Applied Language Learning

1996 Volume 7 Numbers 1 & 2

Articles

General Information

[Back to DLI Publications]

Applied Language Learning

PB 65-96-1 & 2

The mission of Professional Bulletin 65-96, *Applied Language Learning* (US ISSN 1041-6791), is to provide a forum for the exchange of ideas and information on instructional methods and techniques, curriculum and materials development, assessment of needs within the profession, testing and evaluation, and implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences.

*Applied Language Lea*rning, published semi-annually by the Defense Language Institute, Foreign Language Center (DLIFLC), presents professional information: The views expressed herein are those of the authors, not the Department of Defense or its elements. The content does not necessarily reflect the official US Army position and does not change or supersede any information in official US Army publications. *Applied Language Learning* reserves the right to edit material.

By Order of the Secretary of the Army:

DENNIS J. REIMER
General, United States Army
Chief of Staff

Official:

JOEL B. HUDSON
Administrative Assistant to the
Secretary of the Army
02136

Daniel D. Devlin
Colonel, United States Army
Commandant, DLIFLC

Eugene F. Beauvais
Colonel, United States Air Force
Assistant Commandant, DLIFLC

 Applied Language Learning is available on microfilm from ERIC Clearinghouse on Language and Linguistics, Center for Applied Linguistics, 1118 22nd Street, NW, Washington, DC 20037.

[Table of Contents]

Applied Language Learning

PB-65-96-1 & 2

Special Acknowledgment

The members of the editorial board of *Applied Language Learning express warm* words of appreciation to Professor **June K. Phillips,** the outgoing member, for her contribution to the journal.

[Table of Contents]

Preface

Applied Language Learning occasionally publishes collections of colloquium papers on adult language learning for functional purposes. The 1991 issue of the journal (Vol. 2, No. 2) presented proceedings from an invitational conference on "Improving Foreign Language Teaching through Technology." The conference was held at the Defense Language Institute on October 30, 1990.

This special issue of *Applied Language Learning* contains colloquium proceedings on "Validity Issues in the Assessment of Second-Language Learner Strategies" presented at the American Association for Applied Linguistics (AAAL) Conference in Long Beach, California, in March 1995.

If you have recently organized a colloquium on adult language learning, you may wish to submit those proceedings for publication in *Applied Language Learning*. To submit your proceedings for consideration, send a proposal briefly describing the purpose of the colloquium and the content of individual papers.

Applied Language Learning
Editor, Dr. L. Woytak
Defense Language Institute
Foreign Language Center
Presidio of Monterey, CA 93944-5006
United States of America

Lidia Woytak
Editor

[Table of Contents]

[Table of Contents]

Applied Language Learning
1996, Vol. 7, Nos. 1 & 2, pp. 1-4

Introduction
Validity Issues in the Assessment of L2 Learner Strategies

Roberta G. Abraham and Roberta J. Vann
Iowa State University

As both producers and consumers of L2 strategy research, we have often wondered about the accuracy of the picture of learner strategies emerging from the literature over the past two decades. During this period a number of methods of assessing learning strategies have been used, beginning with informal observation and "talking with people" reported by Rubin (1975) and including think-alouds (e.g., Cohen & Hosenfeld, 1981), interviews (e.g., O'Malley, Chamot, Stewner-Manzanares, Kupper, & Russo, 1985), questionnaires (e.g., Politzer, 1983; Oxford, 1986), computer assessment (e.g., Jamieson & Chapelle, 1987) and, more recently, product analysis (Vann & Schmidt, 1993; Abraham & Vann, 1994). Our questions about these methods have included the following:

- Do learners responding to questions in an interview or on a questionnaire really understand what they are being asked?
- Do learners tell the truth in responding to these questions (rather than providing the answer they think the researcher wants to hear)?
- How complete and accurate a picture of the processes are learners able to self-report?
- Why do two different methods of assessing learner strategies sometimes not give the same answers (as in LoCastro, 1994)?
- To what extent do "unnatural" methods like think-alouds reflect the strategies learners normally use in performing a task?
- How accurately and completely do performance data collected on computers or in learner products reflect learner processes?

Thus, when we were invited to organize a colloquium for the 1995 American Association for Applied Linguistics conference, we knew immediately what our topic should be: the validity issues surrounding the assessment of L2 learner strategies. As a basis for exploring these issues, we turned to Messick's (1989) insightful discussion in which he notes that "[v]alidity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (p. 13, italics in original). He points out that it is not the method or observation device per se that is to be validated, but rather "the inferences derived fromtest scores or other indicators—inferences about score meaning or interpretation and about the implications for action that the

interpretation entails" (p. 13). Thus, in validation, we must look not only at the method of collecting data, but also, and more importantly, at how the data are interpreted, that is, what inferences are drawn from the results and how these inferences are justified, and what uses can legitimately be made of these interpretations. Messick stresses that validity is a "unitary concept," and that it is a matter of degree, not of all or none. Furthermore, over time, the existing validity evidence becomes enhanced (or contravened) by new findings" (p.13). [1]

For the colloquium, we asked three colleagues, each of whom has been associated with a key method for investigating L2 learner strategies, to join us in this discussion: Andrew Cohen, Rebecca Oxford, and Carol Chapelle. We all considered the following questions in our presentations:

- What have we learned about the use of our particular method for obtaining information about learner strategies over the past 15-20 years?
- How have particular users of the method defined the term "strategy"?
- What can one legitimately infer about strategy use from data obtained from the method? Conversely, what inferences cannot or should not be drawn from data obtained from the method?
- In what way has the method been particularly useful (to researchers and practitioners)?
- What problems, perhaps unforeseen, have arisen with the method in assessing strategy use?

The articles in this issue are revised versions of the papers these questions elicited. The first two articles deal with relatively well established methods of eliciting information about learner strategies (verbal reports and questionnaires), while the last two discus methods less frequently associated with strategy research (computer-assisted assessment and analysis of learner products).

In the first article, Andrew Cohen contrasts three types of verbal report methods that have been used to illuminate L2 learner strategies, and then discusses several research practices that should be taken into account in interpreting the data. The last portion of his paper addresses ways in which verbal report methods can become more "robust" and the reporting of research using verbal report more complete, thus permitting more valid interpretation and more appropriate use of the results.

Rebecca Oxford's article on the use of questionnaires to assess the use of L2 learning strategies traces the development of her *Strategy Inventory for Language Learning (SILL)* and provides various types of support for the validity claim that it "measures what it purports to measure." Oxford is careful to note what kind of information questionnaires can and cannot provide, and concludes with a number of uses that can legitimately be made of the SILL.

The third article, by Carol Chapelle, reviews computer-assisted strategy assessment over the past fifteen years. Chapelle then looks at two studies where learnerstrategies were inferred from performance data collected by the computer, describing the researchers' justifications for their inferences and suggesting further arguments that could have been made. While there are problems in computer-assisted strategy assess- ment, Chapelle sees this method as a non-obtrusive means of observing learner behavior that can complement other methods in integrating SLA and pedagogical research.

The final article describes two of our own studies in which learner strategies are inferred from learner products. Building on research in communication strategies that shows how products examined within appropriate conceptual frameworks can reveal learner intentions and processes, we describe evidence for the validity of our interpretations of our products, cloze responses and academic reading notes, as reflections

of learner processes. In this article we hope to dispel the notion that process and product are dichotomous and to suggest that products provide yet another means of discovering how learners learn.

Note

[1] According to Messick (1989), this view of validity has evolved from the more traditional one in which three types of validity were identified: content, criterion-related (predictive and concurrent), and construct. He shows how content and criterion-related evidence can provide support for construct validity, which in turn should be complemented by consideration of the implications of using a specific instrument in a specific situation. All of these types of evidence can contribute to the on-going validation process.

References

Abraham, R. G., & Vann, R. J. (1994, March). *Metacognition in the cloze performance of nine Lebanese ESL learners.* Paper presented at the Annual Conference of the American Association for Applied Linguistics, Baltimore.

Cohen, A., & Hosenfeld, C. (1981). Some uses of mentalistic data in second language research. *Language Learning, 31*(2), 285-313.

Jamieson, J., & Chapelle, C. (1987). Working styles on computers as evidence of second language learning strategies. *Language Learning, 37*(4), 523-544.

LoCastro, V. (1994). Learning strategies and learning environments. *TESOL Quarterly, 28*(2), 409-414.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement,* 3rd ed. (pp. 13-103). New York: American Council on Education & MacMillan.

O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Kupper, L., & Russo, R. P. (1985). Learning strategies used by beginning and intermediate ESL students. *Language Learning, 35*(1), 21-46.

Oxford, R. L. (1986). *Development and psychometric testing of the Strategy Inventory for Language Learning (SILL).* ARI Technical Report 728. Alexandria, VA: Army Research Institute. Appendices as ARI Research Note 86-92.

Politzer, R. (1983). An exploratory study of self-reported language learning behaviors and their relation to achievement. *Studies in Second Language Acquisition, 6*(1), 54-65.

Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly, 9*(1), 41-51.

Vann, R. J. & Schmidt, H. H. (1993, April). *Using student notes to assess reading strategies*. Paper presented at the 27th Annual Convention and Exposition of Teachers of English to Speakers of Other Languages, Atlanta.

[Table of Contents]

[Table of Contents]

Applied Language Learning
1996, Vol. 7, Nos. 1 & 2 , pp. 5-24

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.[1]
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.[2]

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 *ubiquitous*, 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 /*ubik*/ with the English /*yub 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).[3] 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 complexify 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 achieved productive control over the tense. She explained that she lived on a street called *Yordei Hasira* 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; Afflerbach & 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.[4] 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 Afflerbach (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., Poulisse, Bongaerts, & Kellerman, 1986; Poulisse, 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; Borsch, 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 Afflerbach (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 heeded 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.[5] In sharp contrast to this methodological position, the Pressley and Afflerbach (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).[6]

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).[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 cued 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, Andreassen, 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 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 con- founding 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.[8] Pressley and Afflerbach 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 Afflerbach 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 Afflerbach (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 respondents are waning. Furthermore, it may be counterproductive for the purposes of the study to have the subjects distracted for even a portion of the time by anxieties concerning the uses to be made of their responses.

[7] Actually both reliability and validity are of concern here. First, there is the concern that the measure 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.

[8] 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

Afflerbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behavior, 16*(4), 307-322.

Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *Modern Language Journal, 75*(4), 460-472.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology, 36A,* 209-231.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly, 20*(3), 463-494.

Borsch, S. (1986). Introspective methods in research on interlingual and inter- cultural communication. In J. House & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication*. Tuebingen: Gunter Narr, 195-209.

Cantor, D. S., Andreassen, C., & Waters, H. S. (1985). Organization in visual episodic memory: Relationships between verbalized knowledge, strategy use, and performance. *Journal of Experimental Child Psychology, 40,* 218-232.

Cavalcanti, M.C. (1987). Investigating FL reading performance through pause protocols. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research*. Clevedon, England: Multilingual Matters, 230-250.

Chern, C-L. (1993). Chinese students' word-solving strategies in reading in English. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning.* Norwood, NJ: Ablex, 67-85.

Cohen, A. D. (1984). On taking language tests: What the students report. *Language Testing, 1*(1), 70-81.

Cohen, A. D. (1987). Studying language learning strategies: How do we get the information? In A. L. Wenden & J. Rubin (Eds.), *Learner strategies in language learning*. Englewood Cliffs, NJ: Prentice-Hall International, 31-40.

Cohen, A. D. (1990). *Language learning: Insights for learners, teachers, and researchers*. NY: Newbury House/Harper & Row.

Cohen, A. D. (1994a). *Assessing language ability in the classroom*. Boston: Newbury House/Heinle & Heinle.

Cohen, A. D. (1994b). English for academic purposes in Brazil: The use of summary tasks. In C. Hill & K. Parry (Eds.), *From testing to assessment: English as an international language*. London: Longman, 174-204.

Cohen, A. D., & Aphek, E. (1979). *Easifying second language learning*. A research report under the auspices of Brandeis University and submitted to the Jacob Hiatt Institute, Jerusalem. Educational Resources Information Center, ERIC ED 163 753.

Cohen, A. D., & Aphek, E. (1980). Retention of second-language vocabulary over time: Investigating the role of mnemonic associations. *System, 8*(3), 221-235.

Cohen, A. D., & Aphek, E. (1981). Easifying second language learning. *Studies in Second Language Acquisition, 3*(2), 221-235.

Cohen, A. D., & Cavalcanti, M. (1987). Giving and getting feedback on compositions: A comparison of teacher and student verbal report. *Evaluation and Research in Education, 1*(2), 63-73.

Cohen, A. D., & Cavalcanti, M.C. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 155-177). Cambridge: Cambridge University Press.

Cohen, A. D., & Hosenfeld, C. (1981). Some uses of mentalistic data in second-language research. *Language Learning, 31*(2), 285-313.

Cohen, A. D., & Olshtain, E. (1993). The production of speech acts by EFL learners. *TESOL Quarterly, 27*(1), 33-56.

Cohen, A. D., Weaver, S. J., & Li, T-Y. (1995). The impact of strategies-based instruction on speaking a foreign language. Unpublished research report. Minneapolis: Center for Advanced Research on Language Acquisition, University of Minnesota.

Cook, V. (1993). *Linguistics and second language acquisition*. New York: St. Martin's Press.

Crutcher, R. J. (1990). The role of mediation in knowledge acquisition and retention: Learning foreign vocabulary using the keyword method. Tech Report No. 90-10. Boulder: University of Colorado, Institute of Cognitive Science.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

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

Faerch, C., & Kasper, G. (1986). One learner--two languages: Investigating types of interlanguage knowledge. In J. House & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication* (pp. 211-227). Tuebingen: Gunter Narr.

Flower, L., & Hayes, J. R. (1984). Images, plans, and prose: The representation of meaning in writing. *Written Communication, 1*(1), 120-160.

Garner, R. (1982). Verbal-report data on reading strategies. *Journal of Reading Behavior, 14*(2), 159-167.

Gerloff, P. (1987). Identifying the unit of analysis in translation: Some uses of think-aloud protocol data. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 135-158). Clevedon, UK: Multilingual Matters.

Gordon, C. (1987). The effect of testing method on achievement in reading comprehension tests in English as a foreign language. Unpublished M.A. Thesis, Tel-Aviv University, Ramat-Aviv, Israel.

Greene, S., & Higgins, L. (1994). "Once upon a time:" The use of retrospective accounts in building theory in composition. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 115-140). Thousand Oaks, CA: Sage.

Hillocks, G., Jr. (1994). Interpreting and counting: Objectivity in discourse analysis. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 185-204). Thousand Oaks, CA: Sage.

Hosenfeld, C. (1984). Case studies of ninth grade readers. In J.C. Alderson & A.H. Urquhart (Eds.), *Reading in a foreign language* (pp. 231-249). London: Longman.

Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153-178). Norwood, NJ: Ablex.

Kern, R. G. (1994). The role of mental translation in L2 reading. *Studies in Second Language Acquisition*, *16*(4), 441-461.

Krings, H. P. (1987). The use of introspective data in translation. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp.159-176). Clevedon, UK: Multilingual Matters.

Lyons, W. (1986). *The disappearance of introspection*. Cambridge, MA: MIT Press.

McDonough, S. (1995). *Strategy and skill in learning a foreign language.* London: Edward Arnold.

Murphy, J. M. (1987). The listening strategies of English as a second language college students. *Research and Teaching in Developmental Education, 4*(1), 27-46.

Naiman, N., Fröhlich, M., Stern, H., & Todesco, A. (1978). *The good language learner*. Research in Education Series, No. 7. Toronto: Ontario Institute for Studies in Education.

Neubach, A., & Cohen, A. D. (1988). Processing strategies and problems encountered in the use of dictionaries. *Dictionaries: Journal of the Dictionary Society of North America, 10*, 1-19.

Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, *6*(2), 199-215.

Nyhus, S. E. (1994). Attitudes of non-native speakers of English toward the use of verbal report to elicit their reading comprehension strategies. Plan B Paper, M.A. in English as a Second Language. Minneapolis, MN: University of Minnesota.

Olson, G.M., Duffy, S.A., & Mack, R.L. (1984). Thinking-out-loud as a method for studying real-time comprehension processes. In D.E. Kieras et al. (Eds.), *New methods in reading comprehension research* (pp.253-286). Hillsdale, NJ: Lawrence Erlbaum.

O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Kupper, L., & Russo, R. (1985). Learning strategies used by beginning and intermediate ESL students. *Language Learning, 35*, 21-46.

Oxford, R., Nyikos, M., & Crookall, D. (1987). *Learning strategies of university foreign language students: A large-scale study*. Washington, D.C.: Center for Applied Linguistics.

Poulisse, N. (1990). *The use of compensatory strategies by Dutch learners of English.* Dordrecht: Foris/Berlin: Mouton De Gruyter.

Poulisse, N., Bongaerts, T., & Kellerman, E. (1986). The use of retrospective verbal reports in the analysis of compensatory strategies. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research*. Clevedon, UK: Multilingual Matters, 213-229.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum.

Radford, J. (1974). Reflections on introspection. *American Psychologist, 29*(4), 245-250.

Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning*, *37*(3), 439-467.

Ramírez, A.G. (1986). Language learning strategies used by adolescents studying French in New York schools. *Foreign Language Annals*, *19*(2), 131-141.

Robinson, M. (1991). Introspective methodology in interlanguage pragmatics research. In G. Kasper (Ed.), *Pragmatics of Japanese as native and target language*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawaii (Technical Report; v. #3), 29-84.

Seliger, H.W. (1983). The language learner as linguist: Of metaphors and realities. *Applied Linguistics*, *4*(3), 179-191.

Skibniewski, L. (1990). The writing processes of advanced foreign language learners: Evidence from thinking aloud and behavior protocols. In J. Fisiak (Ed.), *Papers and studies in contrastive linguistics* (pp. 193-202). Poznan: Adam Mickiewicz University.

Smagorinsky, P. (Ed.). (1994). *Speaking about writing: Reflections on research methodology*. Thousand Oaks: Sage Publications.

Stemmer, B. (1991). *What's on a C-test taker's mind? Mental processes in C-test taking*. Bochum: Universitatsverlag, Dr. N. Brockmeyer.

Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89-112). Thousand Oaks, CA: Sage.

Swanson-Owens, D., & Newell, G. E. (1994). Using intervention protocols to study the effects of instructional scaffolding on writing and learning. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 141-162). Thousand Oaks, CA: Sage.

Towell, R., & Hawkins, R. (1994). *Approaches to second language acquisition*. Clevedon: Multilingual Matters.

Wenden, A. L. (1985). Learner strategies. *TESOL Newsletter*, *19*(5), 1-7.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly*, *17*(2), 165-187.

Acknowledgment

I wish to acknowledge Elaine Tarone, Michael McCarthy, Roberta Abraham, and Rebecca Oxford for helpful feedback on earlier drafts of this article, and Sonya Yoder for assistance in editing the paper.

Author

ANDREW D. COHEN, Director, Institute of Linguistics and Asian and Slavic Languages & Literatures, University of Minnesota, 196 Klaeber Court, 320 16th Ave SE, Minneapolis, MN 55455. Specializations: language learning strategies, foreign language education.

[Table of Contents]

[Table of Contents]

Applied Language Learning
1996, Vol. 7, Nos. 1 & 2, pp. 25-45

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 than did any of the other strategies, even though monitoring and inferencing were used more often. Formal practice with rules and structures was less effective as students advanced to higher levels of learning, but functional practice had no such limitation. Reliability and validity data were absent for this instrument.

Politzer (1983) published an untitled, 1-4-scaled strategy scale including 51 items divided into three groups: general behaviors, classroom behaviors, and interactions outside of class. Using this survey with U.S. university students of French, German, and Spanish, Politzer found that course level influenced strategy use, with

higher-level students using more so-called "positive" strategies (i.e., strategies related to communicative language proficiency); and that females used social learning strategies more often than males. No reliability or validity data were given.

Politzer and McGroarty (1985) used a somewhat similar *Behavior Questionnaire* containing 66 items divided into three groups: individual study behaviors, classroom behaviors, and interactions outside of class. Reliability was marginally acceptable (.51, .61, and .63). The survey was used with students learning intensive ESL in an eight-week course. Improvements in ESL achievement were related to individual strategies, such as asking questions for clarification. Successful strategies for grammar differed from those for listening and speaking. Major academic field had a significant effect on strategy choice, with engineers avoiding strategies that were deemed "positive" for gaining communicative language proficiency; but there was an overlap with nationality, since many engineers were also Asian.

McGroarty (1987) used a 56-item *Language Learning Strategy Student Ques-tionnaire* with a 0-6 range, divided into the same three groups as in the Politzer and McGroarty study above. No reliability or validity data were published. University students of Spanish, although taught by communicative methods, nevertheless avoided authentic practice strategies and used traditional learning strategies, such as relying heavily on the dictionary.

The *Learning Strategies Inventory* (Chamot, O'Malley, Kupper, & Impink-Hernandez, 1987) is a 48-item, 1-4-scaled instrument divided into five parts: listening in class, speaking in class, listening and speaking outside of class, writing, and reading. The items reflected a variety of ways of applying a total of 16 strategies. Results showed that students of Russian used more strategies than students of Spanish, while Spanish and Russian students used somewhat different strategies across language levels (beginning and intermediate or advanced). No data were published on reliability or validity.

