Exploring the Uses and Usefulness of ACTFL Oral Proficiency Ratings and Standards in College Foreign Language Departments

John M. Norris
Northern Arizona University

Peter C. Pfeiffer
Georgetown University

Abstract: *Oral proficiency ratings, based on the ACTFL Guidelines and derived from live or simulated oral proficiency interviews (SOPIs), enjoy widespread use in the United States. In particular, college foreign language departments have increasingly adopted oral proficiency ratings as a way of establishing standards for language or graduation requirements. In the study reported here, the authors explored the intended uses of proficiency-based standards in foreign language departments and reviewed the research on which specific ACTFL-level standards are based. They then examined the results of more than 100 SOPIs administered across all levels of instruction within one German foreign language department. The findings suggest that recommended proficiency standards may underestimate the potential and actual achievements of German language learners and miss other valued learning outcomes. The implications of these findings for the valid use of oral proficiency ratings in collegiate settings are discussed.*

Introduction

Oral proficiency ratings, based on the ACTFL Proficiency Guidelines (ACTFL, 1986; Breiner-Sanders et al., 2000) and derived from live or simulated oral proficiency interviews (SOPIs), enjoy widespread use in secondary and higher education as well as in numerous professional contexts within the United States. As curriculum-independent, widely recognized assessment procedures, oral proficiency ratings are attractive tools at a time when accountability for learning outcomes is an increasingly urgent need for teachers, administrators, legislators, students, and the public. Within this context, college foreign language departments have adopted oral proficiency ratings and assessment methods to meet several purposes within their programs and institutions. Foremost among these is that of establishing standards that help define institutional foreign language requirements or graduation requirements for language majors.

In this article, we first explore why and how proficiency standards have been adopted in collegiate foreign language departments and briefly review the research on which specific ACTFL Guidelines' proficiency standards are based, paying particular attention to German foreign

---

John M. Norris (PhD, ABD, University of Hawaii) is an adjunct faculty member in the English Department, Northern Arizona University, Flagstaff, Arizona.

Peter C. Pfeiffer (PhD, University of California, Irvine) is Associate Professor of German at Georgetown University, Washington, DC.

language contexts. We then examine the relationship between semesters of college foreign language study and students' ratings based on the ACTFL Guidelines' oral proficiency scales, reporting the results of more than 100 SOPIs collected across all levels of instruction within one German department.

In light of our findings, which suggest that recommended proficiency standards may seriously underestimate and delimit the potential and actual achievements of students of German, we discuss fundamental implications for the valid use of oral proficiency ratings in collegiate settings. We conclude by pointing out the need for college foreign language departments to address the critical relationship between setting valuable learning standards, developing curriculum and instruction that enables students to attain these standards, and engaging in assessment that illuminates and fosters student learning.

Proficiency Standards and Assessment in College Foreign Language Programs

Many institutions continue to define foreign language requirements using a "seat-time" standard, according to which students must enroll in and pass a certain language course sequence or its equivalent to fulfill mandated degree or general education credit hours. Although the origins of the foreign language requirement in U.S. institutions of higher education are not entirely clear, the two years of language study standard has had a lengthy (Wolfe, 1959), at times controversial (Fleck, 1970; Klayman, 1978), but nevertheless persistent presence in mainstream undergraduate education (Huber, 1992; Brod & Huber, 1996).

Since the 1980s, a number of foreign language programs and institutions have established language proficiency or competency graduation standards in lieu of the typical two-year study requirement. These moves have been prompted by perceived weaknesses and problems in the seat-time standard, by evolving concepts of effective undergraduate education, and by the foreign language proficiency movement (Barnes, Klee, & Wakefield, 1990; Freed, 1981, 1984, 1987; Lange, 1990; Schulz, 1988). Problems such as student and instructor apathy, divisiveness between lower-level ('language' teaching) faculty and upper-level ('literature' teaching) faculty, time and monetary costs of credit-heavy lower-division language courses, and the concern that students are not acquiring adequate levels of language ability have led some institutions to supplant seat-time standards with language proficiency standards, to be demonstrated via assessment. Such changes have also been supported in some institutions by a transition to competency-based standards and assessment for all general education or liberal studies requirements (e.g., Erwin & Wise, 2001; Voorhees, 2001). Given these developments, a widely recognized, program-independent, and easily communi-

cated set of standards in the form of the ACTFL Proficiency Guidelines (ACTFL, 1986; Breiner-Sanders et al., 2000) and associated assessment procedures has provided a readily available solution for many foreign language programs.

