

# 1 *Language testing – SLA interfaces: An update*

Lyle F. Bachman  
Andrew D. Cohen

For some years, language testing (LT) research and second language acquisition (SLA) research have largely been viewed as distinct areas of inquiry in applied linguistics. Since the late 1980s, however, we have seen an increasing number of studies in which these two subfields of applied linguistics come together, both in terms of the substantive issues being investigated and the methodological approaches used. An overview of the areas of overlap, or interfaces, was provided in Bachman (1989; reprinted as an appendix in this volume). There have been many developments in both areas since then, and in this chapter we identify what we see as some of the continuing contrasts – contrasts that we view as natural and healthy outgrowths of the differing interests and foci of LT and SLA research. We then discuss several areas and issues that have been discussed more or less independently by LT and SLA researchers, and that we believe constitute *interfaces* between these two fields of applied linguistics, briefly illustrating these interfaces with some recent research studies. Finally, we list some of the challenges and questions that we feel suggest areas for future research for both LT and SLA. It is not our purpose to present a comprehensive review of the research literature in LT and SLA. Rather, we would like to suggest a conceptual framework and approach for building more extensive and substantial interfaces between LT and SLA research in the future, and point to some directions for addressing many of the complex and thorny issues that have, in the past, generally been dealt with in a rather compartmentalized manner by LT and SLA researchers.

## Different foci of SLA and LT research

Although the making of broad contrasts between traditional SLA research and traditional LT research necessitates some oversimplification, the following are several of the more noticeable tendencies that were mentioned by Bachman in 1989, and that we feel are still relevant today. SLA research has tended to concern itself with the description and explanation of how second language ability develops, focusing on interlanguage

2 *Lyle F. Bachman and Andrew D. Cohen*

as a language with rules and conventions in its own right and on the processes and factors that are involved in the development of interlanguage. LT research, on the other hand, has attempted to arrive at a model of language ability that can provide a basis for describing and assessing this ability for a given individual or group of individuals at a given stage of development, using a given norm or standard of target language use as a point of reference. While SLA research has concerned itself more with the factors and processes that affect or are part of language acquisition, LT research has tended to focus on components and strategies that are part of language ability. Thus, while SLA has looked for antecedents of language ability, LT research has studied the results of acquisition. The two fields also differ in their theoretical goals: SLA research has concerned itself with theories of language acquisition, while LT research has focused on theories of language test performance and the correspondence between test performance and nontest language use. In addition to having different substantive foci, the two fields have tended to use different research methodologies. SLA research has historically utilized the linguistic analysis of learners' interlanguage utterances, descriptive case studies, ethnographic research, and experimental and quasi-experimental designs, while LT research has more typically employed *ex post facto* correlational methods. These contrasts are summarized below:

*SLA research**Research perspective*

Longitudinal view of interlanguage development.

*Focus of research*

Antecedents of interlanguage ability: factors and processes that affect or are part of language acquisition, e.g., contextual features, learner characteristics, processes.

*Goals of research*

To develop and empirically validate a theory of SLA that will (1) describe how SLA takes place and (2) explain why SLA takes place.

*LT research*

“Slice of life” view of language ability at a given stage of development, with reference to a given norm or standard of language use.

Results of language acquisition: components and strategies that are part of language ability, e.g., grammatical competence, pragmatic competence, strategic competence.

(1) To develop and empirically validate a theory of language test performance that will describe and explain variations in language test performance; and (2) to demonstrate the ways in which language test performance corresponds to nontest language use.

*Research methodology*

Variety of research approaches: discourse analysis, case studies, ethnography, experimental, quasi-experimental, *ex post facto* correlational.

Dominant research approach: *ex post facto* correlational; increasing use of qualitative analysis of test content and test takers' responses, verbal reports of test taking.

## Interfaces: Some areas of common interest and cross-fertilization

Despite the differences in foci, several researchers (e.g., Bachman 1989; Shohamy 1994) have recognized that there is considerable commonality of interests and potential for cross-fertilization between LT and SLA researchers, both in terms of the research questions they address and the empirical approaches they take in dealing with these questions. These common areas of interest, or interfaces, can be seen as deriving from an issue that is of central concern to both LT and SLA research: describing and explaining variability in language acquisition (SLA) and test performance (LT). Ellis (1985) and Bachman (1990) have proposed frameworks for considering sources of variability in SLA and LT, respectively. Ellis considers all variations in language-learner language to be part of variability, and classifies these into two main categories: systematic variability, which includes individual variability and contextual variability, and nonsystematic variability, which includes what he calls *free variability* and *performance variability* (Ellis 1985: 75ff).