Padron and Waxman (1988) developed a 14-item, 1-3-scaled instrument to assess reading strategies of Hispanic ESL students in grades 3-5. Seven of the items were expected to be positively related to learning and seven negatively related. Results showed that six of the seven most-used strategies were in the predicted-positive group. However, only two strategies were significantly related to learning outcomes, and these were both in the negative direction; no strategies significantly helped learning to occur. No reliability or validity data were offered.

Bedell (1993) points out a number of additional strategy scales. Huang (1984) and Huang and van Naerssen (1987) used a *Strategies Questionnaire* for Chinese EFL learners. This instrument includes some scaled items and some yes-no items, as well as free-response questions. Most of the items concern strategies for improving listening and speaking skills. Wangsotorn, Sripaipan, Rattanaprucks, Jarunggidanan, Singkalwanij, and Vejaphurti (1986) used the *Chulalongkorn University Language Institute Learning Strategy Form A* (consisting of 42 yes-no statements about students behaviors) for Thai learners of EFL. Kim (1991) designed a *Perceptual Learning Strategy Questionnaire*, including 18 items. Noguchi's (1991) *Questionnaire for Learners* is an instrument with 24 items on a 3-point scale followed by 24 on a 4-point scale, based largely on items from the *SILL*. Wen and Johnson's (1991) strategy scale is also adapted from the *SILL*.

Few of the above instruments have any published reliability or validity data. This is the key reason that the *SILL* was developed. If the psychometric properties of reliability and validity have not been explored, it is impossible to know whether we can put faith in the results of the research. Another reason for developing the *SILL* is that the preceding instruments do not always systematically represent the wide variety of strategies viewed as important to language learning; often they stop with cognitive and metacognitive strategies. Thus a more comprehensive scale was needed for measuring strategy use among ESL and EFL students.

The Strategy Inventory for Language Learning (SILL)

Development

The *SILL* (Oxford, 1986-1990) was first designed as an instrument for assessing the frequency of use of language learning strategies by students at the Defense Language Institute Foreign Language Center in Monterey, California. Two revised versions of the *SILL*—one for foreign-language learners whose native language is English (80 items) and the other for learners of English as a second or foreign language (ESL/EFL, 50 items)—were published in an appendix to Oxford's (1990b) learning strategy book for language teachers. This article deals only with *research done using the 50-item (short) version*. For details on the longer version, see Ehrman and Oxford (1989, 1990), Nyikos and Oxford (1993), Oxford (1986), Oxford and Ehrman (1995), Oxford and Nyikos (1989), Wildner-Bassett (1992a), and Bedell (1993).

It is estimated that 40 to 50 major studies, including a dozen dissertations and theses, have been done using the *SILL* These studies have, by late 1995, involved approximately 10,000 language learners. According to research reports and articles published in the English language within the last ten to fifteen years, the *SILL* appears to be the only language learning strategy questionnaire that has been extensively checked for reliability and validated in multiple ways.

The *SILL* uses a choice of five Likert-scale responses for each strategy described: *never or almost never true of me, generally not true of me, somewhat true of me, generally true of me,* and *always or almost always true of me*. The *SILL* response options are based on the widely used and well accepted response options of the *Learn- ing and Study Strategies Inventory* described by Weinstein, Palmer, and Schulte (1987). On the *SILL*, learners are asked to indicate their response (1, 2, 3, 4, or 5) to a strategy description, such as "I try to find patterns in English" or "I plan my schedule so I will have enough time to study English." In addition to the original English version, the ESL/EFL *SILL* has been translated and used in the following languages: Arabic, Chinese, French, German, Japanese, Korean, Portuguese, Russian, Spanish, Thai, and Ukrainian.

Two strategy experts matched the *SILL* items with agreement at .99 against entries in a comprehensive strategy taxonomy of language learning. This taxonomy was built from a detailed blueprint of a range of over 200 possible strategy types (for complete details see Oxford, 1986).

One important note is that the *SILL* conceptualizes language-learning strategies in a broad way to include the social and affective sides of the learner as well as the more intellectual (cognitive) and "executive-managerial" (metacognitive). Therefore, when the *SILL* is related to language performance, the "whole learner," rather than just the cognitive and metacognitive aspects of the learner, is usually involved. This implies that language learning, as much as or more than almost any other discipline, is an adventure of the whole learner, not just a mental exercise.

In 1989, the *SILL* was organized according to strategy groups using a factor analysis. This procedure allowed the researcher to divide the instrument into dimensions usually referred to as subscales or factors. Six subscales were developed based on the early factor analyses, with the intent that each subscale would have an adequate number of items to facilitate more in-depth understanding of the learning strategies for ESL/EFL. These subscales included:

1. *Memory strategies*, such as grouping, imagery, rhyming, and structured reviewing (9 items).
2. *Cognitive strategies*, such as reasoning, analyzing, summarizing (all reflective of deep processing), as well as general practicing (14 items).
3. *Compensation strategies* (to compensate for limited knowledge), such as guessing meanings from the context in reading and listening and using synonyms and gestures to convey meaning when the precise expression is not known (6 items).
4. *Metacognitive strategies*, such as paying attention, consciously searching for practice opportunities, planning for language tasks, self-evaluating one's progress, and monitoring errors (9 items).
5. *Affective (emotional, motivation-related) strategies*, such as anxiety reduction, self-encouragement, and self-reward (6 items).
6. *Social strategies*, such as asking questions, cooperating with native speakers of the language, and becoming culturally aware (6 items).

As shown above, the largest group of items is the cognitive strategies. This stands to reason, because research on learning strategies suggests that cognitive strategies possess the greatest variety, covering strategies related to practice and to the all-important "deep processing" in which learners analyze, synthesize, and transform new information (Oxford & Ehrman, 1995).

A *SILL* package includes a short set of directions to the student with a sample item, the 50-item instrument, a scoring worksheet on which students record their answers and calculate their averages for each strategy subscale and their overall average, a summary profile that shows their results and provides examples for self-interpretation, and a strategy graph that allows each learner to graph results from the *SILL*. A background questionnaire is also available to document age, sex, language experience, motivation, and other information (see Oxford, 1990b).

Psychometric Qualities of the ESL/EFL SILL

This section describes the psychometric qualities of the 50-item ESL/EFL *SILL*. Normally, such quality is established and presented in terms of reliability and validity. (Note that psychometric quality data are also available for the longer form of the *SILL* that was designed for native English speakers learning foreign languages; see especially Oxford, 1992 and Oxford & Ehrman, 1995.)

Reliability

Reliability refers to the *degree of precision or accuracy of scores on an instrument*. In the case of the *SILL*, Cronbach alpha, a measure of internal consistency, was chosen as the most appropriate reliability index. The Cronbach alpha reliability coefficient is used on continuous data such as the Likert-type scale in the *SILL*.

Though the current ESL/EFL *SILL* was constructed using six subscales, reliability of the *SILL* is determined with the whole instrument. This is because the six subscales are strongly correlated with the *SILL* mean (.66 to .81) and moderately correlated with each other (.35 to .61); see Oxford and Ehrman (1995).

In general, the ESL/EFL *SILL* reliabilities have been high. With the ESL/EFL *SILL*, Cronbach alphas have been .94 using the Chinese translation with a sample of 590 Taiwanese university EFL learners (Yang, 1992a); .92 using the Japanese translation with 255 Japanese university and college EFL students (Watanabe, 1990); .91 using the Korean translation with 59 Korean university EFL learners (Oh, 1992); .93 using the researcher-revised Korean translation with 332 Korean university EFL learners (Park, 1994); and .91 using the Puerto Rican Spanish translation with 374 EFL learners on the island of Puerto Rico. (These reliabilities are similar to the range of .91 to .95 found for the 80-item foreign language *SILL* given in the native language of the respondent; see Bedell, 1993; Ehrman & Oxford, 1989, 1990; Nyikos & Oxford, 1993; Oxford, 1986; Oxford & Burry, 1993; Oxford & Ehrman, 1995; Oxford & Nyikos, 1989; Wildner-Bassett, 1992a).

Slightly lower but still very acceptable reliabilities are found for the ESL/EFL *SILL* when it is *not* administered in the native language of the respondents but is given in English instead. All the reliabilities in this paragraph refer to heterogeneous (multi-language) groups of ESL learners in the U.S. Phillips' (1990, 1991) data had a reliability of .87 with 141 students. *SILL* data from Oxford, Nyikos, Nyikos, Lezhnev, Eyring, and Rossi-Le (1989) showed a reliability of .86 with 159 students. Anderson's (1993) data on 95 students had a reliability of .91. Involving 31 learners, Talbott's (1993) data had a reliability of .85. A three-study combination (merging ESL data from Anderson [1993]; Talbott [1993]; and Oxford, Talbott, & Halleck [1990]) showed a reliability of .88 with 137 students.

Thus, reliability of the ESL/EFL *SILL* goes down, but not greatly, when the *SILL* is administered in the target language, English, rather than in the respondent's native language. These reliabilities are very respectable, and the *SILL* can be administered in the respondent's native language or a foreign or second language with confidence that measurement error is minimal.

Validity

Validity refers to *the degree to which an instrument measures what it purports to measure.* In the past, several bases have existed for justifying validity: content validity, criterion-related validity (predictive or concurrent), and construct validity. At the current time, these aspects of validity have been condensed to a single general validity (Chapelle, 1994; Messick, 1989).

Justifications of this broad-scale validity, according to Messick (1989), are comprised of construct validity, utility, value implications, social consequences, interpretation, and real-world action. Chapelle (1994, p. 161) asserts that "construct validity is central to all facets of validity inquiry, as most researchers have agreed for some time" and that "researchers should also consider justifications [for instrument validity] pertaining to test utility and the consequences of testing. . . . Researchers are obligated to use construct validity evidence as a basis for considering how their instruments impact the contexts in which they are used." This makes validation an ongoing process, which is the opposite of the notion of an instrument "validated once and for all time." Evidence of validation is not singular but additive. Further, it is not just the instrument that is validated, but also its use in a far bigger context of interpretation and action.

The data on the *SILL* show an ongoing effort at validation, using evidence based partially on construct validity.

Construct Validity as Seen in Relationships between Strategies and Language

Performance

SILL construct validity is partially shown in relationships between the *SILL* on the one hand and language performance on the other. This evidence is probably the strongest support possible for the assertion of the construct validity of the *SILL*.

A number of ESL/EFL studies have demonstrated this relationship. In these studies, language performance is measured in various ways: general language proficiency tests (Rossi-Le, 1989; Wen & Johnson, 1991; Green & Oxford, 1992; Phillips, 1990, 1991; Chang, 1991; Park, 1994), oral language proficiency tests (Chang, 1991), grades in a language course (Mullins, 1991), language achievement tests directly related to course content (Oxford, Park-Oh, Ito, & Sumrall, 1993a, 1993b), proficiency self-ratings (Oxford & Nyikos, 1989; Watanabe, 1990; Chang, 1991), and professional language career status (Ehrman & Oxford, 1989).

Here are some examples of the relationship between strategy use on the ESL/EFL *SILL* and language performance. Rossi-Le (1989) found that for 147 adult ESL students in the Midwestern and the Northeastern parts of the U.S., language proficiency level (on a standardized test) predicted strategy use in multiple-regression analyses. More proficient ESL students used self-management strategies like planning and evaluating ($p < .006$) and formal practice ($p < .02$) significantly more often than less proficient ESL students.

Strategy use was related to language achievement scores (final test grades) in a study involving 107 high school students of Japanese. The ESL/EFL *SILL* was modified slightly for the distance education students in this study by Oxford, Park-Oh, Ito, and Sumrall (1993a, 1993b). In a multiple regression analysis, learning strategy use was a moderate but significant predictor of Japanese language achievement (.20, $p < .04$). The only other significant predictor was the degree of learner motivation (.30, $p < .003$). (Note that this study adapted the ESL/EFL *SILL* for a group learning a language other than English.)

Using a modified Japanese version of the ESL/EFL *SILL*, Wen and Johnson (1991) studied the learning strategies of 242 second-year English majors at seven post-secondary institutions in Nanjing and Shanghai. These subjects had recent national English proficiency scores that averaged 10 points higher than the country's mean. Using partial least squares, the researchers found that one-third of the variance in English proficiency was related to combined effects of six variables, three of which were groups of strategies taken from the *SILL*.

Takeuchi (1993a) used multiple regression and found that eight *SILL* items predicted 58% of the variance in scores on the *Comprehensive English Language Test (CELT)*. The *CELT* was used in that study to measure English achievement among 78 Japanese first-year students of English at a women's college in Kyoto. The figure of 58% is unusually high for just eight strategies. Four strategies positively predicted language achievement: writing notes, messages, letters, or reports in English; trying not to translate verbatim; dividing words into parts to find meaning; and paying attention when someone is speaking English. This means that students who used English for multiple forms of writing, avoided translation, used word-analysis, and paid close attention were more likely to have high *CELT* scores. Four strategies negatively predicted language achievement: asking questions in English; using flashcards; writing down feelings in a language learning diary; and trying to find as many ways as possible to use English. In other words, the more these strategies were used, the lower the *CELT* score.

This is explainable partly because lower-proficiency students might ask more questions, have more turbulent feelings and be more willing to write them down, use very basic strategies like flashcards, or reach out (perhaps in desperation) to many different ways of using English. Takeuchi (1991a, 1991b, 1993a, 1993b) explained some of these findings based on cultural influences.

Watanabe (1990) asked university and college EFL students in Japan to rate from low to high their own proficiency in English. These proficiency self-ratings correlated moderately (average r = .30) with *SILL* strategies (p < .0005-.001), except for those in the category of social/affective strategies. This trend indicates that most *SILL* strategies were used more often by students who rated their language proficiency higher and they were used less often by students who rated their language proficiency lower.

Chang (1991) used the *SILL* to investigate the learning strategies and English proficiency of 50 mainland Chinese and Taiwanese ESL students at a southeastern university in the U.S. Three measures of proficiency (self-ratings and two standardized tests) showed different statistical effects on strategy use. Students who rated themselves above average in proficiency used more strategies overall than those who rated themselves below average. Neither the scores on the *Test of English as a Foreign Language* (*TOEFL*) nor the *Ilyin Oral Interview* were significantly related to overall strategy use, but students with high scores on the oral interview used significantly more social strategies than those with low scores.

Park (1994) employed the *SILL* to determine the relationship between strategy use and proficiency among 332 students of EFL at the Korea Maritime University and Inha University. Park divided the subjects into three groups according to their strategy use: low, medium, and high. Then Park calculated *TOEFL* scores for each group. According to an ANOVA, the *TOEFL* mean scores of these three groups differed significantly from each other. Post-hoc tests showed that the high strategy use group had a language proficiency score that was significantly higher than that of the medium strategy use group, which in turn had a slightly higher language proficiency score than that of the low strategy use group. Thus, a linear relationship was shown between strategy use and language proficiency. In addition, Park found that the correlation between total *TOEFL* scores and strategy use was r = .34 (p < .0001). Cognitive, social, and metacognitive strategies had a slightly higher relationship (r =.33, .30, and .28 respectively) to *TOEFL* scores than did other kinds of strategies (memory, r =.24; affective, r =.23; compensation, r =.21).

Phillips (1990, 1991) found strong relationships between ESL/EFL *SILL* frequencies and English proficiency levels (measured by the *TOEFL*) among 141 adult ESL learners in seven western states in the U.S. She found no consistent differences between high-proficiency students and low-proficiency students on entire strategy categories, so she looked at strategies singly. She found that middle scorers on the *TOEFL*, who thus had moderate proficiency in English, showed significantly higher overall strategy use than did the high-proficiency or the low-proficiency group, when strategy use was defined as the mean number of strategies used frequently and the mean number of strategy categories that had at least one frequently used strategy. The profile of medium-proficiency students using more strategies more often than high-proficiency or low-proficiency students produced a curvilinear pattern. Additionally, Phillips discovered that high *TOEFL* scorers used such learning strategies as paraphrasing, defining clear goals for learning English, and avoiding verbatim translation significantly more often than low *TOEFL* scorers. The low *TOEFL* scorers, many of whom would logically be found among beginning students, reported significantly greater use of such strategies as using flashcards, finding out how to be a better speaker, looking for conversation partners, noticing tension or nervousness, and writing down feelings in a journal.

Green (1991) investigated 213 Spanish-speaking students learning English on the island of Puerto Rico. The *English as a Second Language Achievement Test* (*ESLAT*), which was used in the study, is a measure of overall English proficiency (not achievement on a given curriculum). Green found moderate and significant correlations, usually in the upper .30s, between *SILL* strategy factors and *ESLAT* proficiency scores, and he discovered the same level of correlations between individual *SILL* items and proficiency scores. In a later analysis of variance, Green (1992) showed that language level had a statistically significant influence on strategy use, with higher-proficiency students in general using strategies more frequently than lower-proficiency students. With a larger sample of 374 students, Green and Oxford (1995) found that language proficiency level had significant effects on the use of the following kinds of strategies: compensation strategies ($p < .0001$), cognitive strategies ($p < .0001$), metacognitive strategies ($p < .0025$), and social strategies ($p < .008$). Two other categories of strategies, memory and affective strategies, displayed no significant difference by proficiency level. In the four significant categories, higher proficiency was associated with more frequent strategy use. Significant variation occurred by gender, with females using strategies significantly more often than males in this study.

In Mullins' (1991) *SILL* study, 110 Thai university-level EFL majors showed linkages between strategy use and various measures of English proficiency. For instance, compensation strategy use correlated at $r = .38$ ($p < .0001$) with language placement scores and at $r = .32$ ($p < .006$) with language course grades. A correlation of $r = .24$ ($p < .03$) was found between metacognitive strategy use and language course grades. However, a negative correlation of $r = -.32$ ($p < .005$) was found between affective strategy use and language entrance examination scores, which are different from language placement scores in this particular Thai university. It is possible that students who are very anxious and who resort to affective strategies do less well on the entrance examination.

As shown by Dreyer and Oxford (1996), approximately 45% of the total variance in language proficiency (*TOEFL* scores) in a South African ESL study was explained by learning strategy use as measured by the *SILL*. A regression analysis demonstrated that the greatest part of the variance stemmed from metacognitive strategies, with much smaller amounts contributed by affective and social strategies. Canonical correlation showed a highly significant relationship between the parts of the *TOEFL* and the categories on the *SILL* ($r = .73$). The sample consisted of 305 Afrikaans first-year university students learning ESL in South Africa (Dreyer, 1992).

What we can learn about construct validity of the SILL based on relationships with language performance

ESL/EFL *SILL* strategy use is related, as expected, to language performance in a number of studies, thus providing construct validity evidence for the *SILL*. (These results agree with earlier research using varied strategy-assessment instruments; for instance, Corrales & Call, 1989; Huang, 1984; Oxford & Nyikos, 1989; O'Malley & Chamot, 1990). In many, but not all instances, the relationship is linear, showing that more advanced or more proficient students use strategies more frequently.

Construct Validity in Strategy Use in Foreign versus Second Language Environments

Some existing *SILL* data indicate that strategy-use patterns often differ between ESL and EFL settings. ESL environments typically show high frequencies of use for at least half of the strategy categories. For example, Oxford, Nyikos, Nyikos, Lezhnev, Eyring, and Rossi-Le (1989) found high frequencies of use for 60% of the strategies on the *SILL* as used by 159 ESL learners in the U.S. Rossi-Le (1989) learned that among 147 adult

ESL learners in two community colleges in the U.S., high frequencies existed for most of the strategies. Oxford, Talbott, and Halleck (1989), with a sample of 43 ESL learners at a large Northeastern university in the U.S., discovered high levels of strategy use for two-thirds of the strategy categories. Phillips (1990, 1991), studying 141 adult ESL learners in seven western states, noted that half the strategy categories were used at a high level. All four of these ESL studies reflected large amounts of high-frequency strategy use.

In contrast, five EFL studies (Klassen [1994], 228 Taiwanese university students; Noguchi [1991], 174 Japanese junior high students; Oh [1992], 59 Korean university students; Park [1994], 332 Korean university students; Yang [1994], 68 Taiwanese university students) indicated that strategy use in these settings was mostly at a medium level, far different from ESL strategy use in the studies noted earlier. In a study in Puerto Rico, where English was not the major language of daily communication but was nevertheless highly available as input (a hybrid ESL/EFL setting), strategy use among 213 university students learning English was much more like the EFL settings than the ESL environments.

Thus we can see that a *second* language environment, which demands daily use of the target language, often calls for (or encourages) more frequent strategy use than a *foreign* language environment, which does not require continual use of the target language. This is a sound generalization for most language students.

However, career interests can override this generalization. For example, Mullins (1992) found that her 110 Thai university EFL students had a rather high level of strategy use; and these students, unlike most of the EFL students mentioned in other studies, were majoring in English and wanted to use English in their careers. Likewise, Davis and Abas' (1991) 64 EFL students had mostly high strategy use; and they were already language teachers. Thus, strong language-related career interests can transform the target-language-impoverished EFL setting and make it a thriving home for language learning strategies.

What we can learn about construct validity of the SILL based on relationships with ESL/EFL setting

In short, unless foreign language students are extremely motivated because of their language-career interests, they will use strategies with less frequency than second-language students. Second-language students have more daily need to use the language, and therefore it is likely that they have greater motivation to use strategies. (Strategy use and motivation have elsewhere been shown to be statistically related by Oxford & Nyikos, 1989.)

