

The ACTFL Proficiency Guidelines and the Oral Proficiency Interview: A Brief History and Analysis of Their Survival

Judith E. Liskin-Gasparro
University of Iowa

Abstract: The ACTFL Proficiency Guidelines (ACTFL, 1982; 1986; Breiner-Sanders et al., 2000) and the Oral Proficiency Interview (OPI) derived from them have stimulated abundant and sustained professional engagement by foreign language teachers at all levels and in all languages, as well as intense and equally sustained criticism by specialists in foreign language testing. This paper presents (1) a brief history of the genesis of the Guidelines, (2) a summary of the criticisms leveled against them, and (3) an analysis of the reasons for the continued significance of the Guidelines in curricular and testing initiatives during the last two decades at both state and national levels.

Introduction

The ACTFL Provisional Proficiency Guidelines (1982), the result of a collaboration between U.S. government testing agencies, the American Council on the Teaching of Foreign Languages (ACTFL), and the Educational Testing Service, brought to academic foreign language professionals a framework for understanding and measuring oral language ability that served as a catalyst for research, teaching, testing, and curriculum development projects. These projects emerged shortly after the provisional guidelines were published and the first workshops on oral proficiency interview testing were held. It is uncontroversial to state that today, just 20 years later, the terms *oral proficiency, OPI, and ACTFL Guidelines* are common currency in the discourse of foreign language teachers and preservice teacher candidates. Evidence of the impact of the proficiency guidelines can be seen at all levels and in all sectors of the foreign language teaching profession.

The creation of models of language proficiency, intended as the basis for theory building, test construction, and the design of instructional programs, is not a contemporary phenomenon, as attested by two recent historical studies of foreign language testing (Barnwell, 1996; Spolsky, 2000). In the recent past and in North America alone, our profession has produced models of proficiency at the rate of at least one or two per decade, including Oller's (1976) unitary competence proficiency model; Cummins's (1979) CALP/BICS model; Canale and Swain's (1980) communicative competence model; the ACTFL Proficiency Guidelines, first published in 1982 and then revised in 1986 and 1999; and Bachman's (1990) communicative language ability model (see Chalhoub-Deville, 1997, for brief descriptions). All of these models have sparked lively critical discussion.

Of the models mentioned here, the one that has given rise to the most sustained and broad-based criticism is the ACTFL Proficiency Guidelines (hereafter "Guidelines" or "ACTFL Guidelines"). Yet 20 years after they appeared in provisional form (ACTFL, 1982), the ACTFL Guidelines for speaking are alive and well. They have been institutionalized in foreign language professional circles in the United States through their prominence in the textbooks used in foreign language teacher preparation programs and, more recently, in the *Standards for Foreign Language Learning in the 21st Century* (National Standards, 1999). The national standards document has, in turn, served as the basis for the curriculum frameworks for foreign language instruction developed by 49 of the 50 states (as of this writing, Iowa is the only holdout).

Consequently, to the degree to which these state-level frameworks are serving as the basis for district- and program-level instructional planning, the ACTFL Guidelines will be implicitly incorporated into foreign language programs at all levels of public instruction. The mark of the Guidelines can also be seen in the first (upcoming in 2004) U.S. national assessment in foreign languages, created under the auspices of the National Assessment of Educational Progress (NAEP).¹

In this article, I will (1) review briefly the positive and negative reactions to the Guidelines following their introduction (1982), and (2) offer a modest proposal for the reasons that, despite the severity and well-foundedness of the criticisms leveled against them, the ACTFL Proficiency Guidelines have not fallen under their own weight, but instead have enjoyed a seminal role in curricular and testing projects at local, state, and national levels.

The ACTFL Proficiency Guidelines: Genesis and Critiques

The ACTFL Proficiency Guidelines and the OPI are based on a rating scale and an assessment procedure originally developed in the 1950s by the Foreign Service Institute of the U.S. Department of State and used since that time in various agencies of the U.S. government that are involved in teaching and measuring proficiency in a second language. As a result of a number of initiatives by government and academic groups, ACTFL and the Educational Testing Service received grants in 1981 to adapt the government's proficiency rating scales for use in schools and colleges. The result of their efforts, the ACTFL Provisional Proficiency Guidelines, was published in 1982. In that same year, ACTFL held the first oral proficiency tester training workshop for 30 college faculty members in Spanish and French. Thanks to an almost unprecedented infusion of federal, state, and institutional funds, tester training workshops proliferated throughout the 1980s and into the 1990s; they continue to be held on a regular basis today.