For Bachman, on the other hand, the distinction of prime interest and importance is not that between systematic variation and nonsystematic, or random, variation, but rather that between two types of systematic variation: (1) variation due to differences across individuals in their language ability, processing strategies, and personal characteristics (e.g., cultural and background knowledge, affective schemata), and (2) variation due to differences in the characteristics of the test method, or test tasks. Tarone (Chapter 3, this volume) makes a similar distinction, arguing that the term *individual difference* should be used for differences in performance across individuals, and the term *variation* should be reserved for “synchronic situation-related variation in the use of a second language.” Variation, she goes on to point out, “ought to be reserved to refer to shifts *within* the performance of any given individual” (p. 73). We agree that this distinction in terminology is useful, and will follow it in this chapter.

Considering the major sources of variability that have been identified and investigated in the two fields, clearly a large number of factors need to be considered in both LT and SLA research. However, we focus

4 *Lyle F. Bachman and Andrew D. Cohen*

here on three sources of variability that we believe are central to both fields:

1. Individual differences in the *language abilities* that are acquired or measured
2. Individual differences in the *strategies and other processes* that individuals employ in language use, as well as on language test tasks and SLA elicitation tasks
3. Variation in the *tasks and context* and their effects on language use, as well as on performance on language test tasks and SLA elicitation tasks.

*The nature of the language abilities acquired or measured*

HISTORICAL OVERVIEW OF THE EVOLUTION  
OF A MULTICOMPONENTIAL, INTERACTIONALIST  
VIEW OF LANGUAGE ABILITY

Let us begin this discussion by making a distinction between language and language ability. We would argue that it is not language per se that is measured or acquired, but language ability. That is, even though we speak of language acquisition and language testing, what we are primarily interested in, in both cases, is not the system of language itself, but rather the learner's capacity for acquiring and using a language system either for, or as part of, various processes, such as socialization, psychological orientation, and communication. We would further argue that both LT and SLA researchers make inferences or assumptions about the nature of language ability in their research, and that a clear definition of this construct is thus essential to both LT and SLA research. (See Chapelle, Chapter 2 of this volume, for a detailed discussion of the issues involved in defining language ability as a construct.)

Because of its historical focus on assessing language ability, with its attendant accountability, in terms of actual practice, LT research has understandably been generally more concerned with defining this construct than has the field of SLA. The view of language ability, or proficiency, that dominated the field of language testing during the 1960s and 1970s was one that derived largely from a structuralist linguistics view that saw language as being composed of discrete components (e.g., grammar, vocabulary) and skills (listening, speaking, reading, writing), and a trait psychology view of ability as a unidimensional attribute of which different individuals have greater or lesser amounts.<sup>1</sup>

1 Chapelle (Chapter 2, this volume) provides an excellent discussion of the trait perspective toward defining constructs. Lado (1961) and Carroll (1961) provide discussions of the so-called skills and components approach to defining second

*Language testing – SLA interfaces: An update* 5

Early SLA research, as did contrastive analysis and error analysis, drew on essentially the same linguistic paradigm for its theoretical grounding.<sup>2</sup> As the field developed, however, many SLA researchers fairly quickly embraced a *homogeneous competence model*, based largely on transformational, generative linguistics, in which linguistic competence is clearly distinguished from nonlinguistic, or pragmatic, knowledge, which “does not need to be taken into account when explaining linguistic competence” (Ellis 1985: 77). At the same time, other SLA researchers broadened their theoretical base to include what Ellis calls a *heterogeneous competence model*, in which “the user’s knowledge of the language rules is interlocked with his knowledge of when, where, and with whom to use them” (Ellis 1985: 77). This broader view of language ability as communicative, rather than linguistic, competence incorporates insights from both functional linguistics and sociolinguistics. In addition to enlarging their theoretical view of language ability, SLA researchers began looking more toward cognitive information-processing models of learning for theoretical inspiration.