Construct Validity in the Relationship Between Strategies and Learning Styles

Strong relationships between learning strategy use and sensory preferences—often viewed as an aspect of learning style—have been posited (Oxford, Ehrman, & Lavine, 1991) as partial evidence of the construct validity of the *SILL*. According to Oxford, Ehrman, and Lavine, visual students use strategies involving reading alone in a quiet place or paying attention to blackboards, movies, computer screens, and other forms of visual stimulation. Auditory students are comfortable without visual input and frequently use strategies that encourage conversation in a noisy, social environment with multiple sources of aural stimulation. Kinesthetic students need movement strategies, and tactile students require strategies that involve manipulating real objects in the classroom; both types need to use the strategy of taking frequent breaks.

ESL/EFL *SILL* data exist supporting the link between learning strategy use and learning styles, thus at the same time strengthening the evidence of construct validity of the *SILL*. Rossi-Le (1989) found a significant relationship ($p < .0005$) between learning styles (visual, auditory, tactile, and kinesthetic) and overall strategy use on the ESL/EFL *SILL* through a MANOVA, and she also found significant predictive relationships through multiple regression.

Rossi-Le's MANOVA results showed that visual learners preferred visualization strategies ($p < .0005$). Auditory-style learners used memory strategies more than did other learners ($p < .0005$). Compared with others, tactile learners demonstrated significant use of strategies for searching for and communicating meaning ($p < .006$) and self-management/metacognitive strategies ($p < .02$). Kinesthetic learners did not use general study strategies ($p < .003$) or self-management/metacognitive strategies ($p < .02$) as often as others did.

The regression results indicated that a visual learning style predicted using visualization strategies (beta =.33, $p < .00005$). Being a visual learner, however, negatively predicted using independent strategies (beta = -.22, $p < .001$), affective strategies (beta = -.23, $p < .009$), and strategies for searching for and communicating meaning (beta =-.22, $p < .008$). Having an auditory learning style significantly predicted memory strategies (beta = .38, $p < .0008$) and self-management or metacognitive strategies (beta = .20, $p < .01$) but was a negative predictor of employing authentic language-use strategies (beta = -.20, $p < .01$). Being a tactile learner significantly predicted employing authentic language-use strategies (beta = .26, $p < .001$) and strategies for meaning (beta =.32, $p < .0002$) but negatively predicted use of memory strategies (beta = -.16, $p < .04$). A kinesthetic learning style predicted infrequent use of general study strategies (beta = -.32, $p < .002$). Thus, these predictions are low-to-moderate and significant.

What we can learn about construct validity of the SILL based on relationships with learning styles

Though existing evidence is sparse, the data we have indicate that learning strategy use is related to learning styles. It is as though learning styles are the underlying or internal construct, and learning strategies are the more "outward" manifestation of learning styles.

The relationship is by no means simple, however. Predictions of strategy use according to learning style are sometimes straightforward (e.g., visual learning style predicts visualization strategy use) and sometimes not so straightforward (e.g., auditory learning style predicts metacognitive strategy use). Clearly more information is needed on the links between learning styles and learning strategies.

Construct Validity in the Relationship Between Strategy Use and Gender

In many ESL/EFLstrategy studies, results have usually favored females as more frequent users of strategies (for instance, Dreyer, 1992; Ehrman & Oxford, 1989; Green, 1991, 1992; Green & Oxford, 1993, 1995; Noguchi, 1991; Oxford, 1993a, 1993b; Oxford, Ehrman, & Nyikos, 1988; Oxford, Park-Oh, Ito, & Sumrall, 1993a, 1993b; Yang, 1992b, 1993). In a few studies, females have had a distinctly different pattern of strategy use from that of males (Bedell, 1993; Watanabe, 1990). Some studies, noted by Bedell and by Green and Oxford, have shown that males surpassed females on a certain number of separate strategies but not on whole clusters or groups of strategies.

What we can learn about construct validity of the SILL based on relationships with gender

Overall, the last decade of studies has shown that females are generally more frequent strategy users than men in a language learning situation. This trend fits in with previous theory and research about females as better, more efficient learners and users of language (native or other) than males; see Oxford (1993a, 1993b) for many biosocial reasons for this difference. Thus, the construct validity of the *SILL* has additional evidence.

Other Aspects of Validity: Utility, Value Implications, Social Consequences,

Interpretation, and Real-World Action

According to Messick (1989) and Chapelle (1994), aspects of general validity (in addition to construct validity) include utility, value implications, social consequences, interpretation, and real-world action.

Utility can be defined as the *usefulness of an instrument in real-world settings for making decisions relevant to people's lives.*The *SILL* has utility, deemed to be a crucial piece of evidence of general validity of the instrument. Utility is demonstrated by the many people around the world who have employed the *SILL* and by the uses to which they have put it. The most frequent venue of use has been the classroom, where the goal has been chiefly to reveal the relationship between strategy use and language performance. This goal is important because if there is a strong relationship between these two variables, perhaps language performance can be improved by enhancing strategy use.

Related to utility are the value implications and social consequences of the questionnaire. Underlying the *SILL* is the value of learning strategies as tools for learner self-direction, autonomy, and achievement. Strategies are a means of enhancing learning for each student. Every student uses strategies, but some strategies are more appropriate than others to a given task and to the student's own learning style (visual, auditory, hands-on; extroverted, introverted; and so on). The social consequences of using the *SILL* are that learners (and their teachers) become more aware of the strategies each learner typically uses. This awareness helps teachers more effectively design language instruction and enables them to provide relevant strategy instruction. This awareness also helps students seek and experiment with new and more efficient strategies.

Interpretation and real-world action relate to outcomes of testing. Interpretation of the *SILL* should be limited to "typical" strategies of a given student (or, when aggregated, strategies of a given group) in a variety of situations and tasks. It should not be applied to assess the strategies used for a single activity, such as the very task-bound strategies that Marilyn uses to read aloud in French the first part of Chapter 3 of *Madame Bovary*.

Real-world actions based on the appropriate interpretation of the *SILL* include increased theory-building concerning the nature of language learning strategies; assessing strategy use at a given point, to be compared with strategy use later (sometimes after strategy improvement interventions); comparing strategy use with proficiency or achievement; comparing the learning strategies of women and men; making the conceptual linkage between strategy use and learning styles; and individualizing classroom instruction based on the strategy use of different students. So far the utility of the *SILL* has not included making placements of individuals into language classes on the basis of strategy use results, although such strategy information could conceivably be combined with other kinds of data for making such placements. See the reference list for dozens of studies showing various applications of the *SILL*.

Appropriate Uses and Limitations of Questionnaires Compared with Other Strategy Instruments

Compared with the other strategy assessment techniques (see Table 1), *student-completed strategy questionnaires* have a very important and appropriate use. These questionnaires provide a general assessment of each student's typical strategies across a variety of possible tasks. However, strategy questionnaires do not describe in detail the language learning strategies a student uses in response to any *specific* language task (as do some specific-strategy interviews or think-aloud protocols). For a researcher or a teacher who wants to discover strategy use on a particular reading comprehension task in a given classroom on Monday morning, a general strategy questionnaire like the *SILL* would not be useful. It is a misuse of the *SILL* (or any other strategy questionnaire) to try to identify task-specific strategies with that instrument.

TABLE 1

Comparisons of strategy-assessment types

Type of assessment	Appropriate uses	Limitations of use
Strategy questionnaires	Identify "typical" strategies used by an individual; can be aggregated into group results; wide array of strategies can be measured by questionnaires	Not useful for identifying specific strategies on a given language task at a given time
Observations	Identify strategies that are readily observable for specific tasks	Not useful for unobservable strategies (e.g., reasoning, analyzing, mental self-talk) or for identifying "typical" strategies
Interviews	Identify strategies used on specific tasks over a given time period or more "typically" used strategies; usually more oriented toward task-specific rather than "typical" strategies of an individual; depends on how interview questions are asked	Usually less useful for identifying "typical" strategies because of how interviews are conducted, but could be used for either task-specific or "typical" strategies
Dialogue journals, diaries	Identify strategies used on specific tasks over a given time period	Less useful for identifying "typical" strategies used more generally
Recollective narratives (language learning histories)	Identify "typical" strategies used in specific settings in the past	Not intended for current strategies; depends on memory of learner
Think-aloud protocols	Identify in-depth the strategies used in a given, ongoing task	Not useful for identifying "typical" strategies used more generally

Strategy checklists	Identify strategies used on a just-completed task	Not useful for identifying "typical" strategies used more generally

Strategy questionnaires have certain advantages. They are quick and easy to administer, may be the most cost-effective mode of strategy assessment, and are almost completely nonthreatening when administered using paper and pencil (or computer) under conditions of confidentiality. Moreover, many students discover a great deal about themselves from taking a strategy questionnaire, especially one like the *SILL* that is self-scoring and that provides immediate learner feedback.

An advantage specifically accruing to the *SILL* is that this questionnaire is free of social desirability response bias. Social desirability response bias, or the tendency to answer in a way that the researcher "wants" one to answer, is usually identified by a moderate to high correlation between the *Marlowe-Crown Social Desirability Scale* and a given instrument like the *SILL*. No such correlation appeared in a large-scale study by Yang (1992b), in which the researcher tested 505 students of EFL on the *SILL* and the *Marlowe-Crown*. Therefore, students appeared to express themselves freely and openly on the *SILL*. In other studies, the current author compared results of informal strategy interviews with the way that respondents answered on the *SILL* and found that respondents had answered the *SILL* honestly (Oxford, 1986). At this writing, no other language learning strategy questionnaire has been studied for social desirability response bias.

Implications for Research and Instruction

First, language researchers must conceptualize language learning strategies in a way that includes the social and affective sides of learning (as shown in the *SILL*) as well as the more intellectual and "executive-managerial" sides. Language learning is not just cognitive and metacognitive. It involves much more from the learner.

Second, through strategy assessment teachers can help their students recognize the power of using language learning strategies for making learning quicker, easier, and more effective. Teachers need to know the appropriate uses and limitations of each assessment technique, as seen in the previous section. Multiple techniques are to be encouraged whenever the time and resources are available. When time and resources are restricted, teachers should use the most reliable and valid strategy assessment measure that they can for the purposes they have defined. When the purposes include tapping the "typical" or general strategy use of an individual student or a group, strategy questionnaires like the *SILL* can be extremely helpful. If much more precise measurement of highly task-based strategy use is the purpose, then other measurement tools are required.

Third, based on the information from strategy assessment, teachers can weave strategy instruction into regular classroom events in a natural, comfortable, but explicit way. Chamot and Kupper (1989), Oxford (1990b), and O'Malley and Chamot (1990) provide helpful details on how to do this. Teachers must also keep in mind differences in motivation, learning style, gender, and other factors that affect learning strategy use.

Fourth, teachers need to be judicious in their selection of strategies to use in instruction, and existing research can provide good clues for this selection. Research indicates that some strategies in certain studies do *not* relate strongly to proficiency. For example, based on Takeuchi's research (1993a), it is possible to say that using flash

cards is clearly *not* a sure-fire strategy to promote proficiency in all cultures and for all kinds of learning styles. Flashcards might work for some students but not for others. On the other hand, research shows that paying attention and actively using the language for writing seem to be widely appropriate strategies in most contexts and for most kinds of learners.

Fifth, strategy assessments using different measurement modes with the same sample of students could be cross-correlated. This would contribute to the validity of various assessment techniques. For instance, it would be useful to correlate results from a think-aloud protocol, an interview, and a survey to see how closely they relate to each other. If results show that an interview and a survey are highly correlated but that they are only weakly correlated with a think-aloud procedure, this information would be useful in selection of an assessment procedure next time.

Sixth, studies will need to be replicated so that more consistent information becomes available within and across populations. Particularly important is more information on how students from different cultural backgrounds and different countries use language learning strategies. Teachers need to have more background on how to use such information in the classroom. Here is a clear opportunity for researchers to better translate their findings into materials to be used in the classroom.

In sum, it is critical that learning strategies be considered when planning courses, teaching students, and designing classroom research. Appropriate learning strategies should be among the first considerations of any ESL/EFL teacher or researcher who wants to enhance student learning.

References

Anderson, N. (1993). Data on adult ESL strategy use. Unpublished manuscript, Ohio University, Athens, OH.

Bedell, D. (1993a). Chinese data on EFL strategy use among university students. Unpublished manuscript, University of Alabama, Tuscaloosa, AL.

Bedell, D. (1993b). Crosscultural variation in the choice of language learning strategies: A mainland Chinese investigation with comparison to previous studies. Unpub- lished master's thesis, University of Alabama, Tuscaloosa, AL.

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning*, *28*, 69-83.

Bialystok, E. (1981). The role of conscious strategies in second language proficiency. *Modern Language Journal, 65*, 24-35.

Boraie, D., Kassabgy, O., & Oxford, R. (1994). Empowering teachers and learners: Style and strategy awareness. Paper presented at the annual meeting of Teachers of English to Speakers of Other Languages, Baltimore, MD.

Chamot, A. U., & Kupper, L. (1989). Learning strategies in foreign language instruction. *Foreign Language Annals, 22*(1), 13-24.

Chamot, A. U., O'Malley, J. M., Kupper, L., & Impink-Hernandez, M. (1987). *A study of learning strategies in foreign language instruction: First year report*. InterAmerica Research Associates, Rosslyn, VA.

Chang, S-J. (1991). A study of language learning behaviors of Chinese students at the University of Georgia and the relation of those behaviors to oral proficiency and other factors. Unpublished doctoral dissertation, University of Georgia, Athens, GA.

Chapelle, C. (1994). Are C-tests valid measures for L2 strategy research? *Second Lang- uage Research, 10* (2), 157-187.

Cohen, A. D. (1987). Studying learner strategies: How we get the information. In A. L. Wenden, & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 31-42). Englewood Cliffs, NJ: Prentice-Hall.

Corrales, O., & Call, M.E. (1989). At a loss for words: The use of communication strat- egies to convey lexical meaning. *Foreign Language Annals, 22*, 227-240.

Davis, E., & Abas, H. (1991). *Second language learning strategies utilized by some members of language departments at four institutions - Sulawesi, Indonesia*. Sulawesi, Indonesia: Summer Institute of Linguistics.

Dreyer, C. (1992). Learner variables as predictors of ESL proficiency. Unpublished doctoral dissertation, Potchefstroom University, South Africa.

Dreyer, C., & Oxford, R. (1996). Learner variables related to ESL proficiency among Afrikaans-speakers in South Africa. In R. L. Oxford (Ed.), *Language learning strategies around the world*. Manoa: University of Hawai'i Press.

Ehrman, M. E., & Oxford, R. L. (1989). Effects of sex differences, career choice, and psychological type on adult language learning strategies. *Modern Language Journal, 73*, 1-13.

Ehrman, M. E., & Oxford, R. L. (1990). Adult language learning styles and strategies in an intensive training setting. *Modern Language Journal, 74*(3), 311-327.

Green, J. M. (1991). Language learning strategies of Puerto Rican university students. Paper presented at the annual meeting of Puerto Rico Teachers of English to Speakers of Other Languages, San Juan, PR.

Green, J. M. (1992). Additional analyses of Puerto Rican strategy data. Unpublished manuscript, University of Puerto Rico at Mayaguez.

Green, J., & Oxford, R. (1993). New analyses on expanded Puerto Rican strategy data. Unpublished manuscript, University of Puerto Rico at Mayaguez and University of Alabama, Tuscaloosa, AL.

Green, J., & Oxford, R. (1995). A closer look at language learning strategies, L2 profi- ciency, and gender. *TESOL Quarterly*, *29*, 261-297.

Huang, X. H. (1984). An investigation of learning strategies in oral communication that Chinese EFL learners in China employ. Unpublished master's thesis, Chinese Uni- versity of Hong Kong.

Huang, X. H., & Van Naerssen, M. (1987). Learning strategies for oral communication. *Applied Linguistics, 8,* 287-307.

Keefe, J. W., & Monk, J. S., with Letteri, C. A., Languis, M., & Dunn, R. (1989). *Learning style profile*. Reston, VA: National Association of Secondary School Principals.

Kim, J-D. (1991). A comparison of learning strategies of college students enrolled in beginning and advanced English as a second language. Unpublished doctoral dissertation, Loma Linda University, Loma Linda, CA.

Klassen, J. (1994). The language learning strategies of freshman English students in Taiwan: A case study. Master's thesis, California State University at Chico, CA.

Lawrence, G. (1984). A synthesis of learning style research involving the *MBTI. Journal of Psychological Type,* 2-15.

LoCastro, V. (1994). Learning strategies and learning environments. *TESOL Quarterly, 28,* 409-414.

McGroarty, M. (1987). Patterns of persistent second language learners: Elementary Spanish. Paper presented at the annual meeting of Teachers of English to Speakers of Other Languages, Miami, FL.

Messick, S. (1989). Validity. In R. E. Linn (Ed.), *Educational measurement* (3rd ed.). New York: Macmillan.

Mullins, P. (1992). Successful English language learning strategies of students enrolled in the Faculty of Arts, Chulalongkorn University, Bangkok, Thailand. Unpublished doctoral dissertation, United States International University, San Diego, CA.

Myers, I., & McCaulley, M. (1985). *A guide to the development and use of the Myers-Briggs Type Indicator.* Palo Alto, CA: Consulting Psychologists Press.

Noguchi, T. (1991a). Review of language learning strategy research and its implications. Unpublished bachelor's thesis, Tottori University, Tottori, Japan.

Noguchi, T. (1991b). *Questionnaire for learners*. Tottori University, Tottori, Japan.

Nyikos, M., & Oxford, R. (1993). A factor analytic study of language-learning strategy use: Interpretations from information-processing and social psychology. *Modern Language Journal, 77,* 11-22.

Oh, J. (1992). Learning strategies used by university EFL students in Korea. *Language Teaching, 1,* 3-53.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.

Oxford, R. L. (1986). *Development and psychometric testing of the Strategy Inventory for Language Learning (SILL).* ARI Technical Report 728. Alexandria, VA: Army Research Institute. Appendices as ARI Research Note 86-92.

Oxford, R. L. (1986-1990). *Strategy Inventory for Language Learning*. Various ver- sions. Tuscaloosa, AL: Oxford Associates.

Oxford, R. L. (1989). Use of language learning strategies: A synthesis of studies with implications for strategy training. *System, 17*, 235-247.

Oxford, R. L. (1990a). Language learning strategies and beyond: A look at strategies in the context of styles. In S. S. Magnan (ed.), *Shifting the instructional focus to the learner* (pp. 35-55). Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages.

Oxford, R. L. (1990b). *Language learning strategies: What every teacher should know*. New York: Newbury House / Harper & Row. *Now* Boston: Heinle & Heinle.

Oxford, R. L. (1993a). Instructional implications of gender differences in language learning styles and strategies. *Applied Language Learning, 4*(1-2), 65-94.

Oxford, R. L. (1993b). *La différence continue . . . :* Gender differences in second/foreign language learning styles and strategies. In J. Sutherland (Ed.), *Exploring gender* (pp. 140-147). Englewood Cliffs, NJ: Prentice-Hall.

Oxford, R. L., & Burry, J. A. (1993). Evolution, norming, and psychometric testing of Oxford's *Strategy Inventory for Language Learning (SILL)*. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.

Oxford, R. L., & Burry-Stock, J. A. (1995). Assessing the use of language learning strategies worldwide with the ESL/EFL version of the *Strategy Inventory for Lang- uage Learning*. *System, 23*(1), 1-23.

Oxford, R. L., & Ehrman, M. E. (1989). Psychological type and adult language learning strategies: A pilot study. *Journal of Psychological Type, 16*, 22-32.

Oxford, R. L., & Ehrman, M. E. (1995). Adults' language learning strategies in an intensive foreign language program in the United States. *System, 23*(3), 359-386.

Oxford, R. L., Ehrman, M. E., & Lavine, R. Z. (1991). Style wars: Teacher-student style conflicts in the language classroom. In S. S. Magnan (Ed.), *Challenges in the 1990s for college foreign language programs* (pp. 1-25). Boston: Heinle & Heinle.

Oxford, R. L., & Green, J. M. (1995). Making sense of learning strategy assessment: Toward a higher standard of research accuracy—a reply to LoCastro's article. *TESOL Quarterly, 29*, 166-171.

Oxford, R. L., & Nyikos, M. (1989). Variables affecting choice of language learning strategies by university students *Modern Language Journal, 73*(2), 291-300.

Oxford, R. L., Nyikos, M., & Ehrman, M. (1988). *Vive la différence?* Reflections on sex differences in use of language learning strategies. *Foreign Language Annals, 21*(4), 321-329.

Oxford, R. L., Nyikos, M., Nyikos, K., Lezhnev, V., Eyring, J., & Rossi-Le, L. (1989). Learning strategies of adult ESL learners in the U.S.: A collaborative study. Unpub- lished manuscript, University of Alabama, Tuscaloosa, AL.

Oxford, R. L., Park-Oh, Y., Ito, S., & Sumrall, M. (1993a). Factors affecting achievement in a satellite-delivered Japanese language program. *American Journal of Distance Education, 7*, 10-25.