Participants in the early workshops were almost uni-

formly enthusiastic about the potential impact of the OPI on their language teaching practices. These participants, who included theoretical and applied linguists, as well as literature specialists from liberal arts colleges who taught both language and literature courses, found themselves listening with new ears to student language and the overall discourse of their classroom. New insights surfaced in reports to ACTFL—about how little teachers let students talk, how much they controlled the classroom discourse, and how heavily they relied on display questions to elicit grammatical forms from students. When the participants began conducting OPIs with their students and the students of their colleagues, they were struck by the mismatch between what they thought students knew and how little of it emerged when instructor controls on student talk were loosened. They were also struck by how hard it was during the OPIs not to instruct, not to correct, and otherwise not to dominate the talk (Scebold, 1982, personal communication).

The useful and productive trend towards reflection and self-examination that emerged from the participants' experiences in intensive OPI workshops and in postworkshop OPI interviews at their home institutions, coupled with a perhaps less useful, almost evangelical fervor, sparked a good deal of pedagogical innovation. The notion of proficiency as an "organizing principle for instruction," a term probably coined by Higgs (1984), became a mantra for the increasing number of language teachers at secondary and postsecondary levels who were actively exploring how to maximize student-to-student communication in their classrooms, as well as how to appropriately position the teaching of grammar as a tool or support skill, rather than as the centerpiece of instruction.

The exuberant enthusiasm for the ACTFL Guidelines and for the OPI was met by equally ardent opposition that could have been devastating. It is beyond the scope of this essay to analyze in detail all criticisms leveled against the Guidelines, but in the paragraphs below I summarize the major criticisms, which fall into two large categories.

The first major category of criticisms referred to the validity of the ACTFL Guidelines or, in the view of the critics, the absence thereof. It was objected that the Guidelines had been constructed based on "intuitive judgments" (Fulcher, 1996) rather than on any documented collection and analysis of empirical evidence. Hence, the features of the levels and the developmental trajectory posited by them were lucky approximations at best and mere fabrications at worst, but unacceptable as a model of speaking proficiency in a second language. It became clear, for example, that the proficiency guidelines in listening, reading, and writing were superficial spin-offs of the speaking guidelines, and that they had no empirical basis. Various studies in which student performance was analyzed in terms of the Guidelines (Dandonoli & Henning, 1990; Henry, 1996; Thompson, 1995; Valdés et al., 1992) could not find sup-

port for the skill hierarchy in writing or listening posited by the ACTFL Guidelines in those areas.

A second, related issue was the problem of circularity, most strongly articulated by Lantolf and Frawley (1985). According to their argument, the proficiency levels exist only in terms of the linguistic criteria that define them; the definitions are expressed exclusively in terms of the descriptions of the levels (Lantolf & Frawley, 1985, p. 340). In other words, I am an Intermediate-High speaker of French because I can perform the tasks associated with the Intermediate-High level with the requisite accuracy; and these tasks and this degree of accuracy are associated with the Intermediate-High level because people like me, who are rated Intermediate-High, are able to do them. This circularity reduces the ACTFL Guidelines, and indeed all models of proficiency that are expressed as hierarchical scales, to useless tautology. The problem recurs when it comes to testing. As Bachman and Savignon (1986) and Bachman (1988) argued, the Guidelines were confounded with the method of measurement. In other words, the OPI is used to measure speech samples in terms of the Guidelines, and the Guidelines represent what is measured in the OPI (Chalhoub-Deville, 1997, p. 9).

A third criticism related to the validity of the rating scale concerned positing the perception of the native speaker as a criterion against which the proficiency of non-native speakers would be measured. This hypothetical native speaker appears in various guises in the Guidelines—speakers can qualify for the Intermediate-High level if, among other things, they “can generally be understood by native speakers unaccustomed to dealing with nonnatives” (Breiner-Sanders et al., 1999, p. 16); and the errors of speakers at the Superior level “do not distract the native interlocutor” (Breiner-Sanders et al., 1999, p. 14). In the view of the critics, the absence of data to operationalize key phrases in the descriptions like “native speakers unaccustomed to dealing with nonnatives” and “distract the native interlocutor” renders these criteria imprecise and indefensible (Barnwell, 1988; Lantolf & Frawley, 1985; Salaberry, 2000).