Thus, as the fields of LT and SLA developed during the seventies and into the eighties, they paid allegiance to different theoretical paradigms, both in linguistics and in psychology. Language testing clung tenaciously to a skills and components model of language proficiency and a trait psychology view of ability, with a brief flirtation, in the early eighties, with the view of language proficiency as a single unitary trait. At the same time, SLA became much broader and more eclectic, embracing a wide range of theoretical perspectives in linguistics, sociolinguistics, sociology, anthropology, psychology, and cognitive science. Nevertheless, even though its theoretical base had expanded a great deal, much of the empirical research in SLA continued to focus rather narrowly on acquisition of the linguistic aspects of language ability, such as morphology, syntax, and lexicon, just as language testing continued to focus on these same elements. That is, even though LT and SLA research had become quite diverse in terms of the theoretical frameworks that generated their research questions and guided their interpretations, the data on language ability that the two sets of researchers collected were still very similar, focusing largely on language skills and language components.

The view of language ability as *communicative competence* currently

language ability. See Oller (1979) for a discussion of the “unitary trait hypothesis” of language ability, a hypothesis that he and the majority of language testers have long since abandoned.

- 2 See Lado (1957) and Wardhaugh (1970) for early discussions of contrastive analysis; see Corder (1967, 1971) and Richards (1974) for early discussions of error analysis. We would be remiss if we did not point out that neither of these concepts, in its original formulation, is current in SLA. The description of interlanguage differences and learner errors does, nevertheless, continue to be a part of SLA research, albeit within more recent linguistic paradigms.

6 *Lyle F. Bachman and Andrew D. Cohen*

accepted by many LT researchers represents a major paradigm shift from the structuralist, “skills and components,” trait view outlined earlier.<sup>3</sup> The notion of communicative competence, or communicative language ability, has evolved in the following four ways in the past decade:

1. Recognition of a much broader range of components
2. Recognition that these components are not discrete, and that they interact with each other
3. Consideration of how components of competence interact with other cognitive abilities and processes
4. Recognition that language ability includes the capacity for interacting with the context.<sup>4</sup>

Current views of communicative competence in applied linguistics can be traced, to a large extent, to Hymes’s (1972) work in this area and to Canale and Swain’s (1980) description, which itself drew heavily on work of Hymes (1972), Savignon (1972), and Halliday (1976). Although Canale and Swain (1980) and Canale (1983) provided an expanded view of the components involved in communicative competence, theirs was essentially a static framework, with little discussion of how these expanded components interacted with each other, if at all, or of how communicative competence enables language users to interact with the characteristics of the context of situation, including other language users, in a speech event. Bachman’s (1990) description of *communicative language ability* built upon the Canale and Swain description, retaining essentially the same components as theirs, but expanding the role of strategic competence, which Canale and Swain had considered to be limited largely to compensatory communication strategies for dealing with breakdowns in communication or for enhancing communication. Bachman’s description of the functions of strategic competence in planning, assessment, and execution provided a means for explaining how the various components of language competence (grammatical, textual, pragmatic, sociolinguistic) interact with each other, and with features of the language use situation.

A recent expansion of this framework is that of Bachman and Palmer (1996), which elaborates further the role of strategic competence as meta-cognitive strategies (goal setting, assessment, and planning) and includes a discussion of the roles of topical knowledge, or knowledge schemata, and affective schemata in language use. According to Bachman and Palmer’s formulation, *goal setting* involves the language user or test taker

3 See McNamara (1996) for a thorough historical treatment and critique of the models of language ability proposed by Hymes (1972), Canale, and Swain (1980), and Bachman (1990). Bachman (1990) also provides an overview of these models, as well as those of Lado (1961) and Carroll (1961).

4 This view of language ability is essentially what Chapelle (Chapter 2, this volume) calls an “interactionalist” approach to defining the construct.

*Language testing – SLA interfaces: An update* 7

in identifying and selecting one or more tasks that he or she might attempt to complete, and deciding whether or not to attempt to complete the task(s). The *assessment component* operates in three ways: (1) assessing the language use or test task to determine whether it is either desirable or feasible to attempt the task, and what linguistic and knowledge resources are likely to be required for this; (2) assessing the individual's own linguistic and knowledge resources to determine whether the knowledge needed is available for use; and (3) assessing the correctness or appropriateness of the response to the task. The *planning component* involves (1) the retrieval of the relevant items from linguistic and topical knowledge, (2) the formulation of one or more plans for responding to the task, and (3) the selection of one plan for initial implementation in a response (Bachman & Palmer 1996: 71–73).