Oxford, R. L., Park-Oh, Y., Ito, S., & Sumrall, M. (1993b). Learning a language by satellite: What influences achievement? *System, 21*(1), 31-48.

Oxford, R. L., Talbott, V., & Halleck, G. (1990). Language learning strategies, attitudes, motivation, and self-image of students in a university intensive ESL program. Paper presented at the annual meeting of International Teachers of English to Speakers of Other Languages, San Francisco, CA.

Padron, Y. N., & Waxman, H. C. (1988). The effects of ESL students' perceptions of their cognitive strategies on reading achievement. *TESOL Quarterly, 22*, 146-150.

Park, G. (1994). Language learning strategies: Why do adults need them? Unpublished manuscript, University of Texas at Austin.

Phillips, V. (1990). English as a second language learner strategies of adult Asian students using the *Strategy Inventory for Language Learning*. Unpublished doctoral dissertation, University of San Francisco, San Francisco, CA.

Phillips, V. (1991, Nov.). A look at learner strategy use and ESL proficiency. *CATESOL Journal*, 57-67.

Politzer, R. (1983). An exploratory study of self-reported language learning behaviors and their relation to achievement. *Studies in Second Language Acquisition, 6*, 54-65.

Politzer, R., & McGroarty, M. (1985). An exploratory study of learning behaviors and their relationship to gains in linguistic and communicative competence. *TESOL Quarterly, 19*, 103-124.

Rossi-Le, L. (1989). Perceptual learning style preferences and their relationship to language learning strategies in adult students of English as a second language. Unpublished doctoral dissertation, Drake University, Des Moines, IA.

Talbott, V. (1993). Strategies of adult ESL students: Data. Unpublished manuscript, Skagit Valley Junior College, Washington.

Takeuchi, O. (1991a). A study of EFL students' use of language learning strategies. *Gohokenkyu to Eigokoyoiku (Studies on English Language Usage and Language Teaching), 13*, 58-65.

Takeuchi, O. (1991b). Language learning strategies in second and foreign language of strategies frequency. *Bulletin of the Institute for Interdisciplinary Studies of Culture, 8*, 64-83.

Takeuchi, O. (1993a). A study of language learning strategies and their relation to achievement in EFL listening comprehension. *Bulletin of the Institute for Interdisciplinary Studies of Culture, 10*, 131-141.

Takeuchi, O. (1993b). Language learning strategies and their relationship to achieve- ment in English as a foreign language. *Language Laboratory, 30*, 17-34.

Wangsotorn, A., Sripaipan, T., Rattanaprucks, N., Jarunggidanan, T., Singkalwanij, P., & Vejaphurti, A. (1986). Relationships between learning modes and the beginners' success in English. Paper presented at the annual meeting of International Teachers of English to Speakers of Other Languages, Anaheim, CA.

Watanabe, Y. (1990). External variables affecting language learning strategies of Japanese EFL learners: Effects of entrance examination, years spent at college/university, and staying overseas. Unpublished master's thesis, Lancaster University, Lancaster, U.K.

Weinstein, C. E., Palmer, D., & Schulte, A. C. (1987). *Learning and Study Strategies Inventory*. Clearwater, FL: H & H Publishers.

Wen, Q., & Johnson, R. K. (1991). Language learning approaches and outcomes: A study of tertiary English majors in China. Paper presented at the Sixth International Conference of the Institute of Language in Education, Hong Kong, China.

Wenden, A., & Rubin, J. (Eds.). (1987). *Learner strategies in language learning*. Englewood Cliffs, NJ: Prentice-Hall.

Wildner-Bassett, M. (1992a). Data set of foreign language students' strategies. Unpublished manuscript, University of Arizona, Tucson, AZ.

Wildner-Bassett, M. (1992b). Relationships among language learning strategies, per- sonality types, and learning styles. Paper presented at the annual meeting of the American Council on the Teaching of Foreign Languages, Chicago.

Yang, N-D. (1992a). Data on students in an intensive EFL program in Taipei. Unpub- lished manuscript, National Taiwan University.

Yang, N-D. (1992b). Second language learners' beliefs about language learning and their use of learning strategies: A study of college students of English in Taiwan. Unpublished doctoral dissertation, University of Texas, Austin, TX.

Yang, N-D. (1993). Understanding Chinese students' language beliefs and learning strategy use. Paper presented at the annual meeting of International Teachers of English to Speakers of Other Languages, Atlanta, GA.

Yang, N-D. (1994). An investigation of Taiwanese college students' use of English learning strategies. Research report, National Taiwan University, Taiwan.

Zhang, W. (1994). Data on university EFL language use. Unpublished manuscript, Ohio University, Athens, OH.

Author

REBECCA OXFORD, Associate Dean, College of Education, University of Alabama, 201 Carmichael Hall, Tuscaloosa, AL 35487. Specializations: language learning strategies, styles, motivation, and anxiety; psychometrics.

[Table of Contents]

[Table of Contents]

Applied Language Learning
1996, Vol. 7, Nos. 1 & 2, pp. 47-60

Validity Issues in Computer-Assisted Strategy Assessment for Language Learners

Carol A. Chapelle
Iowa State University

This article probes validity issues of computer-assisted strategy assessment (CASA) in second-language (L2) acquisition research. Using two CASA studies, it reviews the metacognitive and cognitive strategies. It discusses trait-oriented and interactionist approaches to the concept of strategy, and examines validity by focusing on "advance preparation" and "resourcing," examples of trait-oriented and interactionist strategies, respectively. It suggests additional forms of validity evidence and describes applications of CASA as well as problems associated with this form of measurement.

Researchers of L2 acquisition and computer-assisted language learning (CALL) have exploited computer technology as means of gathering performance data of learners working on language activities. These data are then used to make inferences about learners' linguistic competence (e.g., Bland, Noblitt, Armington, & Gay, 1990) and strategies for L2 acquisition. As computer technology becomes more widespread throughout educational settings, these research practices are likely to expand in their popularity. In anticipation that more researchers will be interested in CASA for L2 learners in the future, this article draws upon past experience to summarize pertinent issues.

CASA Research

Throughout the 1980's and early 1990's, a number of researchers have used computer-assisted methods for strategy research by collecting data generated by learners while using interactive software. In such studies, the recorded data were generated through the learner input to the computer program and the program responses to the learner. This work has not been reexamined and interpreted from a measurement perspective to identify common issues across individual studies and to define "CASA" as a method.

Although there are many possible roles for the computer in the learning environment and therefore in strategy research (e.g., Chun, 1994), the basic learner-computer interaction that occurs during individualized work has formed the basis for most L2 strategy research to date. I will focus on assessment of strategies familiar from other L2 research rather than the fundamental cognitive processes that have been the object of inquiry in psychology laboratories for many years.[1]

When a computer program is used to control or facilitate a language task, it can unobtrusively document learners' behavior such as the time they spend at various points in a problem sequence, the order in which they complete steps, and the editing they do to produce a linguistic product. Such a program can collect as sequenced data everything the learner sees on the screen, all keypresses and mouse-clicks that the learner

makes, and the time that each action takes place. For the most part, software in such research has been designed or modified to capture and store the desired information. The question for the researcher is which of the obtainable data can be used as indicators of learner capacities. The CASA studies listed in Table 1 provide some tentative answers to this question.

One study of English as a second language (ESL) learners (Jamieson & Chapelle, 1987) investigated "advance preparation," a metacognitive strategy defined as "planning for and rehearsing linguistic components necessary to carry out an upcoming language task" (O'Malley et al., 1985, p. 33). Using ESL dictation tasks over the course of six weeks, learners listened to words (on the "spelling" task) and individual sentences (on the "dictation" task) and then typed what they had heard. In this setting, it was assumed that the time the learner spent after hearing the input and before responding was spent planning performance; therefore, advance preparation was inferred from the amount of time that elapsed between the end of the input signal and the time that the learner pressed the first key to begin to answer. The actual indicator of the degree to which each learner used advance preparation was obtained by having the computer store the time it took to respond to each item, and calculating the mean "time-to-begin" by dividing each learner's total by the total number of items she or he had completed. It is important to note that the learners in this study were not instructed to complete the exercise as quickly as possible; they were working routinely over the course of several weeks.

Another metacognitive strategy, monitoring output, was assessed in the same instructional setting. On the dictation tasks, learners were able to edit the response they had typed by deleting, inserting, and changing characters or words before the response was evaluated by the computer. This behavior was documented in the computer records; it was considered an indicator of monitoring output, in the sense described by Krashen (1982), Bialystok (1981), Wenden (1985), and O'Malley et al. (1985). The number of times a learner edited was divided by the total number of completed items to obtain the average number of times a learner edited each item.

A third metacognitive strategy investigated with the same software was "monitoring input," defined by Bialystok (1981) as, reflecting on the formal aspects of a message as it was comprehended. The dictation tasks allowed learners to listen to the input as many times as they wanted before typing it. The demands of the task required learners to focus on formal aspects of the input. When they had not comprehended a sentence or word the first time it was presented, they requested to hear it again one or more times. Subsequently, those requests were used as evidence for monitoring input. In other words, "monitoring input" was inferred on the basis of observed requests.

TABLE 1

Studies using computer-assisted strategy assessment

Assessment context		**Learner**	
Task goal	**Behavior**	**Learner's assessment of task***	**Strategy**

construct & edit sentences [1]	completing extra sentences and experimenting	instruction	exploration
construct & edit sentences [2]	requesting any help	instruction	resourcing
read & answer questions or summarize [3]	requesting dictionary responses	experiment definitions	resourcing
complete grammar exercise [4]	requesting rule / example help	instruction	inferencing / hypothesis-testing
complete grammar exercise [4]	requesting / not requesting help	instruction	controlled / automatic processing
dictation: listen & write word / sentence [5]	latency before responding	instruction	advance preparation
dictation: listen & write word / sentence [5]	editing response	instruction	monitoring output
dictation: listen & write word / sentence [5]	requesting repeated input	instruction	monitoring input

[1] Hsu, Chapelle, & Thompson (1993)
[2] Chapelle & Mizuno (1989)
[3] Hulstijn (1993)
[4] Doughty (1987)
[5] Jamieson & Chapelle (1987)

* This refers to the learner's perception of what he/she is doing while participating in the activity. The learner's perception is believed to affect performance, and therefore should be considered in the interpretation of the performance. For example, "instruction" means the learner perceived the task as taking place for instructional purposes.

Other interaction observation programs have been used to measure "resourcing" defined by O'Malley, et al. (1985) as the cognitive strategy of using reference materials to obtain information about the L2. Consistent with this definition, researchers have considered each request for on-line help as evidence for the use of resourcing. In one study (Hulstijn, 1993), learners had access to an on-line dictionary that they could consult while reading a passage followed by questions. The resulting data documented the words that they had looked up in the dictionary and the time that each was looked up. Also while investigating resourcing, Chapelle and Mizuno (1989) collected data as learners worked on a task requiring them to construct and edit sentences in the exercises. As learners worked, help was given only at their request; they could ask for help with vocabulary, grammar, or the semantic facts pertaining to the sentences.

Doughty and Fought (1984) also documented learners' help requests on grammar items but interpreted them differently. Learners' requests for help while they were used as indicators of learners' "controlled access of explicitly learned knowledge." In addition, the researchers reasoned that "attempts to complete tasks without any help from the program [they reasoned] reflect automatic access to implicit knowledge in memory" (Doughty, 1987, p. 151). Other strategies evidenced by learners using particular types of help in software were

"hypothesis-testing" and "inferencing based on L1." Doughty and Fought operationalized the definitions of these strategies as the type of grammar help requested by learners. When learners consulted help consisting of examples, they were considered to be displaying evidence of "hypothesis-testing." When they chose grammar help consisting of formal rules or the correct answer, learners were considered to be "inferencing [about syntactic forms in the L2] based on L1."

Using a sentence-constructing and -editing task, Hsu, Chapelle and Thompson (1993) assessed another strategy—exploration, the use of software to experiment and test hypotheses about the target language (Higgins & Johns, 1984). Exploration was operationalized in two ways: the number of sentences learners constructed after having completed the number required by their assignment and the number of times they edited an answer after receiving a message that it was correct—which the software allowed but did not require them to do. The operational definition of exploration in this study was derived from the theoretical definition that included the notion that learners would be motivated and interested in experimenting with the software.

In summary, a number of strategies have been investigated using computer-assisted methods. To understand CASA as a method, however, it is informative to interpret the types of strategies from a measurement perspective and, consequently, to explore the nature of the strategy construct.

Definition of "Strategy" in Computer-Assisted Strategy Research

To examine this method of assessment, I consider strategies to be theoretical constructs that are themselves not observable, but are hypothesized to be responsible for observed behavioral data.[2] Since there is more than one way to define a theoretical construct, I examine carefully how a particular strategy such as "monitoring" is defined in a research study. By looking for similarities in strategy definition across computer-assisted research studies, I query the extent to which an approach to strategy definition is tied to CASA. I provide means for considering the "nature of the construct," by distinguishing two approaches to construct definition. I then illustrate computer-assisted strategy research that has taken each approach to construct definition, thereby demonstrating that the method of measurement does not preclude either one.

I will distinguish between two approaches defining a strategy as a theoretical construct (approaches explained in Chapelle, forthcoming). The first is a *trait-oriented* definition that conceptualizes a strategy as an attribute of an individual independent of the context in which it is observed. One thinks of a strategy as a trait when one talks about "monitoring" as something learners do all of the time regardless of whether they are listening to an academic lecture, writing an e-mail message to a colleague, or speaking to a close friend. A trait-oriented construct definition assumes that a researcher is able to generalize the inferences made about a construct on the basis of performance on an assessment (i.e., performance in one context) to inferences about the construct in other contexts. In other words, if an individual is a strong monitor user in a test of monitoring, the trait definition would assume that the individual would also be a strong monitor user in the other contexts, such as instructional settings.

A second and contrasting way of defining a strategy as a theoretical construct is an *interactionalist* definition. This definition presents a strategy as a context-dependent attribute of an individual. From an interactionalist approach, one could not define "monitoring" in a global sense. Instead, one would refer to "monitoring while listening to academic lectures," for example.[3] The definition of the strategy would include the context in which the strategy is used.[4] To interpret results of a test of "monitoring while listening to academic lectures," the

researcher would generalize results only to monitoring in this context. In short, an important distinction between the two approaches rests on how far the strategy definition assumes the researchers can generalize the results of strategy assessment.

Both approaches to strategy definition have been used in computer-assisted strategy research. The assessment of "advance preparation" illustrates a trait-oriented definition in computer-assisted strategy research (Jamieson & Chapelle, 1987). The strategy is defined in a general way; even though the definition mentions the word "task," it does not refer to any particular task, implying that the strategy is conceived as one that could apply equally to a linguistic task in any context. Table 2 summarizes the key measurement facets of this strategy.

TABLE 2

Examples of trait and interactionalist approaches to definitions of L2 strategies

	Trait-oriented definition	**Interactionalist definition**
Example	advance preparation (Jamieson & Chapelle, 1987)	resourcing (Chapelle & Mizuno, 1989)
Definition	"planning for and rehearsing linguistic components necessary to carry out an upcoming linguistic task" (O'Malley, Chamot, Stewner-Manzanares, Kupper, and Russo, 1985)	"use of target language reference materials" (pp. 28-29) in the context of learner-controlled CALL materials (p. 26)
Measurement	the amount of time (to .5 second accuracy) between the time that a prompt was given (in a CALL activity) and the time that the student began to respond (averaged over the number of items that student responded to over the course of the semester)	he frequency of the number of requests for help a student made divided by the number of sentences the student produced in a sentence constructing and editing CALL activity (help=dictionary, semantic/pragmatic facts, and grammar)
Inference	performance was assumed to indicate the degree to which the learner was an "advance preparer"	performance was assumed to indicate the degree to which learners used resourcing within the learner-controlled CALL activity
Use	to investigate the relationship between advance preparation and cognitive style as well as the relationship between advance preparation and subsequent language proficiency	to investigate the extent to which learners use resourcing in a set of learner- controlled CALL activities for practicing grammar and editing (as a means of evaluating the pedagogical potential of optional help)

Advance preparation was measured through response latency in an instructional setting in which learners were working at their own pace. The inferences made on the basis of summed response latencies were the degree to which the language learners were advance preparers in general, rather than the degree to which they used

advance preparation while working on this type of software. Accordingly, the scores for advance preparation were used in this research context to investigate the relationship between this strategy and other variables, which were also defined in a context-independent manner.

The assessment of "resourcing" provides an example of an interactional approach to strategy definition (Chapelle & Mizuno, 1989). It is defined in this research as a learner's use of target language reference materials in learner-controlled CALL materials. The definition is interactionalist because it includes the "learner-controlled CALL materials" as the context to which we wish to generalize. Measurement of the strategy was calculated by tabulating the number of times the learners requested help per unit of activity (defined by construction of one sentence). The inferences made were intended to be limited to contexts of learner-controlled CALL, and the scores were used to evaluate the value of offering learners optional help in learner-controlled CALL.

Justifying Inferences about Strategies from Observed Performance Data

In the research cited above, as in any research, one is ultimately concerned about the validity of the uses that are made of the strategy assessment. Validity of test use rests on justifying the inferences made from observed behaviors. Justifications of inferences about strategies constitute evidence for the construct validity of those inferences. Table 3 summarizes and defines types of construct-validity evidence suggested by Messick (1989).

To examine these types of construct-validity evidence, I return to the examples of advance preparation and resourcing. Then I will suggest for each example additional evidence that could be used to make a stronger case for construct validity. Finally, I will underscore two fundamental points about construct validity: (1) the nature of the evidence depends on the way "strategy" is defined in the research and (2) construct-validity evidence refers to the justifications provided for interpretations and therefore it can be evaluated as a strong or weak relative to particular inferences rather than as an all or nothing quality of an assessment.

In the first study, Jamieson and Chapelle (1987) provided three types of justifications that might be used to argue for inferring "advance preparation" from response-latency data. First, content-validity evidence consisted of the authors' judgment that this behavior in the instructional context logically fit with the definition of the construct:

> [A]dvance preparation was inferred from the amount of time it took for the student to press the first key of his or her answer. O'Malley et al. (1985) defined advance preparation as a metacognitive strategy that means "planning for and rehearsing linguistic components necessary for an upcoming language task" (p. 33). The student behavior of consistently waiting before answering may indicate the degree to which he or she was engaging in preparation to answer (p. 531).

TABLE 3

*Potential methods for justifying construct-validity of inferences from tests/assessments (based on Messick, 1989)**

Content analysis	Experts' judgments of what they believe a test measures

Empirical item investigation	Investigation of factors affecting item difficulty and discrimination to provide statistical evidence relevant to researchers' understanding of what the test measures
Internal consistency	Estimation of the consistency of learners' performance
Empirical task analysis	Documentation of the metacognitive strategies that learners use as they complete test tasks through qualitative methods such as think-alouds
Correlations with other tests and behaviors	Identification of theoretically predicted levels of covariance *and behaviors* among tests and behavior in appropriate contexts
Experimental studies	Investigation of changes in performance which accompany systematic changes in test design

* See Chapelle (1994) for an example of how these approaches to construct validity are applied to the evaluation of inferences and uses made from a test.

Second, the authors found a sufficient degree of consistency in strategy use over a six-week period to provide evidence that a construct was assessed through the observed data. "Mean working-style scores from six randomly selected weeks were treated like 6-item scales on which ... reliability estimates were made" (a=.72 and .82 on the two activities, p. 535). Third, a theoretically predicted correlation with a style variable "reflectivity-impulsivity" was found. Advance preparation was significantly, positively related to reflectivity ($r=.50$; $p<.001$); one would expect that this strategy "would logically be associated with the slow, careful aspect of the reflective learner" (p. 538).

These three arguments provide some justification for the advance preparation inferences made from learner performance data. However, the argument would be stronger if additional sources of evidence had been provided. First, ideally, evidence consisting of learner verbal reports might indicate that they were thinking about and planning what they would type between the time they heard the input and the time they began to respond. Second, the authors might also have provided correlations not only with a measure of a related construct but also with another measure of the same construct, "advance preparation." Because of the trait-oriented definition of advance preparation, the other measure should assess advance preparation through a different method of measurement to ensure that performance is the same regardless of the context in which it is measured. Third, some form of experimental data could also contribute to the validity argument. For example, an experiment might compare response latency data of subjects who had been trained to stop and plan with those who were told not to think before responding. If performance reflected the expected patterns for the two groups, results could be used as justification for making inferences about advance preparation from performance.

In the second example, the justification for Chapelle and Mizuno's (1989) use of performance data to assess resourcing rested solely on content evidence for validity. The authors used the following justification: "The computer provides help only upon request so learners must ask for the help they need when they need it. Learners' requests for help are [therefore] evidence of their use of resourcing..." (p. 28). This provides only one argument for their strategy interpretations.

Other justifications that would have made their argument stronger would include the following: First, they might have consulted learners' verbal reports indicating that they had chosen help in order to obtain assistance with the sentence-construction task, rather than for other purposes (e.g., to see what the help looked like). Second, they could have demonstrated consistency in the use of help over the several weeks the activity was used. Third, they might have supplied correlations between use of help on their learner-controlled CALL activity and another one. To act as correlational validity evidence for the interactionalist definition of the strategy, the covariate must be similar to the original assessment in terms of assessing resourcing in learner-controlled CALL as well. Fourth, they could have used an experimental study comparing subjects who had been trained to use help with those who were not told to use it.