What Proficiency Standards Have Been Set?

By way of example, we highlight the use and assessment of standards based on the ACTFL Proficiency Guidelines in four institutions that reflect the range of collegiate contexts in which such practices have been implemented. While many foreign language programs also utilize proficiency standards and oral proficiency assessment for certifying the exit-level abilities of language majors, language educators, and the like, we focus here on the prevalent use of proficiency standards in conjunction with the undergraduate language requirement.

At the University of Minnesota, in the College of Liberal Arts, undergraduate students meet the second language requirement by passing the Graduation Proficiency Test, a set of listening, reading, writing, and speaking assessments based on the ACTFL Proficiency Guidelines (Center for Advanced Research on Language Acquisition, 2003). Proficiency standards, originally set for French, German, and Spanish (Barnes, Klee, & Wakefield, 1990), mandate that students achieve the Intermediate-High level in reading/listening and the Intermediate-Mid level in speaking/writing. These levels are intended to represent abilities at the level equivalent to studying four college semesters.

At California State University Monterey Bay (2003), the University Learning Requirements include the ability "to communicate in a designated language other than English with native speakers of that language." Students may meet the requirement via several routes (course work, testing out, etc.), all of which culminate in assessments (e.g., oral proficiency interviews) that are referenced to the ACTFL Proficiency Guidelines. For languages classified by the Foreign Service Institute as "category 1" (cognate languages for English speakers, such as Spanish and Italian), students are expected to demonstrate Intermediate-Mid proficiency if they entered the university as first-year students or sophomores or Intermediate-Low proficiency if they entered as Junior transfers. Standards for category 4 languages (e.g., Japanese) are one sublevel lower on the ACTFL scale.

At Rice University, the Guidelines for Foreign Language and Literature Teaching Responsibilities (Stokes, 2002) set out the following standards for language competency:

> Specifically, the proficiency goals should be measured in terms of ACTFL standards. That is, lecturers teaching in the Center for the Study of Languages should plan their courses so that the majority of students achieve novice [high] by the

end of 102, intermediate [mid] by the end of 202, and advanced [low] by the end of 302. [Note: these goals may be refined in terms of the four modalities, and in terms of the level of difficulty of the languages concerned.]

Measurement procedures, including oral proficiency assessment based on the ACTFL Guidelines, provide the means for ensuring that students meet these standards.

Finally, at Weber State University (WSU), graduation requirements for the Bachelor of Arts degree include "Documentation of a proficiency level of 'Intermediate Low' or better through an examination administered by the WSU Foreign Language and Literatures Department (2003) or through an examination by a recognized testing agency." In addition, the fact that Weber's foreign language curriculum is based on a nationally recognized standard for language proficiency is listed under the "Points of Pride" of the department, and all courses in the program are gauged according to the ACTFL level at which students are expected to perform (Weber State University, 2002).

In all of these cases, the ACTFL Guidelines proficiency levels and associated assessments clearly serve as the primary means for determining students' completion of language requirements. They also function as measures for ensuring the accountability of the institution and foreign language programs to a set of explicit learning outcomes. In these settings, the ACTFL Guidelines scales have become a *de facto* standards-setting device. It is clear that by prescribing such requirements and implementing their assessment, these programs communicate to students and teachers (as well as to other faculty and administrators at the institution and to the broader public) that a particular type of language learning outcome is prioritized over other possibilities, even over other ACTFL-endorsed learning outcomes in the form of the national standards for foreign language learning (National Standards, 1999). Teaching and testing according to proficiency standards, and in particular teaching and testing for oral proficiency as defined by the ACTFL Guidelines, has no doubt effected positive changes within many foreign language programs (e.g., Barnes, Klee, & Wakefield, 1990). However, it remains to be seen whether this use of proficiency standards and assessments helps foreign language programs achieve the kinds and range of educational goals they actually value.