A related issue, interrater reliability, provided an additional source of skepticism about the OPI. Early studies conducted at the Educational Testing Service (ETS) (Liskin-Gasparro, 1983) and by Magnan (1987) found high correlations in ratings of interviews in French (and, in the ETS study, in Spanish) assigned by apprentice testers and their expert trainers. Magnan (1987) found an even higher correlation between ratings assigned by two master certified testers. The results of her two studies led Magnan to conclude that “the ACTFL proficiency interview can be rated in a reliable fashion, and that the ACTFL intensive workshops, follow-up practicums, and certification procedures can produce a pool of testers who give reliable ratings” (p. 536). At the same time, she cautioned that these encouraging results needed to be tempered by acknowl-

edgment of small sample size, as well as by concern for other reliability issues, such as inter-interviewer reliability and the constancy of rating standards over time (p. 536).

Both of the 1980s studies cited here were conducted using a small number of testers and languages, an understandable limitation when academic use of the OPI was in its infancy. Thompson (1995) undertook a large-scale study to examine the interrater reliability of ACTFL-certified testers in five European languages that was based on 795 double-rated interviews (p. 407). Although Thompson, like Magnan nearly a decade earlier, found an overall high correlation between the ratings assigned by certified testers, she also found patterns of discrepancy between ratings assigned by interviewers and second raters: The interviewers, who rated the performance after conducting the face-to-face interview, tended to rate the speech performances higher than did the second raters, who assigned their ratings after listening to the interviews on audiotape. Thompson also found differential patterns of interrater agreement across languages and across levels within a language. It also appears that disagreements between raters across the so-called major borders (Novice, Intermediate, Advanced, Superior) were frequent, more so than indicated by earlier reliability studies. Thompson’s results highlight the complexity of rating OPI interviews and suggest that when the OPI is used for high-stakes decision making, such as in the work contexts of hiring and promotion, current practices may put certain classes of examinees (e.g., those whose proficiency is close to a major border on the ACTFL scale) at a disadvantage.

A second major category of criticism concerned the nature of the language elicited by the OPI, particularly in light of claims that the OPI measured “language performance in terms of ability to use the language effectively and appropriately in real-life situations” (ACTFL, 1986, 1–1, as cited by Salaberry, 2000, p. 299). Critics charged that the OPI assessed only a limited range of language tasks. Savignon (1985), for example, noted the absence of such language-use contexts as small-group discussions and games, which tapped discourse strategies that the OPI did not sample (p. 132). Similarly, Shohamy et al. (1986) advocated multiple-measures proficiency assessments on the grounds that tasks like discussions, reports, interviews, and conversations were different enough to merit being included in an assessment of global language proficiency. Given that test results are meant to predict performance, promoters of the OPI were called to task for their willingness to “predict performance across [a range of] conversational contexts, interlocutors, topics, and purposes” (Kramsch, 1987, p. 358, as cited by Salaberry, 2000, p. 299) on the basis of a single discourse event; that is, the face-to-face interview.

Criticisms of the OPI from a critical discourse perspective claimed that although the OPI purported to resemble a casual, friendly conversation, the face-to-face

interview format in fact instantiates an unequal power relationship between interviewer and interviewee, in which the rules of interview discourse constrain the right to such discourse management moves as nominating topics, eliciting information, and interrupting (Johnson, 2001; Kramsch, 1986; van Lier, 1989). According to these critics, the interview format of the OPI, broken only by the role play, created a limited and artificial picture of the speakers' interactional skills.

A corollary to the charge that the OPI failed to address significant features of interactional discourse was the criticism that it put excessive emphasis on grammatical accuracy. In an early article on the OPI, Higgs and Clifford (1982) introduced what they termed the "relative contribution model" (p. 69), which depicted graphically the prominence of vocabulary and grammar—relative to pronunciation, fluency, and sociolinguistic competence—in the profile of speaking proficiency at all but the highest level (i.e., level 5) on the U.S. government's oral proficiency scale. Participants in early OPI testing workshops in the European languages received copies of Lowe's "grammar grids," tables consisting of "lists arranged by proficiency level of sample grammatical features whose absence, semi-mastery, or level of mastery tended to contribute to the global rating" (Lowe, 1985, p. 26). These tables of grammatical features were considerably more developed than any similar documents outlining the hypothesized features at the various proficiency levels of the other factors.