These examples indicate that the nature of the construct-validity evidence is related to the construct definition, and where correlational evidence is concerned the type of construct definition (trait-oriented vs. interactionalist) impacts the choice of covariate in validation research. Further, they also indicate that the legitimacy of inferences made from the data is a matter of degree rather than an all or nothing proposition. Validity justification consists of an argument relevant to understanding the meaning of observed data for making inferences about strategies—an argument that is essential for justifying the use of these assessments in SLA research.

Applications of Computer-Assisted Strategy Research

Although very few research projects have used CASA, the method has shown promise in several ways. First, CASA has been useful in extending the researcher's ability to document behaviors in language-learning contexts particularly when large amounts of precise data must be tabulated to make valid inferences about strategies. For years, classroom researchers attempting to study learner strategies have been frustrated by the amount of relevant performance data that they were able to obtain through observation of behaviors (e.g., Naiman, Frohlich, Stern, & Todesco, 1978). Consequently, computer-assisted assessment offers a useful addition to strategy assessment methods.

Second, CASA allows for gathering strategy data during actual instructional exchanges. Given our increasing appreciation of the effects of learners' perception of the task on their performance (e.g., Bruner, 1990), researchers increasingly value data that are obtained in genuine instructional contexts. Despite the usefulness of data obtained through obtrusive methods such as think-aloud protocols (Ericsson & Simon, 1984), and retrospective self-reports (Cohen & Hosenfeld, 1981), there is a need to complement them using observation programs that document behavior unobtrusively in classroom settings. Thus, researchers begin to integrate SLA and pedagogical research and strengthen the possibility of identifying strategies that might provide useful feedback to learners (Scott & New, 1994).

Problems with Computer-Assisted Strategy Research

I would characterize the current problems of computer-assisted strategy research as consisting of two types: analytic measurement problems and practical problems. Measurement problems refer to the need to investigate validity from the perspectives of both construct inferences and consequences of CASA use.

Methods for justifying construct validity are outlined above. As I illustrated, the validity of the inferences and uses of assessment needs to be justified. Because of the precision, accuracy, and directness of data collection in computer-assisted strategy research, however, researchers might attempt to argue away the need for construct validity justification as Chapelle and Mizuno (1989) did: "The behavior exhibited for [resourcing] provides unequivocal evidence for students' use of [this strategy]; in a sense, the behavior is the strategy" (p. 34). From the perspective of the interactionalist construct definition for the resourcing strategy, the behavior is not equivalent to the strategy itself. A second construct validity problem arises when justifications are inappropriate for the type of inferences and uses made from the assessments. For example, if Chapelle and Mizuno (1989) had, on the basis of their content-validity argument, used their data to make inferences about the extent to which learners were resourcers across contexts and had used that data to decide who needed training in resourcing, these inferences and uses would have been inappropriate.

Consequential validity refers to justifications for the usefulness of an assessment for its intended purposes as well as for its unintended outcomes beyond the immediate assessment event and context. Potential consequential validity problems arise when learners are disturbed, rather than facilitated through the assessment process. For example, data gathered and used in a way that violates learners' rights to privacy would argue against consequential validity, as would computer-assisted assessment of learners who feel uncomfortable using the computer. The potential detrimental effects of computer-assisted language learning in general have not been explored rigorously; however, studies in critical pedagogy (e.g., Bowers, 1988) provide some useful directions that may also pertain to the consequential validity of strategy assessment.[5]

Practical problems in CASA are also worth noting. First, in constructing a computer-assisted assessment, researchers may find it difficult to identify software that simultaneously provides relevant language learning activities for instruction and strategy assessment. Unfortunately, to date little work has attempted to combine efforts in instructional design with those of assessment of either strategies or language. A second practical hurdle for CASA is the challenge of modifying software to get it to gather the appropriate data. Some of the commercial software contains data collection capabilities, but there is no guarantee that a given piece of software will collect the data of interest to the researcher. Many of the CASA projects to date have been conducted by researchers who developed their own software. A third practical problem becomes the management of the large quantities of process data that can be generated by recording the details of learners' interaction. Because disk space is limited, there is always a need to make rational decisions about how to summarize and store the data throughout the assessment process (Goodfellow & Laurillard, 1994). Past research has shown that while these practical problems present significant challenges, they are not insurmountable.

Conclusion

Research on interaction-observation programs has been found useful for assessing some SLA strategies. Methodologically, it is particularly interesting that the researcher can construct the type of learning environments which learners would use routinely for instructional activities but which simultaneously serve as a laboratory for data collection (Doughty, 1992). Moreover, the capability to investigate longitudinally learners' routine "working styles" (Jamieson & Chapelle, 1987) offers an ideal setting for investigating important questions about learners' strategies. For example, how do learners' strategies change as task demands are manipulated or as they accrue experience with a task? How consistent (reliable) is strategy measurement on the same task across different

occasions? Can the accuracy of self-report data be substantiated by observation of computer-documented protocols? The investigation of these and other strategy questions, however, relies on the validity of the measurement used to assess strategies and therefore rests in part on validity issues.

Notes

[1] For years, psychologists (e.g., Sternberg, 1977; Snow, 1981) have used response-latency measures for assessing psychological processes in laboratory experiments. In second language acquisition similar research questions have been raised particularly by those approaching SLA from an information processing perspective (McLaughlin, 1987). Automaticity in language processing is hypothesized to be indicative of language knowledge efficiently stored for expedient retrieval; as an aspect of the target language is better learned, restructuring of knowledge occurs making access more automatic. In experimental settings, the amount of time subjects take to respond to a task has been used as an indicator of how automatic subjects are in the use of the linguistic knowledge necessary for performance on the task, or whether knowledge is implicitly or explicitly stored. For example, Hagen (1994) used a computer program to present items to subjects and to time their responses to make such inferences. Crucial to the interpretation of time-to-respond in this case was the learners' understanding that they were to respond as quickly as they could—a task demand the researcher had to make clear to the subjects.

[2] This is not the only way of viewing a strategy. Another is to define it in terms of the actual observed behaviors (see Cohen, this volume).

[3] Of course, the context of the academic lecture could be defined in greater detail to state a more specific construct definition.

[4] What is needed to better express the interactionalist construct definition is a more complete and systematic way of defining "context." Proposals for approaching this problem have been suggested by researchers in language testing (Bachman, 1990), language instruction (Skehan, 1992), and SLA research (Duff, 1993).

[5] Bowers (1988) points out the need to examine the range of possible activities learners engage in through computer-assisted instruction in order to shed light on what students are missing by spending time at the computer. She also guides us to examine the cultural ideologies—such as the value Western societies place on information and individualization—inherent in our educational uses of technology.

References

Abraham, R. A. (1996). Using task products to assess second language learning processes. *Applied Language Learning, 7*(1).

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bialystok, E. (1981). The role of conscious strategies in second language proficiency. *Modern Language Journal, 65*, 24-35.

Bland, S. K., Noblitt, J. S., Armington, S., & Gay, G. (1990). The naive lexical hypothesis: Evidence from computer-assisted language learning. *Modern Language Journal, 74*, 440-450.

Bowers, C. A. (1988). *The cultural dimensions of educational computing*. New York: Teachers College Press.

Bruner, C. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Chapelle, C. A. (forthcoming). Construct definition and validity inquiry in SLA research. In L. Bachman & A. Cohen (Eds.), *Second language acquisition and language testing interfaces*.

Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research, 10*(2), 157-187.

Chapelle, C., & Mizuno, S. (1989). Students' strategies with learner-controlled CALL. *CALICO Journal, 7*(2), 25-47.

Chun, D. M. (1994). Using computer networking to facilitate the acquisition of inter- active competence. *System, 22*(1), 17-31.

Cohen, A. D. (1996). Verbal reports as a source of insights into second-language learner strategies. *Applied Language Learning, 7*(1).

Cohen, A.D., & Hosenfeld, C. (1981). Some uses of mentalistic data in second language research. *Language Learning, 31*, 285-313.

Doughty, C. (1992). Computer applications in second language acquisition research: Design, description, and discovery. In M. Pennington, and V. Stevens (Eds.), *Computers in applied linguistics: An international perspective* (pp. 127-154). Clevedon, Avon: Multilingual Matters, Ltd.

Doughty, C. (1987). Relating second-language acquisition theory to CALL research and application. In W.F. Smith (Ed.), *Modern media in foreign language education: Theory and implementation* (pp. 133-167). Lincolnwood, IL: National Textbook.

Doughty, C., & Fought, C. (1984). On investigating variable learner response: Toward achieving better CALL courseware design. Report from the language analysis project. Philadelphia: University of Pennsylvania. (cited in Doughty, 1987)

Duff, P. (1993). Tasks and interlanguage performance: An SLA perspective. In G. Crookes & S. M. Gass (Eds.), *Tasks and Language Learning: Integrating Theory and Practice* (pp. 57-95). Philadelphia: Multilingual Matters.

Ericsson, K., & Simon, H. (1984). *Protocol analysis—Verbal reports as data.* Cambridge, MA: The MIT Press.

Goodfellow, R., & Laurillard, D. (1994). Modeling learning processes in lexical CALL. *CALICO Journal, 11*(3), 19-46.

Hagen, L. K. (1994). Constructs and measurement in parameter models of second language acquisition. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.). *Research methodology in second-language acquisition* (pp. 61-87). Hillsdale, NJ: Lawrence Erlbaum Associates.

Higgins, J., & Johns, T. (1984). *Computers and language learning*. New York: Addison-Wesley.

Hsu, J., Chapelle, C. &, Thompson, A. (1993). Exploratory environments: What are they and do students explore? *Journal of Educational Computing Research, 9*(1), 1-15.

Hulstijn, J. (1993). When do foreign language learners look up the meaning of unfamiliar words? The influence of task and learner variables. *Modern Language Journal, 77*(2), 139-147.

Jamieson, J., & Chapelle, C. (1987). Working styles on computers as evidence of second language learning strategies. *Language Learning, 37*, 523-544.

Krashen, S. (1982). *Principles and practices in second language acquisition*. Elmsford, NY: Pergamon Press.

McLaughlin, B. (1987). *Theories of second language acquisition*. London: Edward Arnold.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement (3rd ed.),* (pp. 13-103). NY: Macmillan.

Naiman, N., Frohlich, M., Stern, H., & Todesco, A. (1978). *The good language learner. Research in education series (Report No. 7)*. Toronto: Ontario Institute for Studies in Education.

O'Malley, J., Chamot, A., Stewner-Manzanares, G., Kupper, L., & Russo, R. (1985). Learning strategies used by beginning and intermediate ESL students. *Language Learning, 35*, 21-46.

Skehan, P. (1992). Second language acquisition strategies and task-based learning. *Thames Valley University Working Papers in English Language Teaching, 1*, 178-208.

Scott, V. M. & New, E. (1994). Computer aided analysis of foreign language writing process. *CALICO Journal, 11*(3), 5-18.

Snow, R. E. (1981). Toward a theory of aptitude for learning: Fluid and crystallized abilities and their correlates. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 345-362). New York: Plenum.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, New Jersey: Lawrence Erlbaum.

Wenden, A. (1985). Learner strategies. *TESOL Newsletter, 19*, 1-7.

Acknowledgment

I am grateful for comments from Roberta Abraham and an anonymous reviewer on earlier versions of this article.

Author

CAROL A. CHAPELLE, Professor, TESL/Applied Linguistics, 203 Ross Hall, Iowa State University, Ames, IA 50011. Specializations: computer-assisted language learning, language testing, SLA research.

[Table of Contents]

[Table of Contents]

Applied Language Learning
1996, Vol. 7, No.1, pp. 61-89

Using Task Products to Assess Second Language Learning Processes

Roberta G. Abraham
Roberta J. Vann
Iowa State University

Since the shift of attention in the 1970s to the processes language learners employ in acquiring second language (L2) proficiency, the products of performance (language or other representation of meaning that a learner formulates in response to a task) have often been ignored. However, as Bialystok (1990a) demonstrates with examples from research on communication strategies, when products are examined within appropriate conceptual frameworks that reflect actual processing options, considerable information about underlying intentions and processes is revealed. This article reviews Bialystok's analysis of the communication strategy work and then describes two studies in which products are interpreted within frameworks that illuminate underlying processes of L2 learning. In the first study, reported by Abraham and Vann (1994), the responses from a cloze test are examined within Bialystok and Ryan's (1985) framework of analyzed knowledge and control to reveal L2 learners' metacognitive skills. In the second, carried out by Vann and Schmidt (Schmidt & Vann, 1992; Vann & Schmidt, 1993), a depth of processing framework derived from work by Craik and Lockhart (1972), Marton and Saljo (1984), and others guides the analysis of L2 learners' reading notes to reveal the strategies the learners use as they interact with an academic text. Important to the discussion are arguments for the validity of the interpretations and proposed uses of the cloze responses and reading notes. We conclude that products, appropriately analyzed, can provide important insights into the processes of L2 learners.

In the late 1970s, an important shift occurred in L2 research, away from interest in the *products* of language learning towards the *processes* learners employ in acquiring the knowledge and skills required of a proficient L2 user. However, even then, a cautionary note was sounded by a group of researchers who were committed to dis- covering the processes, also referred to as "strategies and techniques" (Naiman, Frohlich, Stern, & Todesco, 1978, p. 65), that learners employed. In their exemplary report on a well-designed and complex study, they noted that while "the practice of inferring language learning *processes* solely on the basis of language *product* is tenuous" (p. 65, italics in original), in future research a combination of process and product data would be needed to understand how second language learners develop proficiency. By "products," they meant learners' performance on individual exercises and learning tasks, which we take in this paper to include the

language or other representation of meaning (e.g., diagrams, sketches, or gestures) that a learner formulates in response to a task (e.g., completing a cloze exercise, writing a composition, taking notes on a reading passage or a lecture, or describing a series of pictures that tell a story).[1]

However, since 1978, relatively few second language acquisition (SLA) researchers have made use of products to identify and understand the processes and strategies[2] that learners use in performing a particular task. Many researchers appear to perceive process and product as dichotomous (product *versus* process), viewing products as irrelevant in investigating learner strategies. Our purpose here is to dispel this notion by reviewing literature discussed in detail by Bialystok (1990a) in which products have been used to discover the communication strategies of L2 learners, and then by describing two of our own studies in which products were important in illuminating L2 learning processes. An important part of this discussion addresses the validity of the inferences drawn from the product data.

Products as Indicators of Process in L2 Communication

The products most frequently used to shed light on L2 communication strategies have been learners' descriptions of drawings of objects and complex pictures (Tarone, 1977; Varadi, 1980; Bialystok & Frohlich, 1980; Bialystok, 1990a). Other products have resulted from picture reconstruction (Bialystok, 1983), translation (Galvan & Campbell, 1979; Varadi, 1980), sentence completion (Blum-Kulka & Levenston, 1983), conversation (Haastrup & Phillipson, 1983), narration (Dechert, 1983; Raupach, 1983), instruction (Wagner, 1983), word transmission (Paribakht, 1985), and interview (Raupach, 1983). In an extended large-scale study referred to as the Nijmegen project (Bongaerts, Kellerman, & Bentlage, 1987; Kellerman, Bongaerts, & Poulisse, 1987; Poulisse, 1987; Kellerman, Ammerlaan, Bongaerts, & Poulisse, 1990), the products from several types of tasks were examined. These tasks included describing concrete objects and abstract shapes, retelling stories, and interviewing.

In some cases, utterances from these products were categorized in taxonomies such as that developed by Tarone (1977).[3] While these taxonomies have been useful in identifying actual behaviors, Bialystok (1990a) reports a study that demonstrates that they are not without problems of reliability and validity. Reliability concerns, of course, arise in the coding of strategies. Using Tarone's taxonomy, Bialystok found that a single utterance might reflect several strategies, raising the question of whether to code each separately or to evaluate the utterance globally. Further, although Tarone's definitions of categories were explicit and her examples clear, a number of ambiguities became apparent in the coding of individual utterances. As a result of these difficulties, Bialystok estimated that in a reliability check, disagreements between coders occurred for at least 50% of the utterances.

Validity problems were encountered in associating a linguistic utterance provided by the learner with the underlying intentions and processes that produced it. As Bialystok (1990a) puts it, "[i]f the taxonomic descriptions are valid, then the distinctions should correspond to real alternatives or real choices experienced (at some level although not necessarily consciously) by the learner" (p. 56), that is, the choice of utterance should be related systematically to some characteristic of the learner or the environment. However, research attempting to link particular types of utterances to learner or situational factors, for example, learner proficiency (Tarone, 1977; Bialystok, 1983; Paribakht, 1985), elicitation task (Bialystok & Frohlich, 1980), and first language (Tarone, 1977), that might affect these choices has produced few meaningful results. Along these same lines,

Bialystok (1990a) asked L2 learners to use descriptions selected as clear representatives of categories in Tarone's taxonomy in identifying objects from an array. She found little relationship between category of strategy and accurate identification of an object by the listener.[4]

Overall, the taxonomic approach to the study of communication strategies, while yielding valuable insights into the types of surface expressions learners use in coping with communication difficulties, has failed to capture the essence of learner choice and intent, and in this sense, lacks validity. To illustrate how one might probe more deeply into such questions, Bialystok describes two large research programs in which the investigators, rather than simply allowing strategic categories to emerge from the data, selected an appropriate conceptual framework within which to interpret the products so that they revealed the options available to the participants. In the first, Snow, Cancini, Gonzalez, and Shriberg (1989a, 1989b) examined definitions of common objects by bilingual children at the United Nations International School in New York within a framework relating children's use of "decontextualized" metalinguistic use of language to "reading skills and 'literacy' in general" (Snow et al., 1989b, p. 5, quoted in Bialystok, 1990a, p. 105). The definitions were classified as "formal" (containing an equivalence and a superordinate, e.g., "a donkey is an animal") and "informal" (failing to specify such semantic relationships). The ability to produce formal definitions was considered to be an important indicator of metalinguistic development because of the "explicit and 'intentional' use of words outside of the contexts which endow them with meaning" (Bialystok, 1990a, p. 105). While Snow and her colleagues quantified the additional descriptors that contributed to the quality and communicative adequacy of the definitions in order to relate them systematically to the children's progress in achieving literacy, the aspect of their work of relevance here was the identification of factors that did (and did not) predict that a definition would be formal. Academic language proficiency, as measured by reading and language scores on the California Achievement Test (Snow et al., 1989a), predicted the use of a formal definition; for children tested in both L1 and L2, the language used did not, nor did oral proficiency in the L2, though the latter was related to the quality of the definition (Bialystok, 1990a). Because the formal and informal categories of response, "rooted in a conception of what these options mean for language processing" (Bialystok, 1990a, p. 109), make a distinction that could be empirically related to general development ofa learner, they are valid in the sense described above.

The second study described by Bialystok was the Nijmegen project, in which the referential (lexical) communication strategies used by Dutch learners of English in L1 and L2 on several tasks were observed. Again, to avoid the weaknesses of the taxonomy approach, the "tendency to confuse the linguistic realization of the referential strategy with the strategy itself"[5] and the confusion of the strategy with the properties of the referent [6] (Kellerman et al., 1990, p. 165), the researchers used a framework based on a theoretical view of communication and language production. In this view, learners unable to communicate their ideas can either "manipulate the concept so that it becomes expressible through their available linguistic . . . resources, or . . . manipulate the language so as to come as close as possible to expressing their original intention" (Kellerman, source unknown, quoted in Bialystok, 1990a, p. 111). Accordingly, the researchers used only two categories in coding the data: *conceptual,* in which the speaker "analyses the concept by decomposing it into its criterial features" either by listing them or referring to a related concept, and *linguistic*, in which the speaker "manipulates his linguistic knowledge" by strategies of "morphological creativity" or transfer from the first language (Poulisse, 1987, pp. 146-147). Again, while the researchers used products as data, their choice of categories was process-oriented, and the results were subsequently related to learner and situational factors. As in the Snow study, the analysis showed that, in most cases, neither the language used in the response (Dutch or English) nor the learners' proficiency predicted the type of strategy chosen (though the quality and effectiveness of the strategy did vary with proficiency). However, the nature of the task did seem to influence the choice of strategy, with descriptions of single objects eliciting a relatively low proportion of "code" strategies (the

"linguistic" category expanded to include certain "ostensive" strategies), story-retellings evoking proportionally more, and interviews producing the highest number (Bialystok, 1990a, p. 113), thus providing evidence for the validity of the classification scheme. Bialystok (1990a) further validated the two-category system by reanalyzing the taxonomy data in her picture identification task within the framework used in the Nijmegen project. She found a distinct difference in the effectiveness of communication for these two strategy types, with those characterized as "code" inducing a higher percentage of correct listener identifications than those characterized as "conceptual."

The work reported here on the use of products to shed light on the communication strategies of L2 learners has important implications for those interested in L2 learning processes. Products can be useful in inferring these processes, but only when the products are analyzed within a conceptual framework that permits their reliable classification and valid interpretation. That is, the framework must allow for and distinguish among different stages of learner development and/or options available to the learner, and the researcher must be able logically to relate the products to these stages or options. The communication strategy work suggests that there will likely be only a few broad product categories growing out of a conceptual framework, but that each of these may include several manifestations of the underlying process (e.g., in the Nijmegen project, the broad "code" and "linguistic" categories each included more than one typeof utterance). However, the framework allows the researcher to interpret these manifestations as the result of a learner's stage of development or his/her exercising a single underlying option (e.g., deciding to amend her language to make herself understood), and thus avoids the validity problems encountered in interpreting taxonomy results where the categories simply emerged from the data.