How Have Proficiency Standards Been Recommended?

Wiggins (1998) defined a true standard as "a worthwhile target irrespective of whether most people can or cannot meet it at the moment" (p. 105), and he suggested that standards rightly used should serve as informative models that enable judgments about the performance of students, teachers, and educational systems vis-à-vis valued principles of education. Clearly, the types of standards that are defined and assessed will largely determine the curriculum and instruction that takes place within educational contexts. As such, standards should pay heed to the range and quality of actual educational outcomes that are valued by learners, teachers, and programs.

There are numerous methods and sources for establishing and assessing standards in education, and these depend largely on the primary purposes of the standard and assessment (Cizek, 2001; Wiggins, 1998). Within language education contexts, standards may be set according to learners' apparent needs, to the demands of employers, or to the expectations of the discipline, and such standards may be put in place to motivate learners, to hold programs and teachers accountable, or to promulgate a particular image of the value and purpose of language education. Of particular recent interest in the United States are standards established in response to perceived national needs for highly proficient language professionals (e.g., Foreign Language Annals, 2002) and standards that are helping to reshape the scope of foreign language education (e.g., National Standards, 1999).

Within collegiate foreign language programs, proficiency standards (of the sort described above) to date have been recommended on the basis of research that suggests what levels of proficiency students should be able to achieve after two years of typical college foreign language instruction as measured according to the ACTFL Guidelines. Wiggins (1998) defined such standards as "benchmarks" that seek to "define a level that we believe a large number of students...not only can but ought to meet if they persist and are well taught" (p. 105). Several studies comparing the oral proficiency levels of students with the number of credit hours or semesters of college instruction have been conducted since the 1980s (e.g., Grosse, 1992; Magnan, 1986; Thompson, 1995, 1996; Tschirner, 1992, 1996), and findings have suggested that, after two years of college language study, ACTFL Intermediate-Low is an appropriate oral proficiency expectation for ostensibly more difficult languages (e.g., Russian for English learners) and Intermediate-Mid is an appropriate expectation for ostensibly easier languages (e.g., German for English learners). These findings have been incorporated into standards for a variety of college foreign language programs, as reflected clearly in the examples above.

More recently, in relation to college study of German, Tschirner and Heilenman (1998) examined the relationship between oral proficiency interview (OPI) ratings and semesters of college study for 20 students at the University of Iowa. They found a median rating of Intermediate-Low (45% of students) after four semesters of study (16 credit/contact semester hours), with roughly equal numbers of learners scoring at Novice-High (25%) and Intermediate-Mid (30%). Based on these findings, they suggested that reasonable expectations for oral proficiency standards in

conjunction with a two-year language requirement be set at the Intermediate-Low level.

It is clear from the available research literature and, more directly, from the types of proficiency standards that have been set within college foreign language programs, that the source for standard setting has been the original two-year foreign language requirement. While proficiency standards and assessments, such as those recommended on the basis of the comparative studies above, have introduced an accountability mechanism into college foreign language study (particularly for students who must pass the proficiency requirement), these standards are nevertheless based on a benchmarking process that has asked what students are typically capable of accomplishing after two years of college language study.

As programs move to embrace this easily implemented language requirement standard, we are moved to ask whether such a benchmark meets Wiggins's definition of a "true standard." Are we enabling judgments about the actual outcomes that reflect what makes foreign language study worthwhile, and what we value as a discipline, by implementing a requirement that students pass an oral proficiency assessment at the Intermediate-Low or Intermediate-Mid ACTFL Guidelines level? If such standards are not warranted as the "true standards" for college foreign language programs, can and should oral proficiency assessment nevertheless play a role within these programs?

Assessing Students' Oral Proficiency in One German Language Program

Clearly, a major objective for any collegiate foreign language program is to foster the development of its students' abilities to communicate orally with the language, and this objective typically applies at the introductory as well as advanced levels of instruction (however these might be defined by individual programs). Assessment of students' oral abilities within foreign language programs, then, may need to be undertaken for a number of reasons, from accountability to achievement to curricular improvement purposes. Such assessment may also use a variety of methods, from classroom performance tasks to curriculum-based measures to global proficiency interviews (Brown, 1996; Brown & Hudson, 1998). What kinds of assessments we utilize depends ultimately on what we hope to accomplish via the assessment process (Norris, 2000). Therefore, administering OPIs and rating students' abilities according to the ACTFL Guidelines, just like any other assessment practice, should proceed on the basis of an explicit understanding of exactly what role this assessment is intended to play within our programs and institutions.