Notwithstanding Byrnes's (1987) explanation of the role of grammatical accuracy in the assignment of a proficiency level to a speech sample and her assertion that "the charge that proficiency testing promotes an additive, discrete-point system of assessing language use that is hung up on grammar and uses grammar as an unrealistic and even elitist distinguisher between speakers [was] simply inappropriate" (p. 111), the perception remained among some critics (e.g., Raffaldini, 1988, p. 199) that, despite its claims to the contrary, the OPI was a test of grammatical accuracy-in-use.

The ACTFL Guidelines as "Proficiency": The Instructional Perspective

Given such broad-based, sustained, and convincing criticism, how can we explain that the ACTFL Proficiency Guidelines are still with us? Not only are they still with us, but they have become an integral part of the professional landscape of foreign language instruction, test development, and policy making in the United States. In the sections below, I discuss three ways that we may understand this perplexing phenomenon.

A Catalyst for Change

With 20–20 hindsight, it is easy to paint the past with a broad brush. In the absence of a thorough historical analysis,

which has yet to be undertaken, to say that the ACTFL Guidelines and the OPI were a catalyst for major change in foreign language teaching at all levels—the kind of broad, systemic change that Swaffar, Arens, and Byrnes (1991) have termed a "paradigm shift"—is to rely on mythology and to attend to only a small piece of our collective professional history. But mythology is a powerful operator in creating and sustaining beliefs and belief systems. During the decade after the introduction of the ACTFL Guidelines and the OPI, the activity surrounding them generated the widespread belief that they had something positive to offer foreign language education.

Consequently, the so-called proficiency movement and the ACTFL Guidelines and the OPI that served as its emblems, sparked significant change in the foreign language field that resonated even at the local classroom level. ACTFL, the principal national professional organization, grew significantly in both membership and funding base as it began to offer workshops on testing for oral proficiency, teaching for proficiency, and incorporating insights from the ACTFL Guidelines into curriculum planning and classroom testing.

A new generation of pedagogical materials that contained such previously unheard-of features as multiple presentations of key grammatical structures in the same textbook, student-to-student interviews, set-ups for role plays and skits, and strategy instruction addressed directly to students in marginal annotations was published. The books, billed as "proficiency oriented," became best sellers in the U.S. foreign language textbook market. Their prominence, along with ACTFL's workshops and publications, contributed to the institutionalization of the OPI and the ACTFL Guidelines (Lantolf & Frawley, 1985, p. 183) and the dissemination of the belief that the profession had entered a new era. People on both sides of the ACTFL Guidelines debate jumped into the controversies in the borderland between theory and practice about such topics as explicit versus implicit teaching of grammar, learnability versus teachability, acquisition versus learning, and the utility of error correction.

Familiarity with the ACTFL Guidelines, exposure to talk about second language acquisition at workshops and conferences, for some, experience with conducting and rating OPIs, gave classroom teachers with little or no background in second language acquisition a voice in the professional discourse. They had seen for themselves the disconnect between students' declarative linguistic knowledge, demonstrated in contexts of controlled production in teacher-centered classrooms, and their more limited ability to communicate in autonomous contexts. They had participated in heated discussions about error correction and grammar instruction, and they had been challenged to think past the received wisdom about methods and classroom practices and to step back and observe

both themselves and the discourse of their classrooms. For the average classroom teacher, the ACTFL Guidelines were a site of creative work and stimulating professionalization.

In their review of approaches to foreign language instruction in the twentieth century, Mitchell and Vidal (2001) used the metaphor of a flowing river to account for the impact of the proficiency movement. According to their analysis, the proficiency movement promoted the conjoining of the streams of communicative language teaching and of language assessment, which had begun to influence each other in the late 1970s and early 1980s (p. 32). As a result, earlier polemics over teaching methodology (audiolingual versus communicative), as well as contemporary divisions about accuracy versus fluency, could be absorbed by the proficiency movement's attention to performance-based outcomes. Rather than promote a single teaching method or approach, the proficiency movement focused on the skills and knowledge that students could deploy in communicative contexts, not on how the knowledge and skills might have been acquired. Within this framework, the ACTFL Guidelines and the OPI offered tools to evaluate teaching practices. By serving as agents of change, they established themselves as part of the bedrock on which future innovations would be based.