In the remainder of this article, we describe two studies of L2 learning processes in which we make use of the insights from the communication strategy work in interpreting task products. In the first, we examine cloze responses to discover the nature of learners' metacognitive skills; in the second, we analyze reading notes to assess the strategies employed as L2 learners interact with academic texts. For each study, we discuss the selection of a framework within which to analyze and interpret the data, the distinctions in the framework that underlie the classification of products, the analysis and coding procedures, and the findings. On the basis of this discussion, we argue for the validity of our interpretations and proposed uses of these products, and suggest that products, appropriately analyzed, can illuminate the processes of L2 learners.

Study One
Cloze Responses as Evidence of Metacognitive Skills of ESL Learners

This study, reported by Abraham and Vann (1994), examined the metacognitive skills displayed on a cloze task administered to a group of Lebanese ESL learners several months after they entered the beginning level of an intensive English program. Although these learners had no prior knowledge of English when they entered the program, their general English proficiency varied considerably at the time the cloze was administered. The major purpose of the study was to compare the metacognitive skills of learners who were successful on the cloze with those of learners who were not. While think-aloud procedures were used to collect data, one goal of the study was to discover how much valid information about metacognition could be gained from analysis of the cloze responses, or *products* of the task. This report focuses on insights gained from the product data.

Background

Although "metacognition" has been defined in many ways and investigated from many points of view in the past 25 years (for example, see Flavell & Wellman, 1977; Brown, Bransford, Ferrara, & Campione, 1983; Yussen, 1985), researchers and educators agree that this "knowing about knowing" (Garner & Alexander, 1989, p. 143) or "thinking about thinking" (Yussen, 1985, p. 253) or "executive control" (Brown et al., 1983, p. 110) plays an important role in learning. Although much of the research on metacognition has dealt with learning and reading in the learner's native language (for example, see reviews in Brown et al., 1983, and Garner, 1987), metacognition has not gone unnoticed in L2 research. Wenden (1987) and Reid (1987) document L2learners' metacognitive knowledge of their own abilities and preferences for learning. Horwitz (1987) describes research using a questionnaire to assess learner beliefs about learning. Jamieson and Chapelle (1987) investigated learners' use of several executive control strategies by means of computer-assisted techniques. O'Malley and Chamot (1990) identified from interview and think-aloud data seven aspects of executive control (planning, directed attention, selective attention, self-management, self-monitoring, problem identification, and self-evaluation) used by foreign language learners. Oxford (1990) advocates the teaching of three types of metacognitive strategies: centering, arranging and planning, and evaluating one's learning. Vann and Abraham (1990) noted the similarity in the cognitive strategies used by successful and unsuccessful learners and pointed to the need for metacognitive explanations for the observed differences in proficiency. Bachman (1991) describes a framework for L2 test development in which one aspect of language ability to be measured is strategic competence, composed of the metacognitive strategies of assessment, goal-setting, and planning, and Block (1992) compared the metacognitive processes of comprehension monitoring in first- and second-language readers.

Most of the L2 research on metacognition has focused on self-understanding and control, that is, what learners know about how they learn and how they oversee and regulate their mental processes as they learn or perform tasks requiring application of what they have learned. Although this work has provided valuable insights into learner differences, its focus neglects an aspect of cognition that seems highly relevant in our quest to understand why some learners are successful and others are not. What is missing in the "control" approach to metacognition is a consideration of the knowledge of language that learners develop and manipulate in learning or performing a task. Metacognitive strategies are not used in a vacuum; learners may be adept at planning, monitoring, and evaluating, but without a firm knowledge base to work from, these metacognitive strategies will not produce success.

In analyzing the data from this study, we used a theoretical framework of metacognition that takes into account both knowledge and control. The framework, proposed by Bialystok and Ryan (1985) to provide insight into broad issues in both first- and second-language development, differs from those used in most discussions of metacognition in that it comprises two "theoretically orthogonal" (p. 208) but pragmatically related dimensions, analyzed knowledge and control,[7] which can be used to characterize both task requirements and learner abilities. Along the first dimension, knowledge of the language ranges from *unanalyzed* information, which is used routinely, with little or no learner awareness of its structure, to *analyzed* information, which the learner is able to manipulate and use creatively. Bialystok and Ryan illustrate this dimension with the example of subject-verb agreement. Some learners simply produce sentences that contain such agreement without understanding the principles involved, thus demonstrating unanalyzed knowledge. Others have access to structural concepts of predicate and arguments, which permit "construction and manipulation" (p. 211) of

sentence parts (e.g., correcting sentences with faulty agreement or stating the rules for agreement), thus demonstrating analyzed knowledge. Cloze tasks such as the one used in this study require a relatively large amount of analyzed knowledge, and learner variation along this dimension can be documented.

The control dimension, as in much of the earlier work on metacognition, "represents an executive function that is responsible for selecting and coordinating the required information within given time constraints" (Bialystok & Ryan, 1985, p. 213). Again, tasks such as the cloze, which require learners to attend to and coordinate multiple characteristics of structure and meaning in order to select appropriate lexical items, make high demands on learners' control abilities and are appropriate in demonstrating individual differences in this respect.

Applying the Bialystok and Ryan framework to the Lebanese study discussed here, we examined not only learners' use of control strategies (including those of predicting, planning, monitoring, and evaluating described by other researchers) but also the nature of their knowledge of English accessed in the process of completing the cloze task (was that knowledge sufficiently analyzed for them to fill each blank with a word that met all the syntactic and semantic constraints of its context?). We expected that learners who performed well on the cloze would show a high degree of control and analyzed knowledge, whereas those who performed less well would show deficiencies in at least one dimension of metacognition, deficiencies that would likely be reflected in their scores on other measures of academic language proficiency such as the TOEFL. The degree to which these expectations were met provides a measure of the validity of using cloze products to reveal L2 learners' metacognitive processes.

Method

Participants

The participants were nine male Lebanese students who, under the sponsorship of the Hariri Foundation, entered the beginning level of the academically oriented Intensive English and Orientation Program (IEOP) at Iowa State University. These scholars had been carefully selected as intelligent, hardworking, and highly motivated to succeed in their respective areas of academic specialization. All were native speakers of Arabic, and none had studied English prior to their arrival in the U.S. Although their scores on the Michigan English Proficiency Test placed all of them into the lowest level (of five) in IEOP, differences in their abilities began to emerge as early as midterm, and by the end of 24 weeks of instruction, their TOEFL scores ranged from 377 to 510.

Procedure

The cloze task was the third of four tasks in which the Lebanese learners were asked to think aloud as they worked. It was administered four to six months after the learners entered the program. Taken from Cohen's (1980) *Testing Language Ability in the Classroom*, the 33-blank cloze passage dealt with cross-cultural differences, presumably a topic with which the Lebanese learners had firsthand experience. As in most clozes, the deleted words differed along a number of dimensions (e.g., whether they were content or function, how much context was needed to restore them, whether they occurred elsewhere in the passage) that affect difficulty (Brown, 1988; Abraham & Chapelle, 1992). A number of items proved to be quite difficult for the learners in this study, thus eliciting the metacognitive skills of interest.

As the learners worked, they were asked to think aloud in the presence of a research assistant, a female Chinese graduate student in TESL.[8] The research assistant's role was limited to probing learners' thoughts when they were not being expressed and answering questions on procedures or vocabulary.

Data Analysis

The analysis of product data consisted of examining the responses to the cloze to identify patterns that could be associated with the metacognitive skills of learners who were successful on the cloze and those who were not. The analysis was organized around three questions, the first focusing primarily on one aspect of control (predicting and planning, in this case selecting appropriate cognitive strategies for the task), and the last two on aspects of both analyzed knowledge and control (selecting and coordinating the required information). The questions and the ways in which products were involved in providing answers are shown below.

1. Did learners appropriately identify the cognitive strategies needed for the task? One important cognitive strategy needed for the task was identifying the main ideas of the passage. Because the cloze proved difficult for these learners, it was assumed that those who demonstrated understanding of the main ideas must have recognized the need to do so. To assess understanding, the researchers individually examined each learner's responses to the key blanks (those carrying important content information) in each of the six main ideas in the passage and rated each idea as "not understood," "partially understood," or "understood." Disagreements were jointly resolved by the researchers.

Another cognitive strategy that could be inferred from products was avoiding guesses that would be misleading in filling in later blanks. Since learners were given no instructions that discouraged them from guessing, unfilled blanks were assumed to reflect the use of this strategy.

2. Did learners demonstrate adequate knowledge of the lexicon and access it appropriately in filling in blanks? The number of correct responses gave one answer to this question, but incorrect responses provided clues concerning learners' analyzed knowledge of the semantic features of lexical items and their ability to access appropriate words for particular contexts. Thus, incorrect answers were classified according to the following criteria: Were they semantically appropriate for the context? Were they in the general semantic field of the passage, though not appropriate in the specific context? Were they totally inappropriate semantically (or too vague to make a judgment)? Were they the opposite of the correct answer? Discrepancies in coding were again jointly resolved.

3. Did learners demonstrate adequate morphological and grammatical knowledge of English and access it appropriately in filling in blanks? As for question 2, both the number of correct responses and the nature of incorrect responses were useful in assessing learners' analyzed knowledge and control. Here, the analyzed knowledge of interest was morphosyntactic, and the researchers individually coded each incorrect answer as syntactically appropriate for the entire sentence, syntactically appropriate if only part of the sentence was considered (operationally, whether the response could fit syntactically with at least one word adjacent to the response), or totally inappropriate syntactically in the sentence or any string of words within it. Again, the researchers jointly resolved discrepancies in coding.

Results

Scores on the cloze ranged from a high of 20 (out of a possible 33) to a low of 6. As expected, the patterning of learners' cloze scores was similar to the patterning of their scores on TOEFL seven months after they entered the program, i.e., Ahmed,[9] Bashir, Khalid, and Daaboul had the highest scores on both tests, whereas Fouad, Ghassan, Hossein, and Jahad had the lowest scores on both (see Table 1).

TABLE 1
Cloze and TOEFL Scores

Subject	Cloze Score	TOEFL Scores after 24 weeks of instruction [a]
Ahmed	20	483 (487)
Bashir	19	477 (510)
Khalid	16	443 (437)
Daaboul	15	510
Essam	13	423
Fouad	10	407
Ghassan	10	377
Hossein	8	400 (420)
Jahad	6	383

[a] First scores shown are for TOEFL administered immediately after instruction was complete. Scores in parentheses are for TOEFL administered three weeks later.

To sharpen the distinction between successful and unsuccessful performers on the cloze, the following analyses contrast the responses of the two highest scorers (with scores of 20 and 19) with those of the four lowest (with scores from 10 to 6). The results of the various product analyses described above are summarized in Table 2.

Of interest here are the qualitative differences between the successful and unsuccessful learners on four key variables that provide evidence of differences in metacognitive abilities. First, the two most successful learners, Ahmed and Bashir, understood more of the six major ideas (4.5) than any of the unsuccessful learners. Second, the two successful learners, while filling nearly two thirds of the blanks with correct responses, omitted a substantial number of the remaining blanks, suggesting that they recognized the dangers of filling in words that they were unsure of. In contrast, the unsuccessful learners left very few blanks unfilled.

TABLE 2
Performance of successful and unsuccessful learners

	Incorrect Responses	
	Semantic	**Syntactic**

Learner	Score	Ideas	Omits	Incorrect	Approp.	Tot. inapprop.	Gen. sem. field	Opp.	Gram.	Tot. ungram.	Too small cntxt.
Successful											
Ahmed	20	4.5	5	8	2	0	6	0	6	0	2
Bashir	19	4.5	12	2	1	0	1	0	2	0	0
Unsuccessful											
Fouad	10	2	0	23	3	9	8	3	8	1	14
Ghassan	10	3.5	0	23	7	7	7	2	7	1	15
Hossein	8	1.5	1	24	3	15	5	1	7	3	14
Jahad	6	1	3	24	3	13	8	0	5	2	17

Ideas = number of main ideas subject appeared to understand (partially understood idea counted as .5)

Omits = number of blanks left empty

Approp. = number of incorrect answers that were semantically appropriate for the context

Tot. inapprop. = number of incorrect answers that were totally inappropriate semantically or too vague to allow a judgment

Gen. sem. field = number of incorrect answers appropriate to general semantic field of passage but not appropriate in specific context

Opp. = number of incorrect answers that were opposites of the correct answers

Gram. = number of incorrect answers that were syntactically appropriate for the context

Tot. ungram. = number of answers totally inappropriate syntactically in any grammatical unit within the sentence

Too small cntxt. = number of answers that were syntactically inappropriate in entire sentence but formed acceptable sequence with at least one adjacent word

Third, the semantic analysis shows that the unsuccessful learners used far more words that were totally inappropriate in the context or were related only to the general topic of the passage than the successful learners. This comparison suggests that the unsuccessful learners had some knowledge of the needed vocabulary, but were unable to select words with precisely the features required by the contexts (because they either lacked these words altogether or were unable to access them on this occasion). Finally, the syntactic analysis shows that the unsuccessful learners used far more words that fit in a local context but not in the sentence as a whole. Like the semantic comparison, this finding suggests that the unsuccessful learners' analyzed knowledge (or control) of English morphology and syntax was inadequate to permit the manipulation required by the cloze or that they simply did not understand overall meaning.

In addition to differentiating the successful from the unsuccessful learners, the product analysis points to at least two differences within the unsuccessful group that would have been useful had follow-up intervention been attempted. The first is in understanding the main ideas of the passage, where Ghassan clearly contrasts with the other three, despite his overall low score. In intervention, it would not have seemed necessary to encourage Ghassan to search for meaning, though the others would likely have benefited from training to develop skills for identifying main ideas. The second is in the number of answers that were totally inappropriate semantically, with Fouad and Ghassan producing only slightly over half as many as Hossein and Jahad. These data suggest an unproductive pattern of wild guessing in Hossein's and Jahad's approaches that would likely be observed in other contexts. In intervention these two learners could have been shown ways to make better guesses by drawing on what they did know.

Discussion

Anchored in Bialystok and Ryan's theoretical model of metacognition, the comparisons reported above support the validity of interpreting products as indicators of learners' metacognitive abilities. The patterns of skills so revealed are related to the learners' degree of success on the cloze and point to important strengths and weaknesses of the unsuccessful learners that could be used in intervention. However, these claims could be strengthened, and would perhaps be modified, by more direct evidence of learners' thought processes.[10] For example, for question 1 above, concerning whether learners identified the cognitive strategies needed for the task, an assumption was made that learners who filled in key blanks correctly recognized the need to discover main ideas, and, indeed, the think-alouds of the successful learners confirmed this assumption. However, those who were less successful in completing the cloze may also have recognized this requirement but have been unable to meet it because of lack of the necessary analyzed knowledge; the think-alouds revealed that this was probably the case for the least successful learner (Jahad). Similarly, the interpretation of unfilled blanks as representing an unwillingness to make potentially misleading guesses could be further validated by learner comments in think-alouds and/or retrospective reports to this effect. Here, the think-alouds of the two successful learners showed that they were intentionally leaving some blanks unfilled because they did not know the answers; additional probing might have revealed why they felt it was desirable to do so.

Revelation of the learners' thought processes would undoubtedly have modified the classification of products in Questions 2 and 3. For example, without knowledge of what the learner was thinking, we classified any word that could *conceivably* fit syntactically with an adjacent word as appropriate in a limited context. The classification of some responses in this category was confirmed in the think-alouds, but in many cases, it was not. Another question that remained unresolved through product analysis was whether an incorrect response resulted from lack of analyzed knowledge or of the control necessary to access the required knowledge. Unfortunately, the think-alouds were of little help in resolving this issue.

Despite these limitations, the analysis of products in this study did illuminate learner processes that would likely remain invisible in other methods of data collection, particularly by showing where learners' guesses were "good" or "bad," and suggesting areas where learners' analyzed knowledge (or control) was deficient. For teachers lacking the time and resources to have learners self-observe or self-report, products from tasks such as the cloze have an obvious advantage in that they are easily available throughout the period of instruction. This study suggests that, appropriately analyzed, these products can provide valuable information pertaining to intervention. The outcomes of such intervention would provide further evidence of the validity of the product

analysis. For researchers seeking detailed information about how learners learn, product analysis can suggest processes that can be confirmed through other data collection methods and sometimes provide insights not available elsewhere.

We turn now to another study in which products provided a different type of information about learner processes.

Study Two
Using Reader Notes to Assess Academic Reading Strategies

Though notetaking is viewed as an essential academic learning strategy, learner notes have rarely been used as a means for *assessing* strategies. We suggest here that learner notes may be a product that can help us view important academic reading strategies that a reader uses in interacting with a text.[11] We draw our support for the validity of using learner notes to assess learner strategies primarily from a study conducted by Schmidt and Vann (1992; Vann & Schmidt, 1993) which investigated the characteristics of notes made by L2 learners while reading academic texts and attempted to link these findings with traits of the learners, including cultural background and level of schooling. In addition to providing evidence that categories of notetaking behavior can be linked with certain influential learner variables, we argue that the naturalness of notetaking enhances its validity in assessment.

Background

Although little direct use has been made of learner notes to assess learner strategies, research has been conducted on the usefulness of notes in enhancing comprehension and learning in first language reading. These studies help reveal the potential of notes as an assessment tool. Notes have been shown to be effective, even when not reviewed, but it is the *quality* of the notes, not merely the act of taking them, that makes a difference in learning (Anderson & Armbruster, 1980). Even when they are not reviewed, notes can enhance retention and comprehension when they focus on text analysis and increase the depth at which material is processed. Anderson and Armbruster (1980) and Bretzing and Kulhavy (1979) suggest that when note-taking seems ineffective it may be because subjects fail to engage in deep and meaningful processing. Further support for the power of certain kinds of notes over others comes from a study done by Hidi and Klaiman (1983) which identified the notetaking strategies of experts and novices, finding that the former tended to "construct certain mental representations" (p. 381), to keep the overall purpose of the text in mind, to perform more critical appraisal of the text, and to paraphrase rather than to copy verbatim. Novices in the study took notes that resembled those made by children (as described by Brown & Smiley, 1978) in relating more closely to the surface structure of the text, both in order of the occurrence of the idea units and in the near verbatim selection of words produced. Depth of processing has also been specifically associated with certain categories of notetaking, especially graphic representations of text, including diagrams or knowledge maps of the content (Jones, Pierce, & Hunter, 1988/89; Tang, 1992/93; and Amer, 1994).

The fact that some kinds of notes may reflect or enable deeper engagement and greater learning connects with a larger body of work relating depth of processing with recall (Craik & Lockhart, 1972; Craik, 1977). This notion that "deep" as opposed to "surface" processes are more likely to involve learning also figures prominently

in the work of Marton and his colleagues (Marton & Saljo, 1984), who originally probed (using questionnaires and interviews) what university students learned from reading an academic text and found that different approaches to the task were associated with qualitative differences in outcomes. While some students fully understood the argument being presented by the author, others could mention only details. Those who set out trying to understand the content of the text and engaged themselves deeply in trying to understand it stood the best chance of getting the main ideas and supporting facts. The original study was subsequently extended to a variety of tasks such as writing assignments and scientific problem solving. Learners taking a "deep approach" to a reading task focus on "what is signified" (i.e., the author's argument) and attempt to organize, find connections, and reformulate and comprehend information, while learners taking a surface approach to the task focus on the "signs" (i.e., the text itself and discrete elements such as vocabulary). "Surface" readers depend on memorizing and rote learning, failing to distinguish major from supporting details or new information from old.

Marton and Saljo (1984) suggest that a surface approach to reading results in poor comprehension, while a deep approach makes understanding the author's message more likely. Applying a similar distinction to writing, Bereiter and Scardamalia (1987) distinguish "knowledge-telling," mere regurgitation of ideas, from knowledge which is transformed in some way. Haas (1990) argues that the extent to which knowledge is transformed distinguishes composing from mere transcription, which involves little or no reformulating, and Spivey (1990) maintains that certain kinds of reformulating (organizing, selecting, and connecting) are key operations in constructing meaning from text.

Taken as a whole, this literature, spanning two decades, paints a convincing picture of the links between certain deep processing operations and the construction of textual meaning on the one hand and surface processing and failure to comprehend on the other. However, because it is not clear to what extent learners are aware of their own use of these strategies, assessment via learner report remains problematic. Bretzing and Kulhavy (1979) suggest that this problem can be solved in part if evidence of deep or surface processing can be discerned in the notes learners take, and this line of reasoning underlies the study by Vann and Schmidt described here.

The researchers, both of whom teach ESL for academic purposes, had noticed that their students spontaneously took notes while reading in class, even when not specifically asked to do so. This observation suggested that notetaking was a by-product of academic reading for most ESL readers, making it an unobtrusive means of gathering data during the reading process when compared with methods that require students to read aloud, answer questions periodically while reading, etc. Further, the researchers observed that students varied remarkably in their notetaking, both in what they chose to note and in their style of conveying it. Prior research suggesting the superiority of some notetaking styles over others encouraged the researchers to examine student notes systematically within the conceptual framework of deep and surface processing to determine the extent to which notes might convey insights into how readers were interacting with texts.