This multicomponential view of language ability has provided a particularly productive theoretical basis for both research and practice in language testing. As part of the background research in preparation for the TOEFL 2000 project, Chapelle, Grabe, and Berns (1997) have provided a description of *communicative language proficiency* and discuss the implications of this model for test development. Several test development projects based explicitly on multicomponential definitions of language ability are included in Brindley (1995). McDowell (1995) uses Bachman's 1990 model of communicative language ability as a basis for developing an English proficiency test as part of a procedure for qualifying nonnative English speakers as language teachers in Australia. Clarkson and Jensen (1995) draw on Bachman's 1990 model to develop rating scales for specific task components, for the purpose of assessing learner achievement in English for occupational purposes courses. Grierson (1995) also utilizes Bachman's 1990 model, along with a systemic functional model of language use (Halliday 1985), to develop an observational checklist for use by teachers in assessing communicative ability in spoken discourse in secondary Intensive English Center classrooms in Australia. McKay (1995) uses Bachman and Palmer's (1996) more elaborated model of language ability, along with considerations of content and contextual features, as a basis for constructing rating scales for use in assessing the ESL proficiency of primary and secondary school students.

STUDIES THAT PROVIDE INSIGHT INTO SPECIFIC  
COMPONENTS OR ASPECTS OF LANGUAGE ABILITY  
AND THEIR MEASUREMENT

Now let us consider some studies in which the interplay of SLA and LT research has helped to enhance our understanding of some specific components or aspects of language ability and its acquisition and measurement. The following are some of the many areas that have been explored: the development and role of discourse domains in language acquisition

8 *Lyle F. Bachman and Andrew D. Cohen*

and test performance; the nature of second language (L2) ability, pedagogical factors that affect its development in classroom settings; and social and individual factors that affect its development in majority and minority language settings; interrelationships among L2 ability, foreign language aptitude, and intelligence; the communicative interaction among speakers and learners of a language; communicative grammar; breadth of lexical knowledge; type of cognitive activity in L2 essay writing; multilingual versus monolingual language learning; sociolinguistic ability; and cross-cultural pragmatic ability.

*Development and role of discourse domains in language acquisition and test performance.* The multicomponential view of language ability has provided a theoretical basis for research not only in LT but in SLA as well. Douglas and Selinker (1985; Selinker & Douglas 1985, 1989; Douglas 1986, and Chapter 6 in this volume), for example, draw on the assessment function of strategic competence to develop the notion of discourse domain as a dynamic, evolving sociopsychological construct that test takers may construct temporarily as a means for dealing with variation across different types of language test tasks. They suggest that these discourse domains may also provide a basis for mediating between the context of language use and language ability, becoming part of the learner's interlanguage competence.

*The nature of L2 ability and factors that affect its development in classroom settings, and in majority and minority language settings.* In the Development of Bilingual Proficiency (DBP) project (Allen, Cummins, Mougeon, & Swain 1983; Harley, Allen, Cummins, & Swain 1987, 1990), a massive longitudinal study that investigated issues relevant to both the assessment and the acquisition of bilingual proficiency, the Canale and Swain model of communicative competence provided the theoretical definition of language proficiency. The DBP project included both quantitative and qualitative studies, which investigated the nature of language proficiency, the pedagogical factors that affect its development in classroom settings, and the social and individual factors that affect its development in majority and minority language settings. With respect to the nature of language proficiency, the results were mixed: Although confirmatory factor analysis did not support the construct hypothesized by the Canale-Swain model, other analyses "did provide some support for the hypothesis that these constructs are distinguishable and also educationally relevant" (Harley, Cummins, Swain, & Allen 1990: 24). Other findings were that academic tests tended to be related, that academic skills were strongly related across languages, that cognitive factors were more strongly related to discourse competence and writing than to grammatical competence, and that "language proficiency must be conceptualized within

*Language testing – SLA interfaces: An update* 9

a developmental context as a function of interactions that students or learners experience in their languages” (p. 25).