The principal change agents of the 1990s are a case in point. The presentation of the *Standards for Foreign Language Learning: Preparing for the 21st Century* (National Standards, 1996), which was followed three years later by the multilanguage volume *Standards for Foreign Language Learning in the 21st Century* (National Standards, 1999), appear to take their inspiration in part from the ACTFL Guidelines but, at the same time, to go beyond the Guidelines in their vision of language teaching, learning, and assessment. Whereas the proficiency guidelines and the OPI had been the exclusive purview of the language teaching profession, the collaborative, federally funded standards project was one of many such initiatives undertaken concurrently in a range of academic subject areas. By joining the K-12 standards movement, the foreign language profession was staking out a place for itself as a core subject, on a par with language arts, math, and social studies. Further, by emphasizing the links between language learning and learning in other subject areas, the language profession was asserting that "language and culture education is part of the core curriculum" (National Standards, 1999, p. 7).

Of the five educational goal areas—communication skills, understanding the cultures associated with languages, interconnectedness of language with other bodies of knowledge, comparisons that offer insight into the nature of language and culture, and participation in multilingual communities—only in the communication standard can the link to the ACTFL Guidelines be clearly seen. Indeed, the writers of the standards document downplay

the connection. In a section that describes the development of the standards, we find this statement:

Teachers will recognize the influence of the guidelines within the standards, particularly in the area of communication. However, in keeping with the attempt to create broadly conceived standards, communication is organized around a framework of interpersonal, interpretive, and presentational modes, rather than carved into separate skill areas of listening, speaking, reading, and writing. (National Standards, 1999, pp. 13–14)

Similarly, in a chapter on the communication standard, Hall (1999) emphasized the contrast between the focus of the ACTFL Guidelines on the four traditional language skills and the standards, which "frame communicative development in terms of three general communicative domains" (p. 45).

Despite the assertions that distance the standards of the 1990s from the proficiency movement of the 1980s, it is notable that a single organization, ACTFL, was at the forefront of both professional initiatives and that the leaders of the standards development project were individuals also well versed in the Guidelines and the OPI. It would not be a stretch to imagine that in the mind of the average classroom teacher, whose professional involvement outside her district may be limited to attendance at the state foreign language association conference, communication looms largest of the five standards and that it bears a close connection to the Guidelines and the OPI.

Testing

A second area in which the impact of the Guidelines and the OPI has persisted is testing. As North (1993) and Chalhoub-Deville (1997) have discussed, there is no automatic path from a theoretical model of proficiency to an operational model for test construction. According to North (1993), "an operational model is always simpler than a theoretical model, and . . . it may reinterpret elements [of the theoretical model] to make them more accessible in a particular context" (p. 7). This process has happened informally, perhaps in theoretically unmotivated ways, with the ACTFL Guidelines. The rating scale, particularly the summary statements about the functions and text types that characterize each of the levels, provide a common "frame of reference to describe achievement in a complex system in terms meaningful to all the different partners in or users of that system" (North, 1993, p. 6). The emphasis in the OPI on formats and techniques that maximize student language production, along with its focus on tasks rather than linguistic accuracy alone, inspired the designations "prochievement tests" (Gonzalez Pino, 1989) and "hybrid tests" (Omaggio Hadley, 1986). These locally developed measures, intended for classroom use, also attempt to maximize the production of language

by the student in free-response item types, while still remaining within the constraints of the particular portion of curricular content for which students are to be held responsible.

The ACTFL Guidelines and the OPI have similarly inspired the development of large-scale tests at state levels and, in the upcoming NAEP Assessment in Foreign Language, at the national level as well. An inspection of state-level performance benchmarks and sample tasks (e.g., Wisconsin Department of Public Instruction, 2001) reveals concepts and verbiage borrowed from the Guidelines, yet another indication of the degree to which they have become ingrained in the professional imagination.

Research: Good Data and Problems of Interest

A third reason that may explain why the OPI has flourished as an evaluation measure, despite the well-documented problems with the rating scale as a theoretical model of proficiency, is that OPI speech samples make for rich data. Examinations of OPI discourse have contributed to the growing body of knowledge about such topics as the nature of the language that is produced in face-to-face and semi-direct oral tests (e.g., Koike, 1998; Shohamy, 1994) and how the discourse of oral tests is constructed jointly by the interviewer and examinee (e.g., Ross & Berwick, 1992; Ross & Kasper, 1998). These topics of theoretical significance also have immediate practical application. Researchers in search of data for studies that employ a cross-sectional design may find that OPI speech samples at different ratings serve as better indicators of levels of acquisition than do course placement or course-based achievement measures.