The researchers thus set out to assess certain academic reading strategies of advanced students in an intensive English program. They were interested in whether notetaking strategies, classified according to the deep/surface processing dichotomy, were predictably related to the context of reading, for example, to different texts or tasks, and whether or not cohorts of strategy users were apparent. For example, did students with similar overall English language proficiency, similar amounts of academic training, and/or related first languages show the same tendencies in notetaking?

A basic question in the study, on which we focus here, is the extent to which the results provide evidence for the claim that notes are a valid means of assessing academic reading strategies.

Method

Participants

The participants were 50 students (27 women and 23 men) enrolled in the advanced levels of the IEOP at Iowa State University. Their TOEFL scores ranged from 420 to 617, with a mean of 498, and their Michigan English Language Placement Test (MELPT) scores from 65 to 92, with a mean of 76. Twenty-six of the participants were graduate students, nineteen were undergraduates, and five were nonacademic. The largest ethnic group, comprising nearly a third of the students, was Japanese, and the second largest group was Latin American. No academic participant in the study was more than two months away from full-time study.

Materials

The researchers selected three passages on scientific topics from a Scholastic Aptitude Test preparation manual (CEEB, 1990) based on their likely interest and language appropriateness for advanced ESL students in a pre-academic program. Though equal in length (approximately 375 words), the passages represented distinctively different expository types. Passage 1 described the life of Lloyd Hall, a pioneer in food chemistry; passage 2 contrasted the lives of ancient Pueblos and Navajos; and passage 3 speculated on possible causes of the Tunguska blast, which left a huge crater in Siberia.

Procedure

On the first day of the study participants answered written questions designed to provide information about their approaches to learning. They returned the second day to read one of the three passages under one of two possible conditions: in preparation to take a multiple-choice test (test condition) or to explain the passage to a fellow student (tutor condition). In all cases students were told that they could take as many notes as they wished in any form on the passage itself and/or on a separate sheet provided. They were allowed to use dictionaries of any kind and to ask for explanations from the two teacher-researchers in the room. Participants were given 20 minutes for each task, after which they completed a questionnaire that was designed to assess their strategies and the feelings they experienced while studying the passage, as well as its interest and difficulty for them. At this point, they either took a test or talked about their passage with a fellow student, according to their prior instructions. On the third day, students read a different passage under contrasting conditions. Remaining procedures were identical. The particular assignments to participants of passage, condition, and order in this balanced design enabled the researchers to determine the relationships between these factors and the notes taken.

Coding of data

After collecting notes on two passages from each student, each researcher independently examined a subset of these products and made lists of various notetaking strategies found, for example, underlining, highlighting, paraphrasing, listing, translating, and copying. Together they characterized each strategy as indicative of deep processing or surface processing (Marton & Saljo, 1976), depending on whether it appeared to represent an attempt to extract and/or relate main ideas (deep) or merely to focus on an isolated element of the text (surface). The observed categories of notes and their classification within the deep/surface processing framework are shown in Figure 1, and notes exemplifying deep and surface strategies are shown in Figures 2 and 3.

Surface Processing

1. Copying or transcribing word for word from original document

2. Focusing on vocabulary

-underlining individual words
-highlighting individual words
-writing down/listing discrete words
-defining individual words

Deep Processing

1. Organizing or diagramming information using graphic or pictorial devices including linking or separating lines, parentheses or brackets, space to separate concepts, arrow, bullets or footnotes

2. Paraphrasing from original article restating the text in another form or other words, either telegraphically (omitting function words) or in full sentences

3. Focusing on main ideas

-underlining main ideas
-highlighting main ideas
-writing down /listing main ideas

FIGURE 1. *Surface and deep processing strategies represented in notes*

FIGURE 2. Examples of notetaking strategies: surface processing (GIF 26K)

FIGURE 3. Examples of notetaking strategies: deep processing (GIF 29K)

The researchers then independently analyzed the remaining notes using the protocol. In most cases, strategies were coded as being either present or not in a particular set of notes, with no attempt made to determine the frequency of occurrence. Only for translation and definitions, where units could be commonly agreed upon, were individual units tabulated and frequency examined. In translation, sets of characters written in another language and separated by white space were counted as single units, and for definitions the meaning of each word or phrase written in English was so regarded.

The overall reliability of coding, determined by the percentage of judgments on which the coders originally agreed, was 80%. Discrepancies were discussed until agreement was reached.

Analysis and Results

Correlations with other standard measures

As indicated in the literature on first-language reading reviewed earlier in this paper, deep processing has been associated with text comprehension and superior learning, and surface processing with lack of comprehension and learning. Depth of processing has also been specifically associated with the type of notes taken, with the notes of novices more likely to relate to the surface structure of the text and the notes of experts reflecting greater depth of processing. The question here is whether notes made by L2 learners while reading academic texts reflected the extent of their deep and surface processing, thus arguing for learner notes as a means of gaining insight into certain key learning processes. It was hypothesized that the notes of more proficient readers would reflect more deep processing strategies and fewer surface processing strategies than the notes of less proficient readers, and, conversely, that the notes of less proficient readers would reflect more surface and fewer deep strategies. Strong relationships between student performance on TOEFL and MELPT total and reading/vocabulary part scores and the type of notes produced would support these hypotheses.

To test the hypotheses, Vann and Schmidt used correlational procedures as described by Bohrnstedt and Knoke (1988; see also Borgatta & Bohrnstedt, 1981). The hypotheses were confirmed in part, with strong correlations in the expected directions for translation, copying, and diagramming. Simple correlation coefficients indicated a significant[12] negative correlation between translation and MELPT score (-.3418), MELPT vocabulary part score (-.3806), MELPT reading part score (-.4840), and TOEFL general score (-.4081), and in the expected direction, though not significant for TOEFL reading (-.2472), showing that translation as apparent in notes was associated with lower proficiency scores as expected. Likewise, copying was negatively correlated with vocabulary scores on the MELPT (-.3606) to a significant degree and negatively correlated with MELPT and TOEFL reading scores, though not significantly. Though vocabulary-focus correlations were in the expected direction, none was significant. For deep processing strategies, diagramming was positively correlated with both the structure (.4591) and reading (.5565) subparts of the TOEFL. Though not all categories functioned equally effectively, overall the correlations between certain aspects of notetaking and proficiency suggest that information we can gain from student notes is in accord with information gained about reading skills from standardized tests. The fact that no statistically significant correlation for paraphrasing was found may have resulted from an excessively broad definition that characterized all notes not verbatim from the text as paraphrasing. True paraphrasing may, in fact, entail more sophisticated skills not captured with this definition.

Relationships with other learner variables

As another step in validating the classification of notes developed in this study, Vann and Schmidt examined whether certain categories of strategies might be connected with other key learner variables besides language and reading proficiency. Because prior studies have shown the importance of expertise (presumably acquired through experience) and suggested the role of language and culture in shaping learner strategies, these variables were examined. Learners with more academic experience (graduates) were expected to show more evidence of deep processing strategies than those with less academic experience (undergraduates). Students who came from cultures that emphasize copying and memorizing (Huang & Van Naerssen, 1987) and that might cause students to face greater challenges with vocabulary (in part because cognates with their L1 would be less common) were expected to show more evidence of surface strategies.

Undergraduates versus graduates. The hypothesis that graduate students would use more deep processing strategies than undergraduates was partially supported, as shown by the X2 values in Table 3.[13] Although both graduates and undergraduates used the surface processing strategies of translating and copying verbatim from the texts, half of all undergraduates used these strategies, in comparison to fewer than 15% of the graduates who translated and 19% who copied. Undergraduates and graduates were, however, almost identical in their focus on vocabulary. For deep strategies, the pattern reversed, with more graduates than undergraduates using diagramming, paraphrasing, and main idea focus. Differences were in the direction predicted in all cases except for vocabulary focus, and were statistically significant for translation, copying, and main idea focus.

These differences provide partial empirical support for the validity of the classification of notes described here. They further suggest that certain key academic strategies may be acquired with academic experience, though it is also conceivable that users of deep processing strategies are more likely to become graduate students.

Asians versus Latin Americans. As expected, the notes of Asians revealed more surface processing strategies than did the notes of Latin Americans. As indicated in Table 4, most students did not translate, but those who did were primarily Asian, 37% of whom showed evidence of translation in their notes. Even more dramatically, 48% of all Asians showed evidence of copying directly from the text in their notes, while there was no evidence of copying among Latin Americans. Similarly, 85% of the Asians showed evidence of focusing on vocabulary in their notes, while only 38% of the Latin Americans did. For deep processing strategies, although a greater percentage of Latin Americans used diagramming, paraphrasing, and main idea focus than did Asians, these differences were not significant. What is clear here is that Asians were more likely to reveal surface strategies in their notes: translating, copying, and focusing on vocabulary. Though we have speculated earlier as to possible reasons, when we keep in mind the relatively high proficiency levels of this group, it seems likely that prior educational training, as suggested by Huang & Van Naerssen (1987), and the extent to which there is shared vocabulary between the learner's first language and English both play a critical role.

TABLE 3

Percentage (and number) of undergraduates and graduates taking notes reflecting surface and deep strategies (n=43)

Type of notes	Undergraduates (n=16)	Graduates (n=27)	X^2 (df)
Surface			
Translating	50% (8)	15% (4)	6.18 (1)
Copying	50% (8)	19% (5)	4.84 (1)
Vocabulary focus	69% (11)	70% (19)	.012 (1)
Deep			
Diagramming	63% (10)	81% (22)	1.90 (1)
Paraphrasing	81% (13)	93% (25)	1.17 (1)
Main idea focus	63% (10)	93% (25)	5.90 (1)

TABLE 4

Percentage (and number) of Asian and Latin American students taking notes reflecting surface and deep strategies (n=40)

Type of notes	Asian (n=27)	Latin American (n=13)	X^2 (df)
Surface			
Translating	37% (10)	8% (1)	3.79 (1)
Copying	48% (13)	0	9.27 (1)
Vocabulary focus	85% (23)	38% (5)	9.13 (1)
Deep			
Diagramming	70% (19)	92% (12)	2.42 (1)
Paraphrasing	85% (23)	92% (12)	0.41 (1)
Main idea focus	74% (20)	92% (12)	1.82 (1)

Contextual Variables

There was no statistical support for the prediction that the passage, the student's assessment of the difficulty of the passage, the condition (test versus tutor), or the order in which the student read the passage influenced the extent of deep or surface processing. These findings would suggest that deep and surface processing are fairly strong traits not greatly influenced by at least some contextual modifications. They do not suggest that context has no effect, only that the kinds of contextual manipulations in this study did not result in perceivable differences in notetaking.

Discussion

This segment of the paper has focused on validity issues surrounding the development of a relatively untested tool for assessing academic reading strategies. The process of validation has so far consisted of attempting to anchor notetaking in relevant prior theory and connect it empirically with factors suggested by that theory. However, there are other features of reader notes that contribute to the validity of their use as a means of assessing learner strategies.

As a by-product of academic reading, notes are readily available to teachers and researchers for evaluation, thus enabling frequent in-class assessment. Comparing notes across individual assignments can provide information about the conditions under which a student chooses one strategy over another, and lead to a better understanding of the relationships between context and strategic decisions. However, the ease of collecting readers' notes is not merely convenient; it allows data to be collected without imposing constraints on the individual's reading such as insisting on oral reading or stopping the reader intermittently and asking for

commentary unrelated to the passage. Reader notes offer two other advantages as well. Because notes, like other learner-initiated products, avoid self-report or retrospective data, they enable the analyst to tap strategies of which learners may not be conscious or which they may be unable or unwilling to report. The observer sees concrete evidence, not testimony. Finally, notes, somewhat like individual fingerprints, can be classified into types, but remain unique, reflecting the individual response of a reader to a text in a way no standardized test can.

Although notes are limited to assessing academic reading strategies, it is difficult to overestimate the importance of academic reading for many of the world's users of English. Assuming that academic reading requires different processes from ordinary everyday reading (Carver, 1992), it would seem that specialized assessment tools are appropriate. As with other tools for strategy assessment, notetaking does not provide a direct view of cognitive processes; the researcher must infer these from the data the learner provides. Because the kinds of inferences we can make from each kind of assessment are different and limited, a variety of methods—think-alouds, interviews, product analysis—should be employed to obtain a comprehensive picture of learner strategies.

Given the limited population, texts, and contexts in the study, these findings must be interpreted tentatively. Further studies should be conducted to pursue key questions about the relationship between reader notes and reader behavior in various contexts. One obvious question currently being examined by Vann and Schmidt is the extent to which different kinds of notetaking behavior might correlate with differences in recalling texts. Other studies might pursue further how the learner's assessment of a task influences strategic behavior and presumably task performance. MacIntyre's (1994) framework for examining strategies in a psychological and social context may prove especially fruitful in helping future researchers formulate the complex relationships involved between individual traits and contextual variables.

Conclusion

The two studies discussed above suggest that common classroom products, responses to cloze tests and notes made by readers, can be generated under various experimental conditions and offer a practical and natural means for gaining access to certain types of learner processes. Following the approach suggested by Bialystok (1990a), the researchers analyzed the products within conceptual frameworks that permitted the illumination of underlying processes. The results were validated by a variety of means that demonstrated predicted relationships with learner and contextual variables. While products cannot reveal the entire breadth of learner processes, they can provide information not obtainable through other types of data collection and validate findings from other sources. Thus, the work reported here bears out the wisdom of Naiman et al.'s (1978) call for the use of both process *and* product data in understanding second language learning. We look forward to future studies demonstrating the validity of other uses of products to assess learner processes.

Notes

[1] This view of product differs from that in which product is "typically defined by global proficiency scores" (Larsen-Freeman & Long, 1991, p. 283). It should be noted that such global measures usually require "selected" responses (in multiple-choice items), whereas the responses of interest here are "constructed" (Bachman, 1990).

[2] A number of SLA researchers, for example, Blum and Levenston (1978), Bialystok (1978), Frauenfelder and Porquier (1979), Faerch and Kasper (1983), and Seliger (1984)—all discussed in Bialystok (1990a)—and Rubin (1981) have distinguished between processes and strategies. Processes are typically described as "the mental steps taken to carry out a cognitive activity" (Bialystok, 1990a, p. 15), whereas strategies have been variously characterized as optional, conscious, and/or problem-oriented occurrences that come into play on a particular occasion. Bialystok (1990a) suggests that the criteria for differentiating processes from strategies in language production are "fragile" (p. 24) and that the two types of events may not be basically different.

[3] Tarone's 1977 taxonomy grew out of earlier attempts by Varadi (1973) and Tarone, Cohen, and Dumas (1976) to characterize communication strategies.

[4] There were no significant differences in the effectiveness of three of the four strategies tested (approximation, circumlocution, and language switch). The fourth strategy, word coinage, used sparingly by the learners who had produced the descriptions used in this study, was statistically more effective than the other three.

[5] To illustrate this confusion, Kellerman et al. (1990) refer to Varadi's (1973) example of a balloon being called an "airball" or a "ball with air in it," (Kellerman et al., 1990, p. 165) by a learner who does not know the correct word. Using typical taxonomic categories, the first would be classified as a word-coinage and the second, as paraphrase. However, the difference between the two attempts is merely one of superficial linguistic form.

[6] Here Kellerman et al.'s example is of a learner who, lacking the words *moon* and *knife*, substitutes "it's in the sky" and "you use it for cutting," respectively. Again, using carefully defined taxonomic categories, the first response would be classified as a description of a location, and the second as a description of function. However, this categorization "fails to capture the obvious point that the contents of the descriptions are only as different from each other as *the referents themselves*" (Kellerman et al., 1990, p. 166, italics in original).

[7] Subsequent revisions of the framework appear in Bialystok (1990a, 1990b, 1994). In these more recent versions, Bialystok regards the analyzed knowledge dimension as a *process*, with emphasis on the *analysis* of knowledge, in which mental representations, initially "loosely organized around meanings (knowledge of the world)," become "rearranged into explicit representations that are organized around formal structures" (Bialystok, 1994, p. 159). Thus, a learner's representations of knowledge at any given time are indicators of the degree of analysis that has taken place.

[8] The reports obtained in this study would fall into Cohen's (this issue) category of *self-observation*. Although participants were not explicitly told to think aloud in English, the context of the data collection strongly encouraged English, and it was exclusively used. If we repeated this study today, we would give participants a clear choice between using their native language or English.

[9] All learner names used in this report are pseudonyms.

[10] Poulisse, Bongaerts, and Kellerman (1987) discuss the usefulness of this kind of validation in the context of communication strategies.

[11] "Academic reading" is defined here as any kind of reading for academic purposes.

[12] Any reference to statistical significance implies alpha = .05.

[13] Of the initial sample of 50 participants, two students failed to complete the second reading and five students had "special student" status. Thus, strategies for a total of 43 undergraduates and graduates were analyzed here.

References

Abraham, R. G., & Chapelle, C. A. (1992). The meaning of cloze test scores: An item difficulty perspective. *Modern Language Journal, 76*(4), 468-479.

Abraham, R. G., & Vann, R. J. (1994, March). *Metacognition in the cloze performance of nine Lebanese ESL learners*. Paper presented at the Annual Conference of the American Association for Applied Linguistics, Baltimore.

Amer, A. A. (1994).The effect of knowledge-map and underlining training on the reading comprehension of scientific texts. *English for Specific Purposes, 13*(1), 35-45.

Anderson, T. H., & Armbruster, B. B. (1980). *Studying* (Technical Report No. 155). University of Illinois: Center for the Study of Reading.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly, 25*(4), 671-704.

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

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning, 28*(1), 69-83.

Bialystok, E. (1983). Some factors in the selection and implementation of communication strategies. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 100-118). London: Longman.

Bialystok, E. (1990a). *Communication strategies: A psychological analysis of second-language use.* Oxford: Basil Blackwell.

Bialystok, E. (1990b). The dangers of dichotomy: A reply to Hulstijn. *Applied Linguistics, 11*(1), 46-51.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16* (2), 157-168.

Bialystok, E,. & Frohlich, M. (1980). Oral communication strategies for lexical difficulties. *Interlanguage Studies Bulletin, 5,* 3-10. Cited in Bialystok (1990a).

Bialystok, E., & Ryan, E. B. (1985). A metacognitive framework for the development of first and second language skills. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition and human performance*, Vol. 1: *Theoretical perspectives* (pp. 207-252). Orlando, FL: Academic Press.

Block, E. L. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, *26*(2), 319 -343.

Blum, S., & Levenston, E. A. (1978). Universals of lexical simplification. *Language Learning*, *28*(2), 399-415.

Blum-Kulka, S., & Levenston, E. A. (1983). Universals of lexical simplification. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 119-139). London: Longman.

Bohrnstedt, G. W., & Knoke, D. (1988). *Statistics for social data analysis* (2nd ed.). Itasca, IL: Peacock.

Bongaerts, T., Kellerman, E., & Bentlage, A. (1987). Perspective and proficiency in L2 referential communication. *Studies in Second Language Acquisition, 9*(2), 171-199.

Borgatta, E. F., & Bohrnstedt, G. W. (1981). Level of measurement: Once over again. *Sociological Methods and Research*, *9*(2), 147-160.

Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *Con- temporary Educational Psychology*, *4*(2), 145-153.

Brown, A. L., Bransford, J. D., Ferrera, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Hand- book of child psychology* (4th ed.). Vol. 3: *Cognitive development* (pp. 77-166). New York: Wiley.

Brown, A. L., & Smiley, S. S. (1978). The development of strategies for studying texts. *Child Development*, *49*(4), 1076-1088.

Brown, J. D. (1988). What makes a cloze item difficult? *University of Hawai'i Working Papers in ESL*, *7*, 17-39.

Carver, R. (1992). Effect of prediction activities, prior knowledge, and text type upon amount comprehended: Using rauding theory to critique schema theory research. *Reading Research Quarterly*, *27*(2), 165-174.

CEEB (1990). *10 SATS* (4th ed.). Princeton: College Entrance Examination Board.

Cohen, A. D. (1980). *Testing language ability in the classroom*. Rowley, MA: Newbury House.

Craik, F. I. M. (1977). Depth of processing in recall and recognition. In S. Dornic (Ed.), *Attention and performance* (pp. 679-697). Hillsdale, NJ: Lawrence Erlbaum.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 175-195). London: Longman.

Faerch, C., & Kasper, G. (1983). Plans and strategies in foreign language communi- cation. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 20-60). London: Longman.

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail, Jr. & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3-33). Hillsdale, NJ: Erlbaum.

Frauenfelder U., & Porquier, R. (1979). Les voies d'apprentissage en langue etrangere. *Working Papers on Bilingualism*, *17*, 37-64. Cited in Bialystok (1990a).

Galvan, J., & Campbell, R. (1979). An examination of the communication strategies of two children in the Culver City Spanish immersion program. In R. Andersen (Ed.), *The acquisition and use of Spanish and English as first and second languages*. Washington, DC: TESOL. Cited in Bialystok (1990a).

Garner, R. (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.

Garner, R., & Alexander, P. A. (1989). Metacognition: Answered and unanswered

questions. *Educational Psychologist, 24*(2), 143-158.