In this section, as an illustration, we discuss the use of oral proficiency assessment based on the ACTFL Guidelines within one collegiate German foreign language program, addressing two questions: (1) at what global oral

proficiency levels are students from all curricular levels of this program rated? and (2) how and why are global oral proficiency ratings used within the program (and how are they not used)? To contextualize answers to these questions, we first provide a brief overview of the Georgetown University German Department (GUGD) program.

Recent curriculum renewal efforts in the GUGD have led to the development of a sequenced undergraduate curriculum, Developing Multiple Literacies (Byrnes, 1998, 2002; Byrnes & Kord, 2001; Pfeiffer, 2002). This curriculum integrates content-oriented instruction from the very beginning (first year, first semester) with an emphasis on language acquisition through the end of the four-year undergraduate sequence of courses. Throughout all years of instruction, the program fosters development in all four modalities (listening, reading, writing, and speaking) by emphasizing a discourse focus that incorporates a wide variety of comprehension and production tasks and engages students in literate language use from the start. A variety of in-house developed materials, instructional tasks, and pedagogical approaches are interwoven with this discourse focus, to maximize both implicit and explicit learning of content and language form.

Ultimately, the Developing Multiple Literacies curriculum was created to better address the background, character, and academic and professional orientations of foreign language learners at Georgetown University. Although not all language students at Georgetown go on to complete four years of language study, those who pursue language study (for however many years) do so with the expectation of acquiring a full range of literate communication abilities that will directly serve their needs (for further details, see Georgetown University German Department, 2003).

Several unique features of this curriculum bear emphasis. First, the curriculum does not prioritize oral proficiency—or global language proficiency of any kind—as its primary instructional objective; rather, the curriculum posits the ability to use German for literate adult communication purposes as the overriding objective for all of its courses. Second, the curriculum does not distinguish between 'lower-' and 'upper-' division courses or between 'language requirement' and other courses. All courses, from the first semester of introductory German through the senior-level courses, are planned and integrated according to the same set of advanced literacy objectives that characterize the program and the needs of the learners. Third, all faculty regularly teach at all levels of the curriculum; there is no distinction between 'language program' and other instructors.

At What Global Oral Proficiency Levels are Students Rated?

In conjunction with a host of other assessment development efforts, and in order to determine students' general speaking abilities across the four years of the Developing

Multiple Literacies curriculum, the GUGD has incorporated assessment of students' oral proficiency levels according to the ACTFL Proficiency Guidelines for speaking (ACTFL, 1986; Breiner-Sanders et al., 2000). The German Speaking Test (GST), one of several audio tape-based SOPIs developed by the Center for Applied Linguistics (CAL) (1996), was selected as a tool that could be easily and simultaneously administered to large numbers of students within a language laboratory setting. The GST elicits spoken performances on a range of narrative, situational, and topical tasks in order to probe examinees' proficiency levels, and the resulting speech samples are rated by trained raters according to the ACTFL Guidelines' rating scales in much the same manner as OPIs (Norris, 1997; Stansfield, 1996). In addition, research has shown consistently that ACTFL Guidelines' ratings based on these SOPIs are substantially comparable to ratings from live OPIs (Clark, 1988; Kenyon & Tschirner, 2000; Shohamy, 1994; Stansfield & Kenyon, 1992, 1993).

In the Spring 2002 semester, the GUGD began administering the GST to all students completing the first 12 hours of instruction (either two intensive semesters or four nonintensive semesters); graduating seniors studying German (at any level) were also given the opportunity to take the GST. In addition, for program evaluation purposes (see discussion below), the GST was administered in academic years 1998/1999 and 1999/2000 during the final weeks of the semester in several undergraduate courses to students whose level of study ranged from 3 semester credit/contact hours (first year, first semester) to 22 semester credit/contact hours (fourth year, first semester on nonintensive track or fourth semester on intensive track) of German instruction. Note that after 22 semester credit/contact hours, students have completed the tightly sequenced course offerings in the Developing Multiple Literacies curriculum.

