts are included in the appendix of the study), Pressley and Afflerbach would request that investigators indicate the fit between the task and the characteristics of the given respondents. Any such details could help other researchers to interpret the findings with greater ease, as well as to attempt replication of the study, if so desired. Perhaps more so in foreign than in native language reading, the genre of the text can make a big difference in the ease of reading. Even if the readers feel comfortable with the genre (e.g., journalistic writing), still they may have difficulty with the specific topic transmitted by means of that genre (e.g., an account of a holiday with which the reader is completely unfamiliar).

**Criterion Task**

It is imperative for the purpose of comparison that the researcher provide a clear indication of the tasks that the respondents were asked to perform (e.g., in reading research, whether it was free recall, recognition, question answering, summarization, or some combination of these), plus the directions given to the subjects. Pressley and Afflerbach found in the studies they reviewed that the instructions were either not provided or that reference to them was vague. The reason that the instructions are considered so crucial in verbal report work is expressly because of the orienting that takes place through instructions. It is also important to have a clear description of any technical equipment employed in the study (e.g., a multimedia program on CD-ROM). Likewise the goals of the language task should be clear,
as well as the modalities utilized.

Guidance in Verbal Reporting

It is valuable both for purpose of comparison across studies and for replication that information be given as to the nature and extent of guidance that the subjects received in verbal reporting. It is useful to know, for example, whether the subjects received feedback in practice sessions, whether they were coached during the data collection sessions, and if so, the length of the guidance—for example, until they got the behavior correct or until they acted as they were supposed to act. It has become more common to instruct respondents in how to provide verbal report, as well as to coach them as they are providing it (e.g., requesting that they not report on the basis of what they usually do, but rather that they stick to what they are actually doing in the given instance).

Methods of Analysis

To help other researchers interpret the findings, it may prove beneficial to include details concerning the development of categories and coding of verbal reports. Further it may be beneficial to include the codes and symbols used in the transcriptions of the verbal report protocols as well—for example, symbols for suprasegmental features, such as tone of voice.* Pressley and Aflerbach found that the reporting of these methods was usually incomplete.

Categories Used to Score Verbal Report Protocols

It is helpful for researchers to indicate how the scoring of verbal report protocols is done, since so much interpretive work is involved. If the respondents themselves listen to their verbal reports in order to assist in the interpretation of protocols, as in the case of the study by Nyhus (1994) on the effects of verbal report on L2 reading, it would be important to highlight this feature and describe it fully in the write up phase. Such a procedure has the value of improving the validity of the measure, since the respondents themselves are verifying the accuracy of what they reported (choice of words, completeness of the report, etc.) and possibly adding what they had neglected to mention the first time around. It might even pay to have researchers provide verbal report while they are engaged in the task of making their decisions about how to score given instances of behavior appearing in the protocols. Verbal report protocols of raters of L2, for example, reveal instances where the raters do not understand the categories that they are supposed to be using in their ratings (e.g., "style," "register," and so forth).

Inter-Rater Reliability Checks

In cases where two or more investigators score the data, it would be advisable to run inter-rater reliability checks to determine the extent to which the investigators are using similar criteria in arriving at scores. Information about such checks should be provided in the research report.

Selection of Verbal Report Excerpts for Inclusion in Research Reports

A somewhat subtle issue is that of how the data are chosen for inclusion in reports. Other researchers would want to know how representative such excerpts are of the data set as a whole. There is a concern that the investigators may slant the findings according to the excerpts from the data that they choose to select for inclusion in any reports that they write. It is for this reason that Greene and Higgins (1994) go to some lengths to demonstrate how to represent verbal report data in an equitable way in their study of retrospective verbal report of L1 writing processes.

Theories Used in Framing Verbal Report Study

The researchers are asked to identify the theoretical principles that the verbal report techniques were intended to investigate. Pressley and Aflerbach consider it the researchers’ responsibility to provide information as to whether the verbal report measures really reflect the cognitive processes that are reported. This information is necessary to validate the verbal report measures of the study. They contend that the researchers should indicate the relationship between the verbal report and the performance outcomes, much as they do in their own book, by demonstrating that theoretical models of reading (e.g., Baker & Brown, Anderson & Pearson, Dijk and Kintsch, and their own models of constructively responsive reading) are supported by verbal report data obtained from reading studies. As Pressley and Aflerbach (1995) put it,

As validation efforts proceed, we urge careful attention to the establishment of clear linkages between theory, verbal process reports, and other measures that can be complementary to verbal self-reports. We believe this work will do much to bring verbal reports from the status of a ‘bootstrap operation’ (Ericsson & Simon, 1993) to a maturing methodology (p. 126).

Most published studies of second language acquisition include a statement of the research questions and the rationale for each one. If the verbal report measures are simply aimed at exploring some aspect(s) of these research questions, then the theoretical underpinnings are probably provided. It is possible, however, that the theoretical rationale for a given verbal report procedure is not overtly clear to the reader of the report. In such cases, the request would be to provide this rationale.
The Validity of Verbal Reports

While the above discussion of nine issues focused mostly on the reliability of the verbal report measures, their validity also comes into play in each and every issue. While larger samples help to make the results more valid, an alternative to increasing the sample size would be to amass a series of well planned and executed small-scale studies. As for the role played by the materials and the tasks in the determination of validity, it is imperative that the consumers of the research results have adequate information about the nature of the materials and about the specific instructions that the respondents were given for performing the task. Such information is crucial in interpreting the verbal report responses received. By the same token, the consumers of the reports need to know the extent to which the respondents were coached on how to perform the task.