Haas, C. (1990). Composing in technological contexts: A study of note-making. *Written Communication, 7*(4), 512-547.

Haastrup, K., & Phillipson, R. (1983). Achievement strategies in learner/native speaker interaction. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 140-158). London: Longman.

Hidi, S., & Klaiman, R. (1983). Notetaking by experts and novices: An attempt to identify teachable strategies. *Curriculum Inquiry, 13*(4), 377-395.

Horwitz, E. K. (1987). Surveying student beliefs about language learning. In A. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 119-129). Englewood Cliffs, NJ: Prentice Hall.

Huang, X-H., & Van Naerssen, M. (1987). Learning strategies for oral communication. *Applied Linguistics, 8*(3), 287-307.

Jamieson, J., & Chapelle, C. (1987). Working styles on computers as evidence of sec- ond-language learning strategies. *Language Learning, 37*(4), 523-544.

Jones, B. F., Pierce, J., & Hunter, B. (1988/89). Teaching students to construct graphic representations. *Educational Leadership, 46*(4), 20-25.

Kellerman, E., Ammerlaan, T., Bongaerts, T., & Poulisse, N. (1990). System and hierarchy in L2 compensatory strategies. In R. C. Scarcella, E. S. Andersen, & S. D. Krashen, *Developing communicative competence in a second language* (pp. 163-178). New York: Newbury House.

Kellerman, E., Bongaerts, T., & Poulisse, N. (1987). Strategy and system in L2 referential communication. In R. Ellis (Ed.), *The social context of second language acquisition.* Englewood Cliffs, NJ: Prentice-Hall. Cited in Bialystok (1990a).

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research.* New York: Longman.

MacIntyre, P. D. (1994). Toward a social psychological model of strategy use. *Foreign Language Annals, 27*(2), 185-195.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology 46*(1), 4-11.

Marton, F., & Saljo, R. (1984). Approaches to learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning* (pp. 36-55). Edinburgh: Scottish Academic Press.

Naiman, N., Frohlich, M., Stern, H. H., & Todesco, A. (1978). *The good language learner.* Toronto: Ontario Institute for Studies in Education.

O'Malley, J . M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition.* Cambridge: Cambridge University Press.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know.* New York: Newbury House.

Paribakht, T. (1985). Strategic competence and language proficiency. *Applied Ling- uistics, 6*(2), 132-146.

Poulisse, N. (1987). Problems and solutions in the classification of compensatory strategies. *Second Language Acquisition, 3*(2), 141-153.

Poulisse, N., Bongaerts, T., & Kellerman, E. (1987). The use of retrospective verbal reports in the analysis of compensatory strategies. In C. Faerch & G. Kasper (Eds.), *Introspection in second language research* (pp. 213-229). Clevedon: Multilingual Matters.

Raupach, M. (1983). Analysis and evaluation of communication strategies. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 199-209). London: Longman.

Reid, J. M. (1987). The learning style preferences of ESL students. *TESOL Quarterly, 21*(1), 87-111.

Rubin, J. (1981). Study of cognitive processes in second language learning. *Applied Linguistics, 2*(2), 117-131.

Schmidt, H. H., & Vann, R. (1992, March). *Classroom format and student reading strategies: A case study*. Paper presented at the 26th Annual Convention & Exposition of Teachers of English to Speakers of Other Languages, Vancouver.

Seliger, H. W. (1984). Processing universals in second language acquisition. In F. R. Eckman, L. H. Bell, & D. Nelson (Eds.), *Universals of second language acquisition* (pp. 36-47). Rowley, MA: Newbury House.

Snow, C., Cancini, H., Gonzalez, P., & Shriberg, E. (1989a). Giving formal definitions: An oral language correlate of school literacy. In D. Bloome (Ed.), *Classrooms and literacy* (pp. 233-249). Norwood, NJ: Ablex.

Snow, C. E., Cancini, H., Gonzalez, P., & Shriberg, E. (1989b). Second language learners' formal definitions: An oral language correlate of school literacy. In D. Bloome (Ed.), *Literacy in functional settings.* Norwood, NJ: Ablex. Cited in Bialystok (1990a).

Spivey, N. N. (1990). Transforming texts: Constructive processes in reading and wri-ting. *Written Communication, 7*(2), 256-287.

Tang, G. M. (1992/1993). Teaching content knowledge and ESOL in multicultural classrooms. *TESOL Journal, 2*(2), 8-12.

Tarone, E. (1977). Conscious communication strategies in interlanguage. In H. D. Brown, C. A. Yorio, and R. C. Crymes (Eds.), *On TESOL '77* (pp. 194-203). Washington, D.C.: TESOL.

Tarone, E., Cohen, A., & Dumas, G. (1976). A closer look at some interlanguage terminology. *Working Papers in Bilingualism, 9*, 76-90. Cited in Tarone (1977).

Vann, R. J., & Abraham, R. G. (1990). Strategies of unsuccessful language learners. *TESOL Quarterly, 24*(2), 177-198.

Vann, R. J., & Schmidt, H. H. (1993, April). *Using student notes to assess reading strategies*. Paper presented at the 27th Annual Convention and Exposition of Teachers of English to Speakers of Other Languages, Atlanta.

Varadi, T. (1973). Strategies of target language learner communication: Message adjustment. Paper presented at the VIth Conference of the Romanian-English Ling- uistics Project, Timisoara. Cited in Kellerman et al. (1990).

Varadi, T. (1980). Strategies of target language learner communication: Message-adjustment. *International Review of Applied Linguistics, 18*(1), 59-72.

Wagner, J. (1983). Dann du tagen eineeeee—weisse Platte—an analysis of interlanguage communication in instructions. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 159-174). London: Longman.

Wenden, A. (1987). Metacognition: An expanded view on the cognitive abilities of L2 learners. *Language Learning, 37*(4), 573-597.

Yussen, S. R. (1985). The role of metacognition in contemporary theories of cognitive development. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance,* Vol. 1: *Theoretical perspectives* (pp. 253-283). Orlando, FL: Academic Press.

Acknowledgment

We wish to thank the anonymous reviewers and Lidia Woytak for helpful suggestions on an earlier version of this article.

Authors

ROBERTA G. ABRAHAM, Professor Emerita, Department of English, Iowa State University, 203 Ross Hall, Ames, IA 50011-1201. Specialization: individual differences in second language learning.

ROBERTA J. VANN, Professor, Department of English, Iowa State University, 203 Ross Hall, Ames, IA 50011-1201. Specialization: methods of teaching English as a second language.

[Table of Contents]

General Information

Calendar of Events*

1996

17-18 October, *Minnesota Council on the Teaching of Foreign Languages,* location to be announced. Information Millie Mellgren, 3055 Rosewood Lane, Plymouth, MN 55441.

18-20 October, *New York State Association of Foreign Language Teachers,* Kiamesha Lake. Information Helene Combopiano, 71 Grand Boulevard, Binghamton, NY 13905.

18-20 October, *Texas Foreign Language Association,* Houston. Information TFLA, 1320 Modiste Drive, Houston, TX 7705; (713)468-4959.

23-25 October, *Foreign Association of North Dakota,* Bismarck. Information Herbert Boswau, P.O. Box 8198, Grand Forks, ND 58202-8198.

24-26 October, *Foreign Language Association of North Carolina,* Greensboro. Information Wayne Figart, P.O. Box 739, Wilmington, NC 28402-0739; (910)763-5733.

24-27 October, *Second Language Acquisition and Teaching,* Tucson. Information SLAT, Modern Languages 347, University of Arizona, Tucson, AZ 85721; Email [SLRF96@ccit.arizona.edu].

1-2 November, *National Association of Self-Instructional Language Programs,* Washington. Information NASILP, Critical Languages, 022-38, Temple University, Philadelphia, PA 19122.

22-24 November, *American Council on the Teaching of Foreign Languages with American Association of Teachers of German,* Philadelphia. Information ACTFL, 6 Executive Plaza, Yonkers, NY 10701-6810; Fax (914)963-1275.

27-30 December, *Modern Language Association of America,* Washington, D.C. Information MLA, 10 Astor Place, New York, NY 10003-6981; Fax (212)477-9863.

1997

9-11 January, *Chronos Conference,* Brussels. Information Svetlana Vogeleer, Institut Libre Marie Haps, rue d'Arlon 3, 1040 Brussels, Belgium;

Fax 32-2-511-98-37.

6-8 March, *Southern Conference on Language Teaching with South Carolina Foreign Language,* Myrtle Beach; Information Lee Bradley, Valdosta State University, Valdosta, GA 31698; (912)333-7358, Fax (912)333-7398, Email [lbradley@grits.valdosta.peachnet.edu].

8-11 March, *American Association of Applied Linguistics,* Orlando. Information AAAL, 7630 West 145th Street, Suite 202, Apple Valley, MN 55124; (612)953-0805, Fax (612)891-1800, Email [howe@mr.net].

11-15 March, *Teachers of English to Speakers of Other Languages,* Orlando. Information TESOL, 1600 Cameron St., Suite 300, Alexandria, VA 22314-2751; (703)836-0774, Fax (703)836-7864, Email [tesol@tesol.edu].

3-5 April, *North American Association of Christian Foreign Language and Literature Faculty,* Grand Rapids. Information Barbara Carvill, Calvin College, Grand Rapids, MI 49546; (616)957-6365, Email [carv@calvin.edu].

3-6 April, *Northeast Conference on the Teaching of Foreign Languages,* New York. Information Northeast Conference, St. Michael's College, 29 Ethan Allen Avenue, Colchester, VT 05446.

3-6 April, *Northeast Conference on the Teaching of Foreign Languages,* New York. Information Northeast Conference, Dickinson College, PO Box 1773, Carlisle, PA 17013-2896; (717)245-1977, Fax (717)245-1976, Email [neconf@dickinson.edu].

10-12 April, *Pacific Northwest Council for Languages,* Eugene. Information PNCFL, Foreign Languages and Literatures, Oregon State University, 210 Kidder Hall, Corvallis, OR 97331-4603; Fax (541)737-3563, Email [verzascr@cla.orst.edu].

10-13 April, *Central States Conference on the Teaching of Foreign Languages,* Columbus. Information CSCTFL, Madison Area Technical College, 3550 Anderson Avenue, Madison, WI 53704; (608)246-6573, Fax (608)246-6880.

10-13 April, *Central States Conference on the Teaching of Foreign Languages,* Columbus. Information CSCTFL, Rosalie Cheatham, University of Arkansas - Little Rock, 2801 S. University Avenue, Little Rock, AR 72204; (501)569-8159, Fax (501)569-3220, Email [rmcheatham@ualr.edu].

16-18 April, *Russia's Golden Age,* Columbus. Information George Kalbouss, Slavic & E. European Languages & Literatures, 232 Cunz Hall, 1841 Millikin Road, Ohio State University, Columbus, OH 43210; (614)292-6733, Fax (614)292-3107.

18-20 April, *Midwest Slavic Conference,* Columbus. Information George Kalbouss, Slavic & E. European Languages & Literatures, 232 Cunz Hall, 1841 Millikin Road, Ohio State University, Columbus, OH 43210; (614)292-6733, Fax (614)292-3107.

25-26 April, *Classical Association of the Atlantic States,* Wilmington. Information R. Leon Fitts, Department of Classical Studies, PO Box 1773, Dickinson College, Carlisle, PA 17013-2896.

8-10 May, *Non-Slavic Languages,* Chicago. Information NSL-10, Department of Slavic Languages & Literatures, University of Chicago, 1130 E. 59th St., Chicago, IL 60637; (312)702-8033, Fax (312)702-9861, Email [hia5@midway.unchicago.edu].

4-7 July, *Society for the History of Authorship, Reading and Publishing,* Cambridge. Information SHARP, 51 Sherlock Close, Cambridge CB3 0HP, United Kingdom.

5-7 August, *Problems of Teaching Modern Hebrew,* Jerusalem. Information Ben-Zion Fischler, Council on the Teaching of Hebrew, P.O.B. 7413, Jerusalem 91073, Israel.

12-16 August, *Foreign Language Education and Technology,* Victoria. Information (604)721-8294, Fax (604)721-8778, Email [fleatIII@call.uvic.ca], WWW [http://ikra.call.uvic.ca/langcen/fleat/fleat1.html].

4 September, *Association of Literary Semantics,* Freiburg. Information Monika Fluderik, English Department, University of Freiburg, D-79085, Germany.

7-11 October, *Foreign Language Association of North Carolina,* Winston-Salem. Information Wayne Figart, P.O. Box 739, Wilmington, NC 28402-0739; (910)763-5733.

7-11 October, *Foreign Language Association of North Carolina,* location to be announced. Information Mary Lynn Redmond, 6 Sun Oak Court, Greensboro, NC 27410; Fax (910)759-4591, Email [redmond@wfu.edu].

21-23 November, *American Council on the Teaching of Foreign Languages,* Nashville. Information ACTFL, 6 Executive Plaza, Yonkers, NY 10701-6801; (914)963-8830, Fax (914)963-1275.

27-30 December, *Modern Language Association of America,* Toronto. Information MLA, 10 Astor Place, New York, NY 10003-6981; Fax (212)477-9863.

27-30 December, *American Association of Teachers of Slavic & E. European Languages,* Toronto. Information AATSEEL, 1981 N. Evelyn Ave., Tucson, AZ 85715; Fax (520)885-2663, Email [76703.2063@compuserve.com].

1998

26-28 February, *Southern Conference on Language Teaching with Foreign Language Association of Georgia*, Savannah. Information Lee Bradley, Valdosta State University, Valdosta, GA 31698; (912)333-7358, Fax (912)333-7389, Email [lbradley@grits.valdosta.peachnet.edu].

17-21 March, *Teachers of English to Speakers of Other Languages*, Seattle. Information TESOL, 1600 Cameron St., Suite 300, Alexandria, VA 22314-2705; (703)518-2521, Fax (703)836-7864, Email [jennifer@tesol.edu].

26-29 March, *Central States Conference on the Teaching of Foreign Languages*, Milwaukee. Information CSCTFL, Rosalie Cheatham, University of Arkansas - Little Rock, 2801 S. University Avenue, Little Rock, AR 72204; (501)569-8159, Fax (501)569-3220, Email [rmcheatham@ualr.edu].

16-19 April, *Northeast Conference on the Teaching of Foreign Languages*, New York. Information Northeast Conference, Dickinson College, PO Box 1773, Carlisle, PA 17013-2896; (717)245-1977, Fax (717)245-1976, Email [neconf@dickinson.edu].

24-26 April, *Pacific Northwest Council for Languages*, Boise. Information PNCFL, Foreign Languages and Literatures, Oregon State University, 210 Kidder Hall, Corvallis, OR 97331-4603; Fax (541)737-3563, Email [verzascr@cla.orst.edu].

September dates to be announced, *International Congress of Slavists*, Cracow. Information Robert Rothstein, Department of Slavic Languages, Herter Hall, University of Massachusetts, Amherst, MA 01003-3940.

27-31 October, *Foreign Language Association of North Carolina*, Winston-Salem. Information Wayne Figart, P.O. Box 739, Wilmington, NC 28402-0739; (910)763-5733.

[Table of Contents]

Information for Contributors

Statement of Purpose

The purpose of *Applied Language Learning* (*ALL*) is to increase and promote professional communication within the Defense Foreign Language Program and academic communities. *ALL* publishes research and review articles, research reports as well as reviews on adult language learning for functional purposes. The editor encourages the submission of research and review manuscripts from such disciplines as: (1) instructional methods and techniques; (2) curriculum and materials development; (3) testing and evaluation; (4) implications and applications of research from related fields such as linguistics, education, communication, psychology, and social sciences; (5) assessment of needs within the profession.

Research Article

Divide your manuscript into the following sections:

- Abstract
- Introduction
- Method
- Results
- Discussion
- Conclusion
- Appendices
- References
- Author

Introduction

In a few paragraphs state the purpose of the study and relate it to the hypothesis and the experimental design. Point out the theoretical implications of the study and relate them to previous work in the area.

Next, under the subsection *Literature Review,* discuss work that had a direct impact on your study. Cite only research pertinent to a specific issue and avoid references with only tangential or general significance. Emphasize pertinent findings and relevant methodological issues. Provide the logical continuity between previous and present work. Whenever appropriate, treat controversial issues fairly. You may state that certain studies support one conclusion and others challenge or contradict it.

Method

Describe how you conducted the study. Give a brief synopsis of the method. Next develop the subsections pertaining to the *participants,* the *materials,* and the *procedure*.

Participants. Identify the number and type of participants. Specify how they were selected and how many participated in each experiment.

Identify the number of experiment dropouts and the reasons they did not continue.

Provide major demographic characteristic such as age, sex, geographic location, and institutional affiliation. Indicate that the treatment of participants was in accordance with the ethical standard of the APA (Principle 9).

Apparatus. Describe briefly the materials used and their function in the experiment.

Procedure. Describe each step in the conduct of the research. Include the instructions to the participants, the formation of the groups, and the specific experimental manipulations.

Results

First state the results. Next describe them in sufficient detail to justify the findings. Mention all relevant results, including those that run counter to the hypothesis.

Tables and figures. Prepare tables to present exact values. Use tables sparingly. Sometimes you can present data more efficiently in a few sentences than in a table. Avoid developing tables for information already presented in other places. Prepare figures to illustrate key interactions, major interdependencies, and general comparisons. Indicate to the reader what to look for in tables and figures.

Discussion

Express your support or nonsupport for the original hypothesis. Next examine, interpret, and qualify the results and draw inferences from them. Do not repeat old statements: Create new statements that further contribute to your position and to readers understanding of it.

Conclusion

Succinctly describe the contribution of the study to the field. State how it has helped to resolve the original problem. Identify conclusions and theoretical implications that can be drawn from your study.

Review Article

It should describe, discuss, and evaluate several publications that fall into a topical category in foreign language education. The relative significance of the publications in the context of teaching realms should be pointed out. A review article should be 15 to 20 double-spaced pages.

Research Report

It should present and discuss data obtained in a research project in the area of foreign language education. A research report should be 5 to 10 double-spaced pages.

Review

Reviews of textbooks, scholarly works on foreign language education, dictionaries, tests, computer software, video tapes, and other non-print materials will be considered for publication. Both positive and negative aspects of the work(s) being considered should be pointed out. The review should give a clear but brief statement of the work's content and a critical assessment of its contribution to the profession. Quotations should be kept short. Reviews that are merely descriptive will not be accepted for publication. The length of the manuscript should be three to five double-spaced pages.

Submission of Manuscripts

All editorial correspondence, including manuscripts for publication should be sent to:

Applied Language Learning
ATFL-*ALL*
ATTN: Editor (Dr. L. Woytak)
Defense Language Institute
Foreign Language Center
Presidio of Monterey, CA 93944-5006

Manuscripts are accepted for consideration with the understanding that they are original material and are not being considered for publication elsewhere.

Specifications for Manuscripts

All material submitted for publication should conform to the style of the *Publication Manual of the American Psychological Association* (4th Ed., 1994) available from the American Psychological Association, P. O. Box 2710, Hyattsville, MD 20784. Manuscripts should be typed on one side only on 8-1/2 x 11 inch paper, double-spaced, with ample margins. Subheads should be used at reasonable intervals. Typescripts should typically run from 30 to 50 pages.

All manuscripts should be sent with an author identification, an abstract, a list of references, and, if necessary, notes.

Author Identification. The title of the article and the author's name should be typed on a separate page to ensure anonymity in the review process. Authors should prepare an autobiographical note indicating: full name, position, department, institution, mailing address, and specialization(s). Example follows:

Author

> JANE C. DOE, Assistant Professor, Foreign Language Education, University of America, 226 N. Madison St, Madison, WI 55306. Specializations: foreign language acquisition, curriculum studies.

Abstract. An abstract of not more than 200 words should identify the purpose of the article, provide an overview of the content, and suggest findings.

A list of references should be submitted on a separate page of the manuscript with the centered heading: *References*. The entries should be arranged alphabetically by surname of authors. The sample list of references below illustrates the format for bibliographic entries:

> Dulay, H. & Burt, M. (1974). Errors and strategies in child second language acquisition. *TESOL Quarterly*, *16* (1), 93-95.
>
> Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.

Reference citations in the text of the manuscript should include the name of the author of the work cited, the date of the work, and when quoting, the page numbers on which the materials that are being quoted originally appeared, e.g., (Jones, 1982, pp. 235-238). All works cited in the manuscripts must appear in the list of references, and conversely, all works included in the list of references must be cited in the manuscript.

Notes should be used for substantive information only, and they should be numbered serially throughout the manuscript. They all should be listed on a separate page entitled *Notes*.

Specifications for Floppy Disks

Where feasible, manuscripts are preferred on Windows-based software. Manuscripts produced on one of the DOS-based or Macintosh systems should be formatted as an ASCII file at double density, if possible.

When mailing a floppy disk, please enclose the following:

1. Word processing software used: _____

2. Disk is formatted as: double__ high density__

3. Remarks:_____

Review Process

Manuscripts will be acknowledged by the editor upon receipt. Following preliminary editorial review, articles will be sent to at least two reviewers whose area of expertise includes the subject of the manuscript. *Applied Language Learning* uses the blind review system. The names of reviewers will be published in the journal annually.

Copyright

[Table of Contents]