Research on language acquisition during study abroad has used speaking samples drawn from OPIs as data on which to base measures of fluency (Freed, 1995) and pragmatic competence (Lafford, 1995), in addition to global measures of gain in speaking ability during a period abroad. One particularly interesting study undertaken after the OPIs of North Americans studying in Russia revealed a puzzling pattern of male students more frequently attaining high oral proficiency ratings than women, even though women were initially the stronger speakers. Through an analysis of students' diary entries, Polanyi (1995) uncovered pervasive differences in the social and, therefore linguistic, experiences of male and female students. Whereas male students managed easily to join social groups, make friends, and engage in a range of recreational activities with Russians, female students did not have the same easy access. They reported in their journals that their attempts to engage Russians in conversation were often rebuffed or interpreted as sexual advances, and that consequently they did not have access to the same range and quantity of social contact and opportunities for talk on a variety of topics with native speakers that their male counterparts enjoyed. In this study, the OPI ratings were the starting point of the

research, which then led to the formation of hypotheses and the examination of other types of data.

Finally, the Guidelines themselves are a rich area for investigation. Researchers who seek to verify the hierarchy or the accuracy of the level descriptions use OPI speech samples as a crucial source of data on which discourse analyses are conducted (e.g., Liskin-Gasparro, 1993, 1996). Similarly, studies of interrater reliability (cf. the discussion of Magnan, 1987, and Thompson, 1995, above) can contribute to an increased understanding of the factors that enter into the evaluation process. Although the many unknown factors that contribute to the assignment of ratings have sparked criticism of the OPI and reliance on it for professional decision making, they also offer rich areas for investigation and the development of new knowledge.

Conclusion

A review of the research and critiques on the ACTFL Guidelines and the OPI (see Freed, 1988; Liskin-Gasparro, 2000) leads to the conclusion that proficiency as described in the ACTFL Guidelines is far less absolute and broad based than the early proponents had claimed. Its empirical basis is shaky, and its claim to be conversational does not hold up. It is clear that oral proficiency ratings predict far less about an individual's future performance in a range of communicative situations than promoters of the OPI would like them to.

I have outlined three possible explanations for the persistence and even expansion of the impact of the Guidelines and the OPI, despite their limitations. To summarize, the discussion suggests that the notion of speaking proficiency that was conveyed by the Guidelines and disseminated by ACTFL and by federal funding agencies has become ingrained in the teaching, testing, programmatic, and research activities of a significant segment of the foreign language profession. Although the drive to give speaking and communication a central place in the foreign language curriculum began with the audiolingual revolution, at least two decades before the advent of the ACTFL Guidelines,² nevertheless a kind of folk pedagogy has emerged in the consciousness of today's teachers that associates with proficiency (and, by extension, with the ACTFL Guidelines) all manner of teaching practices that are considered communicative, educationally progressive, and culturally authentic. For example, in a 1993 survey conducted with university- and secondary-level foreign language supervisors (Birckbichler & Corl, 1993), respondents were asked to indicate the degree to which particular activities reflected a "proficiency orientation." The following surfaced as highest on the list: pair and small-group activities, role plays and simulations, information gap activities, cooperative learning, free writing, and cultural units.

These results seem quite odd at first blush. Some of the highest-rated items—pair and small-group activities, role

plays and simulations, and information gap activities—are directly related to the development of oral skills. They seem to reflect a consensus that speaking should occupy a central place in the language classroom, and it is not hard to trace them to the ACTFL Proficiency Guidelines. But free writing and cultural units, which do not figure in the Guidelines, also received high ratings for proficiency orientation. This seems to reflect a different understanding—that the ACTFL Guidelines are somehow responsible for all practices in language instruction that focus on something other than linguistic accuracy.³ If the goal of instruction is to have students reason, interact, and learn through the medium of the target language, even in early levels of instruction (Swaffar et al., 1991), then approaches and activities that enhance processes of cognition and communication naturally fall under the rubric of proficiency. Free writing, which is a kind of brainstorming on paper that enables students to clear their minds of linguistic concerns and focus on ideas, would logically fit within the cognition and communication framework, as would culture learning and the use of authentic materials in language teaching.

In conclusion, we see that the notion of proficiency as construed by the ACTFL Guidelines, which have been heavily criticized as a theoretical construct in second language acquisition and language testing research, seems to have found its legitimacy in the arenas of policy, program development, and classroom instruction. This frankly non-technical and atheoretical stance to speaking as proficiency and, by extension, to foreign language instruction as proficiency, presents us with interesting areas for future investigation.

Acknowledgments

An earlier version of this article was presented at the joint colloquium of the American Association for Applied Linguistics/Language Testing Research Colloquium (AAAL/LTRC) in St. Louis, MO, in March 2001. I am grateful to an anonymous reviewer of this paper for insightful observations and helpful suggestions.