Once the data are collected, the analysis procedures also have direct impact on whether the data measure what they purport to measure—that is to say, the rationale for the construction of the analysis categories and then the actual process of data analysis. Did the raters understand and properly use all of the rating categories? With regard to inter-rater reliability (if there is more than one rater), a low correlation would call into question not only the reliability of the ratings but their validity as well.

Furthermore, there is the issue of whether the reported data are comprehensive or selective, and if selective, what this says about the validity of the reporting process. Finally, there is concern that the study not use verbal report simply for its own sake, but rather because the data collection method does, in fact, help to gather data bearing on the theoretical issue(s) at hand.

Summary and Conclusion

The article started by defining and then illustrating the important split between language-learning strategies on the one hand and language use strategies on the other. It then contrasted the three forms of verbal report—self-report, self-observation, and self-revelation—and briefly indicated the contribution that verbal report methods have made to the understanding of language-learning and use strategies. It then focused on concerns about the appropriate use of these measures and about the nature of reports that include the findings from the use of such measures. The issues included the immediacy of the verbal reporting, the respondents’ role in interpreting the data, prompting for specifics in verbal report, guidance in verbal reporting, and the reactive effects of verbal reporting.

The lengthy focus on both refining verbal report methods and on improving the write up of verbal report procedures was intended to underscore the importance of being rigorous both in design and in description. The purpose would be not only to improve the data, but also to assist others in understanding fully what was done, in being able to make comparisons to other studies, and in being able to replicate the studies. In addition, the point was made that care in the write up can help to dispel arguments that such methodological approaches are not adequately rigorous.

While previous studies have tended to focus on justifying verbal report in the face of criticism from those opposed to it, this article has instead focused on the fine-tuning of verbal report methods. Since by now so many studies using verbal report techniques have emerged, the time has come to provide greater systematicity both in the collection of such data and in the reporting of such studies through the research literature. This article has intended to help researchers ask and get answers to more finely-tuned questions, so that the already valuable findings of verbal report studies might be enhanced by the extra methodological rigor.

Notes

1 See Cook (1993, Ch. 6), Ellis (1994, Ch. 12), Towel & Hawkins (1994, Ch. 13), and McDonough (1995) for recent reviews of the learning and communication strategy literature, and for discussion of its terminology.

2 A keyword mnemonic is a word or phrase, usually in the native language of the learner, that is linked by similar sounds to the word to be learned in the target language. The learner then creates an interacting image between this keyword and the target word.

3 I am grateful to Tim McNamara for suggesting the term “cover” strategies (Personal Communication, July 9, 1996).

4 While critics have often referred to verbal report data as too qualitative in nature, Hillocks (1994) argues that quantitative studies, while taking the stance of being dispassionate and objective, inherently involve biased interpretations. By the same token, verbal report often relies on counting instances of activity in order to arrive at conclusions. Hence, he would argue against categorically labeling verbal report data as qualitative.

5 The Ericsson and Simon book was originally written in 1984 and was reissued intact in 1993 with a 53-page preface, intended to update the book. The 1984 volume has served for many as the authority on how verbal reports are supposed to be conducted. The Pressley and Afflerbach volume constitutes perhaps the first effort to determine the fit between Ericsson and Simon’s methodological recommendations and actual uses made of the methodology in the field.

6 Current research policies at many institutions now require that respondents be fully informed as to what they will be asked to do and that they give their written consent. So in essence, the days of concealing the true motives from the respon-
dents are waning. Furthermore, it may be counterproductive for the purposes of the study to have the students distracted for even a portion of the time by anxieties concerning the uses to be made of their responses.

Actually both reliability and validity are of concern here. First, there is the concern that the measures produce data that are consistent within a given verbal report session and across sessions of a similar nature. The second concern is that the data be valid—i.e., that they actually constitute examples of what they purport to be. Hence, reliability is a contributing factor in the determination of validity.

Of course, verbal report data may also be collected in the written modality, as has been done in various studies (e.g., Robinson 1991). In such cases, there would be no suprasegmentals.

References


Employing a Questionnaire to Assess the Use of Language Learning Strategies

Rebecca L. Oxford
University of Alabama

Questionnaires are among the most efficient and comprehensive ways to assess frequency of language learning strategy use. This article discusses the validity of the most widely employed strategy questionnaire, the ESL/EFL version of the Strategy Inventory for Language Learning (SILL). Validity of the SILL rests on its link with language performance (course grades, standardized test scores, ratings of proficiency), as well as its relationship to learning styles. Reliability of the SILL is high across many cultural groups. Appropriate uses and limitations of questionnaires for strategy assessment are detailed, along with implications for research and instruction.

One of the most prevalent ways to assess the use of language learning strategies is to use a questionnaire (otherwise known as an inventory or a summative rating scale). The strategy questionnaire most often used around the world at this time is the Strategy Inventory for Language Learning (SILL, Oxford, 1986-1990). This article has five purposes: (1) to discuss strategy questionnaires other than the SILL, (2) to describe the SILL’s purpose and nature, (3) to provide detailed psychometric results concerning the ESL/EFL (English as a second or foreign language) version of the SILL, (4) to present information on the appropriate uses and limitations of a strategy questionnaire in comparison with other means of strategy assessment, and (5) to provide implications for research and instruction.

Strategy Questionnaires Other than the SILL

To present a context, we turn first to strategy questionnaires other than the SILL. Nearly a dozen have been used in published studies. For example, Bialystok (1981) used a 12-item, structured, untitled rating scale to assess strategy use. The scale asked questions about the extent to which strategies were used on both oral and written tasks in communicative settings (the strategies were functional practice and inferencing or guessing) and in formal classroom settings (the strategies were formal practice and monitoring). Using the scale with students of French in grades 10 and 12 in Canada, Bialystok found that functional practice had a stronger relationship with achievement.