At the time of this writing, more than 150 GST exams had been collected and scored within the GUGD. Of these, 128 are of primary interest for this article. The remaining exams were administered to students who had just been placed into the program via a curriculum-based placement exam, and whose abilities may not reflect the levels of oral proficiency attributable to students proceeding through the curriculum (see below).

Student performances on Form A of the GST (a non-secure version available for public use) were rated following standard procedures in the GST Rater Training Kit (CAL, 1995) by raters who had been certified on the German Speaking Test rater certification exam (also administered by CAL). Each examinee tape was independently scored by two raters and assigned a holistic level rating according to the ACTFL Proficiency Guidelines. Where the two ratings differed, a third certified rater also scored the tape. Any remaining discrepancies were resolved by discussion among the raters. It is important to point out that raters were not aware of the GUGD curriculum status of the students being assessed. Across all examinees, pairs of raters agreed with each other—within one ACTFL Guidelines sublevel—96% of the time.

Table 1 displays the final ACTFL Guidelines' proficiency ratings for the 128 students whose scores are of interest for the reasons mentioned above. Note that the regular administration of the GST to all students at the end of 12 hours of instruction (equivalent to level 2.2 in the curriculum), which began in Spring 2002, is reflected in the larger number of scores that appear at this level. Not reflected in Table 1, however, are the ratings of German majors graduating from the GUGD, all of whom took substantial course work beyond the sequence of courses ending at curricular level 4.1.

There were several noteworthy patterns in the oral proficiency levels of students observed across the first four sequential levels of the GUGD curriculum. First, virtually all students apparently were capable of reaching the Intermediate-Low level very quickly, by as early as the end of the first semester of nonintensive instruction (3 semester credit/contact hours in the GUGD), despite the fact that the courses did not have an explicit oral proficiency orientation. Second, the central tendency of oral proficiency ratings (represented here by medians) remained constant at the Intermediate-Mid level from the second through fourth nonintensive semesters of instruction (6 to 12 credit/contact hours) and then gradually increased from the fifth (Intermediate-High level) through seventh (Advanced-High level) nonintensive semesters. However, a substantial number of students were rated both below and above the Intermediate-Mid level through the fourth semester of instruction. Indeed, at the end of 12 credit/contact hours, there remained a wide dispersion of ratings, with one student rated as low as Novice-High and another as high as Advanced-High.

The ACTFL Guidelines' oral proficiency ratings, then, did not provide a very consistent indication of the curricular level of a student, given the wide dispersion of ratings. Overall patterns did suggest a quick initial rise to the Intermediate-Low to Intermediate-Mid levels for most students, followed by a leveling off at Intermediate-Mid and then increases during the third and fourth years.

Based on the current data, the Intermediate-Low level is clearly an absolute minimum oral proficiency level for students, *from the end of the first semester of nonintensive instruction*. Nevertheless, many students will surpass this oral proficiency level as early as the end of the first year of nonintensive instruction, with the majority reaching the Intermediate-Mid level after 6 to 12 hours of instruction and 25% or more reaching Intermediate-High or higher by the end of 12 semester credit/contact hours. Clearly, students who study beyond 12 hours will continue to achieve higher oral proficiency levels.

As many as half of GUGD students enroll in intensive

Table 1

ACTFL GUIDELINES ORAL PROFICIENCY RATINGS FOR STUDENTS IN
THE GEORGETOWN UNIVERSITY GERMAN DEPARTMENT

	GU German Curricular Level Completed (corresponds to years of nonintensive [3 semester contact hours] instruction or semesters of intensive instruction [6 semester contact hours])						
ACTFL Level	1.1 (3 hours) n = 7	1.2 (6 hours) n = 26	2.1 (9 hours) n = 19	2.2 (12 hours) n = 39	3.1 (15 hours) n = 20	3.2 (18 hours) n = 12	4.1 (22 hours) n = 5
Novice-High	1			1			
Intermediate-Low	4	7	5	10	2		
Intermediate-Mid	2	15	10	18	4		
Intermediate-High		4	4	7	10	7	
Advanced				2	3	3	
Advanced-High[2]				1	1	2	4
Superior							1
Median rating	IL	IM	IM	IM	IH	IH	AH
Mode rating	IL	IM	IM	IM	IH	IH	AH

[2]Because a number of ratings were completed before the publication of the revised (1999) ACTFL Proficiency Guidelines for Speaking, when the Advanced-Low and Advanced-Mid designations were added to the proficiency scale, all ratings are reported here according to the ACTFL (1986) Proficiency Guidelines.