Notes

1. For information on the NAEP Foreign Language assessment, see www.nagb.org.
2. I am grateful to an anonymous reviewer of this paper for making this observation.
3. An anonymous reviewer of this paper has pointed out the irony that the Guidelines appear to have been both criticized for being overly grammar driven and lauded for moving instruction away from a focus on grammar.

References

- American Council on the Teaching of Foreign Languages. (1982). *ACTFL provisional proficiency guidelines*. Yonkers, NY: ACTFL.
- American Council on the Teaching of Foreign Languages. (1986). *ACTFL proficiency guidelines*. Yonkers, NY: ACTFL.
- Bachman, L. F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition*, 10, 149–64.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Savignon, S. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal*, 70, 380–90.
- Barnwell, D. (1988). Proficiency and the native speaker. *ADFL Bulletin*, 20(1), 42–46.
- Barnwell, D. (1996). *A history of foreign language testing in the United States from its beginnings to the present*. Tempe, AZ: Bilingual Press.
- Birckbirchler, D. W., & Corl, K. A. (1993). Perspectives on proficiency: Teachers, students, and the materials that they use. In J. K. Phillips (Ed.), *Reflecting on proficiency from a classroom perspective* (pp. 115–58). Lincolnwood, IL: National Textbook Co.
- Breiner-Sanders, K. E., Lowe, Jr., P., Miles, J., & Swender, E. (2000). ACTFL proficiency guidelines—speaking, revised. *Foreign Language Annals*, 33, 13–18.
- Byrnes, H. (1987). Second language acquisition: Insights from a proficiency orientation. In H. Byrnes & M. Canale (Eds.), *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp. 107–31). Lincolnwood, IL: National Textbook Co.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47.
- Chalhoub-Deville, M. (1997). Theoretical models, assessment frameworks and test construction. *Language Testing*, 14, 3–22.
- Cummins, J. (1979). Cognitive academic language proficiency, linguistic interdependence, the optimum age question, and some other matters. *Working Papers on Bilingualism*, 19, 197–205.
- Dandonoli, P., & Henning, G. (1990). An investigation of the construct validity of the ACTFL proficiency guidelines and oral proficiency procedure. *Foreign Language Annals*, 23, 11–22.
- Freed, B. (1988). Perspectives on the future of proficiency-based teaching and testing. *ADFL Bulletin*, 20, 52–57.
- Freed, B. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–48). Philadelphia: Benjamins.
- Fulcher, G. (1996). Invalidating Validity Claims for the ACTFL Oral Rating Scale. *System*, 24, 163–72.
- Gonzalez Pino, B. (1989). Prochievement testing of speaking. *Foreign Language Annals*, 22, 487–96.
- Hall, J. K. (1999). The communication standards. In J. K. Phillips (Ed.), *Foreign language standards: Linking research, theories, and practices* (pp. 15–56). Lincolnwood, IL: National Textbook Co.

- Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *Modern Language Journal*, 80, 309–26.
- Higgs, T. V. (1984). Introduction: Language teaching and the quest for the holy grail. In T. V. Higgs (Ed.), *Teaching for proficiency, the organizing principle* (pp. 1–9). Lincolnwood, IL: National Textbook Co.
- Higgs, T. V., & Clifford, R. T. (1982). The push toward communication. In T. V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57–79). Lincolnwood, IL: National Textbook Co.
- Johnson, M. (2001). *The art of non-conversation: A reexamination of the validity of the oral proficiency interview*. New Haven, CT: Yale University Press.
- Koike, D. A. (1998). What happens when there's no one to talk to? Spanish foreign language discourse in oral proficiency interviews. In R. Thompson & A. Weiyun-He (Eds.), *Talking and testing: Discourse Approaches to the Assessment of Oral Proficiency* (pp. 69–98). Philadelphia: Benjamins.
- Kramsch, C. (1986). From language proficiency to interactional competence. *Modern Language Journal*, 70, 366–71.
- Kramsch, C. (1987). The proficiency movement: SLA perspectives. [Review of *Teaching language in context: Proficiency-oriented instruction and Defining and developing proficiency: Guidelines, implementations, concepts.*] *Studies in Second Language Acquisition*, 9, 355–62.
- Lafford, B. (1995). Getting into, through and out of a situation: A comparison of communicative strategies used by students studying Spanish abroad and 'at home.' In B. F Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 97–121). Philadelphia: Benjamins.
- Lantolf, J. P., & Frawley, W. (1985). Oral proficiency testing: A critical analysis. *Modern Language Journal*, 69, 337–45.
- Liskin-Gasparro, J. E. (1983). *Professional development: Oral proficiency testing and rating*. Final report of U.S. Department of Education Grant. Princeton, NJ: Educational Testing Service.
- Liskin-Gasparro, J. E. (1993). Talking about the past: An analysis of the discourse of intermediate high and advanced level speakers of Spanish. Doctoral dissertation; University of Texas at Austin.
- Liskin-Gasparro, J. E. (1996). Circumlocution, communication strategies, and the ACTFL proficiency guidelines: An analysis of student discourse. *Foreign Language Annals*, 29, 317–30.
- Liskin-Gasparro, J. E. (2000). The proficiency movement: Current trends and a view to the future. In O. Kagan & B. Rifkin (Eds.), *The learning and teaching of Slavic languages and cultures* (pp. 9–28). Bloomington, IN: Slavica.
- Lowe, Jr., P. (1985). The ILR proficiency scale as a synthesizing research principle: The view from the mountain. In C. J. James (Ed.), *Foreign language proficiency in the classroom and beyond* (pp. 9–53). Lincolnwood, IL: National Textbook Co.
- Magnan, S. S. (1987). Rater reliability of the ACTFL oral proficiency interview. *French Review*, 43, 525–37.
- Mitchell, C. B., & Vidal, K. E. (2001). Weighing the ways of the flow: Twentieth century language instruction. *Modern Language Journal*, 85, 26–38.
- National Standards in Foreign Language Education Project (1999). *Standards for foreign language learning: Preparing for the 21st century*. (1996). Yonkers, NY: Author.
- National Standards in Foreign Language Education Project. *Standards for foreign language learning in the 21st century*. Yonkers, NY: Author.
- North, B. (1993). *The development of descriptors on scales of language proficiency*. Washington, DC: National Foreign Language Center.
- Oller, J. W., Jr. (1976). Evidence of a general language proficiency factor: An expectancy grammar. In J. W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 3–10). Rowley, MA: Newbury House.
- Omaggio Hadley, A. (1986). *Teaching language in context: Proficiency-oriented instruction*. Boston: Heinle & Heinle.
- Polanyi, L. (1995). Language learning and living abroad. In B. F Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 271–91). Philadelphia: Benjamins.
- Raffaldini, T. (1988). The use of situation tests as measures of communicative ability. *Studies in Second Language Acquisition*, 10, 197–216.
- Ross, S., & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14, 159–76.
- Ross, S., & Kasper, G. (1998). Topic management in oral proficiency interviews. Paper presented at AAAL conference, Seattle.
- Salaberry, R. (2000). Revising the revised format of the ACTFL oral proficiency interview. *Language Testing*, 17, 289–310.
- Savignon, S. J. (1985). Evaluation of communicative competence: The ACTFL provisional proficiency guidelines. *Modern Language Journal*, 69, 129–34.
- Shohamy, E. (1988). A proposed framework for testing the oral language of second/foreign language learners. *Studies in Second Language Acquisition*, 10, 165–79.
- Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing*, 11, 99–123.
- Shohamy, E., Reves, T., & Bejerano, Y. (1986). Introducing a new comprehensive test of oral proficiency. *ELT Journal*, 40, 212–20.
- Spolsky, B. (2000). Language testing in *The Modern Language Journal*. *Modern Language Journal*, 84, 536–52.
- Swaffar, J., Arens, K., & Byrnes, H. (1991). *Reading for meaning: An integrated approach to language learning*. Englewood Cliffs, NJ: Prentice Hall.
- Thompson, I. (1995). A study of interrater reliability of the ACTFL oral proficiency interview in five European languages: Data from ESL, French, German, Russian, and Spanish. *Foreign Language Annals*, 28, 407–22.
- Valdés, G., Haro, P., & Echevarriarza, M. P. (1992). The development of writing ability in a foreign language: Contributions toward a general theory of L2 writing. *Modern Language Journal*, 76, 333–52.
- van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversations. *TESOL Quarterly*, 23, 489–508.
- Wisconsin Department of Public Instruction. (2001). *Wisconsin's model academic standards for foreign languages*. Available online at <http://www.dpi.state.wi.us/dpi/standards/pdf/fl.pdf>. Accessed on September 10, 2003.