*Interrelationships among L2 ability, foreign language aptitude, and intelligence.* In a study that included both quantitative and qualitative research approaches, Sasaki (1993a, b) investigated the relationships among components of second language proficiency and two individual factors that have long been investigated in the SLA literature: foreign language aptitude and intelligence. The results of her quantitative analyses provide support for a multicomponential model of second language proficiency, for a cognitive ability that includes both intelligence and language aptitude, and for Bachman’s (1990) hypothesis that second language ability “is related to but not identical with, general cognitive abilities” (Sasaki 1993a: 337). The results of her qualitative analyses of verbal report protocols suggest that a general factor of second language proficiency is related to the amount of information processing involved in correctly solving certain types of tasks. In addition, the verbal report protocols reveal that high proficiency subjects were better able to use available information to answer cloze test items correctly, spent more time planning, and used a greater variety of strategies than did the low proficiency subjects. This study thus provides some tantalizing insights into the nature of language proficiency and its relationship to general cognitive abilities. In addition, it demonstrates that by combining quantitative and qualitative approaches to research one can investigate both the process of problem solving and its product. Since SLA tends to focus on the process of language acquisition, and LT on its product, the combining of qualitative and quantitative approaches would appear to provide a particularly powerful paradigm for future research in this interface.

*Assessing word order through communicative interaction.* Perhaps one of the first areas in which SLA research informed the design of more communicative speaking tests was that of interactional analysis. SLA researchers wanted to capture interaction, but felt that traditional approaches to testing speaking were not assessing these interactions well, if at all. There is now a series of SLA studies that inform language testing because they suggest the kinds of interactive tasks that are likely to produce interactional data. An example of these studies is that reported by Ellis (1989), which looked at the extent to which classroom and naturalistic acquisition are the same.

The article reports a study of the classroom acquisition of German word order rules by adult foreign language learners of varied language backgrounds in two institutions in London. Data elicited by an information gap task performed by 39 foreign learners of German at two points in time were used to describe the sequence of acquisition of three

10 *Lyle F. Bachman and Andrew D. Cohen*

obligatory word order rules. The learners worked in pairs. A picture composition was cut up and two pictures were given to each learner. Each learner was asked to describe his or her pictures in German without showing them to the partner so that they could jointly work out the story, creating a two-way communication task. The pairs completed the task three times, the first serving as a warm-up, followed by each learner's producing a monologue of one of the complete stories. The tasks were conducted in a language laboratory and the stories were recorded, both as a pre- and a posttest.

Ellis compared the word order sequence observed with that reported in the literature on naturalistic learners of German and found no difference, despite the fact that the order in which the rules were introduced in the curriculum and the degree of emphasis given to the rules in the instruction differed from the naturalistic order. The classroom learners, however, did appear to be more successful than the naturalistic learners in that they reached higher levels of acquisition in a shorter period of time. The results of this study supported the claim that the classroom and naturalistic L2 acquisition of complex grammatical features such as word order follow similar routes. The results also suggest that classroom learners may learn more rapidly. In addition to the substantive findings about SLA, this study's use of paired tasks constituted an effort to use LT tasks that were genuinely communicative. The learners were recorded on a pre- and posttest basis by performing a task in a language laboratory whereby they had to communicate back and forth until they reconstituted picture compositions for which each learner had been given two pictures. Then in each case one of the learners had to tell the complete story in German. The elicitation task was designed to tap unmonitored, informal language use.

*Assessing grammar in communicative contexts.* Some studies have by their very nature been suggestive of how to design measurement tasks in areas that have been a challenge to assess. One such area is that of assessing grammatical competence in immersion education programs. In an effort to produce L2 speakers who have truly functional ability, immersion programs have at the same time produced speakers who lack control of grammar. Day and Shapson (1991) conducted a study to evaluate the degree to which focusing on form affects French immersion students' proficiency. Their contribution to the language testing literature was to demonstrate an approach to measuring form within a communicative situation.

Although far from unique, their efforts do constitute an example of applying SLA insights to test construction. The researchers designed a teaching/learning unit focusing on the conditional in natural, communicative situations, and administered it for three 45- to 60-minute peri-