GU = Georgetown University; IL = Intermediate-Low; IM = Intermediate-Mid; IH = Intermediate-High; AH = Advanced-High

language instruction courses, which cover the identical curriculum offered in two semesters of nonintensive instruction in one semester (6 credit/contact hours per week). We compared the levels of oral proficiency achieved by intensive- and nonintensive-track students. Figure 1 shows the distribution of oral proficiency ratings for students completing 12 hours of instruction in either one (intensive) or two (nonintensive) years. Patterns reveal that students in intensive and nonintensive groups generally achieved the same oral proficiency levels, although a higher percentage of intensive-track students (35%) reached Intermediate-High or above compared with nonintensive-track students (15%).

### How and Why are Global Oral Proficiency Ratings Used in the GUGD?

Given these observations of the relationship between students' oral proficiency ratings and their curricular status, we turn to how such information is used within the GUGD. In conjunction with the initial implementation of the Developing Multiple Literacies curriculum, the specification of desired learning outcomes for each curricular level, and the creation of pedagogy and materials for fostering the achievement of these outcomes, the faculty and administration of the department became aware of the need for a variety of assessment procedures to meet several purposes. Among these, oral proficiency assessment based on the ACTFL Guidelines and employing the GST was undertaken with the following specific set of intended uses, audiences, and consequences in mind.

1. While a number of curriculum-based performance assessments were developed for evaluating the extent to which students are achieving desired learning outcomes in terms of language ability and content knowledge, it was determined that a program-external indicator of students' language abilities would provide a valuable additional perspective for summative and formative evaluation purposes. The ACTFL Guidelines" proficiency ratings from the ongoing administration of the GST thus provide continuous feedback on performance of learners, teachers, the curriculum, and the institutional learning context. As such, they offer an independent "check" on the efforts of the program as a whole, and they are used by the program faculty and administration in revising and improving curriculum and instruction. For example, as Table 1 shows clearly,

Figure 1

ORAL PROFICIENCY RATINGS FOR INTENSIVE AND NONINTENSIVE STUDENTS
COMPLETING 12 SEMESTER CONTACT HOURS OF INSTRUCTION
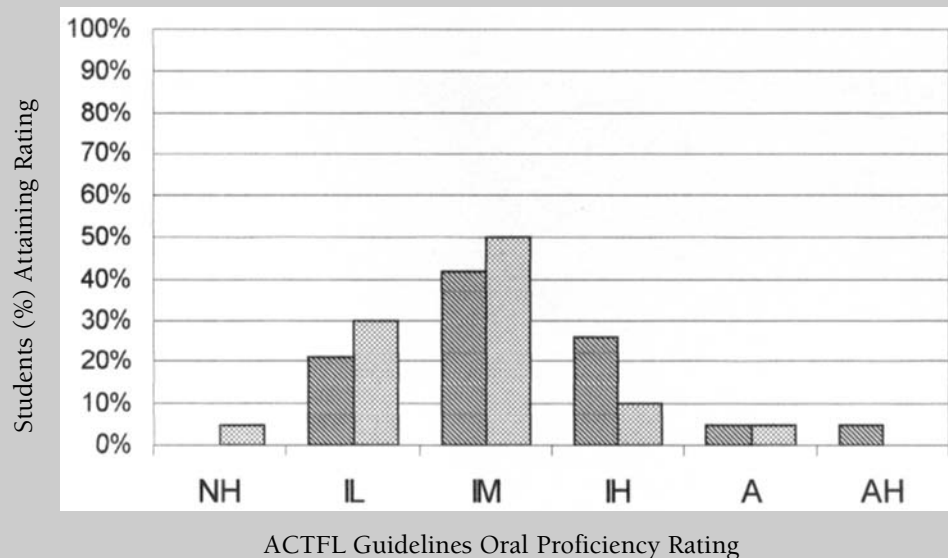
Intensive (N = 19)
Nonintensive (N = 20)

GUGD = Georgetown University German Department; NH = Novice-High; IL = Intermediate-Low; IM = Intermediate-Mid; IH = Intermediate-High; A = Advanced; AH = Advanced High

a substantial proportion of students in the second year of the curriculum did not reach the Intermediate-Mid level or higher, despite the fact that most students did attain this level. In part as a result of this observation, recent year-two curricular revisions have included the development of carefully sequenced instruction on valued spoken genres (e.g., public speech, narration of past events), increased opportunities for oral performance, and a reformulation of the expectations associated with students' speaking abilities at this level.

In addition to such formative uses, students' oral proficiency ratings are also used as a program-independent indicator for summative evaluation purposes in conjunction with periodic program review. In this regard, ratings based on the ACTFL Guidelines prove particularly useful, as they are easily communicated and widely understood by individuals from outside of the GUGD context. Such ratings also provide a measure of comparability with other German programs and with programs in other languages. We should add, however, that ACTFL Guidelines ratings are never conveyed as the sole indicator of program outcomes in the GUGD; rather, they are used as one independent source in conjunction with other internal sources (see

the curriculum Web page (GUGD, 2003) for the full range of assessment instruments and perspectives on learning outcomes and program assessment).

2. The majority of students who enroll in courses in the GUGD complete the equivalent of either two or four-plus years of the sequenced curriculum. Current practice is to administer the GST to all students at these points of exiting the program. Resulting ACTFL Guidelines oral proficiency ratings are provided to students as one indication of the German language abilities they have achieved within the program. These ratings are of direct value to the students for communicating with other academic units within the institution, with other institutions, and with employers, due primarily to their ease of interpretation and widespread recognition as a standard indicator of global oral language proficiency. The ratings also give students an idea of how they will fare according to expectations about language ability beyond the GUGD. In addition, the option of an undergraduate portfolio assessment of learning outcomes (beginning with majors)—one component of which will be GST results from different stages of a student's career—is under current development.

3. Recorded GST performances from across the first four years of the curriculum provide a highly consistent language corpus for investigating features of GUGD students' oral language use (e.g., Norris, 1996, 2001). Because these tape-recorded performances are directly linked to the end of particular levels of instruction within the sequenced curriculum, and because students respond to essentially the same tasks regardless of curricular level, these performances provide an excellent locus for investigating features of language knowledge and ability that are expected to develop in particular ways throughout the program. For example, where the curriculum posits that students will develop the ability to produce increasingly complex and accurate sentences over the first three years of the program, direct measures of syntactic complexity and accuracy may be taken from the existing GST corpus and compared with expected levels of development. Findings are then applied to the ongoing revision of curriculum and instruction.

4. As one component of the validity evaluation of a newly developed curriculum-based placement exam, GST ratings were also used as an external criterion for the abilities of students placed into the upper levels of the curriculum. Thus, in addition to the students described in Table 1, we assessed the oral proficiency of 20 year-3, year-4, and beyond year-4 students at the beginning of the semester of instruction into which they had been placed via the new exam (i.e., they had not yet been exposed to the GUGD curriculum). Students placing into the first semester of instruction at year 3 (i.e., equivalent to those students having completed level 2.2 in the curriculum) were rated Intermediate-Low, Intermediate-Mid, or Intermediate-High; students placing into the second semester of instruction at year 3 were rated Intermediate-Mid, Intermediate-High, or Advanced; students placing into the first semester of the fourth year of instruction were rated Intermediate-High, Advanced, or Advanced-High; and students placing beyond the fourth year (graduate students) were rated Advanced-High or Superior. These patterns of consistently increasing oral proficiency levels in conjunction with consistently higher placements at the upper levels of the curriculum contributed additional evidence to determinations about the effectiveness and accuracy of the placement exam (see Norris, 2004).

How are Global Oral Proficiency Ratings Not Used in the GUGD?

The GST and ACTFL Guidelines' oral proficiency ratings have proved particularly useful for the purposes listed above; however, the GUGD determined that they should not be used for several purposes. The oral proficiency ratings are not used as a proxy or gate-keeping device for stu-

dents to meet any institutional or degree-program language requirements. Whereas a number of foreign language programs in the United States have set proficiency standards in lieu of the typical two-year language requirement in exactly this manner, utilizing an Intermediate-Low or Intermediate-Mid standard as recommended by the available literature, we do not find this use of proficiency ratings warranted within the GUGD program and with GUGD students. First, the majority of students would have little trouble in quickly meeting an Intermediate-Mid oral proficiency requirement by the end of 6 credit/contact hours of instruction. However, in doing so, they would most definitely not be achieving the array of desired learning outcomes associated with the second year of the curriculum, including both oral and written literacy outcomes that constitute the core values of the GUGD undergraduate program. Setting a proficiency standard such as that recommended in the literature would thus dramatically underestimate and delimit the value of the program and our expectations for students.

Second, we find that establishing a proficiency requirement that must be met by students removes the responsibility for effective and useful instruction from the GUGD program and faculty and places the onus for learning directly on the shoulders of the individual learner. Adopting such a deficit approach to student learning could lead to negative washback effects for students, teachers, and the program, rather than the positive uses of assessment in the GUGD as formative for teachers and evaluative of the curriculum. Such a requirement would also direct learners to attend only to those portions of the curriculum that will help them "get through the language requirement.'" Within the GUGD, it is preferred that learners attend to all aspects of the curriculum that are designed to develop their abilities as literate users of the German language.

The ACTFL Guidelines' proficiency ratings are also not used as standards for GUGD undergraduate program outcomes expectations. First, within a focus on oral language abilities, OPI or SOPI assessments would not provide a good indicator of the full range of German speaking tasks that the GUGD prioritizes, which are frequently integrated with other modalities and always based on communication with content that is meaningful to the individual students and relevant to the curricular topics/themes. Prioritized tasks within the curriculum are also generally interactive group-level tasks and call upon competencies beyond those reflected in OPI or SOPI assessments. Task-based performance assessments (Byrnes, 2002; Norris et al., 1998) rather than global proficiency assessments, prove much more useful in this regard as operationalizations of desired learning outcomes.

Second, setting a proficiency standard of Intermediate-Low or Intermediate-Mid communicates the wrong message about the value and scope of foreign language learning

within the GUGD, not only to students, but also to our colleagues in other degree programs and the university administration. To embrace such a standard as the sole or even a major indicator of learning outcomes is to short-change students, teachers, and the foreign language discipline as part of the humanities. Such assessment reinforces inappropriate notions of the nonacademic nature of foreign language study at the college level and thus undermines its standing within the higher education academic community.

Conclusion and Implications

As the demand for enhanced accountability of all participants in higher education intensifies, the development of "true standards" in the sense proposed by Wiggins poses an increasingly urgent task for foreign language departments. In this article, we have provided a quick overview of current trends to utilize ACTFL Proficiency Guidelines (ACTFL, 1986; Breiner-Sanders et al., 1999) for setting standards and especially for identifying learning outcomes. Against that backdrop and on the basis of an analysis and interpretation of SOPI data collected from GUGD students, we have laid out several reasons for not following this model. Instead, we argue for developing curriculum-dependent standards and the use of a range of assessment instruments that reflect the kinds of learning valued and fostered by the program.

At the same time, we argue that curriculum-independent assessment tools such as the GST and other oral proficiency interviews rated according to the ACTFL Guidelines can offer a useful balance to such program-specific standards, by providing a measure of comparability and outside accountability to the program. In this balancing act, such standards maintain the need to continue and strengthen a diversity and distinctiveness of institutional educational purposes in this country while being cognizant of and appropriately responsive to the pressure towards more universal standards of learning and its assessment.

We also hope to have made clear that setting educational outcomes standards based solely on the ACTFL Guidelines and oral proficiency scales overly constrains the outlook of the educational purposes and goals of collegiate foreign language departments. We therefore argue for the careful incorporation of curriculum-independent measures of this sort to meet clearly defined and delimited roles within program assessment.

As we have suggested, the findings and interpretations provided in this article are part of ongoing assessment activities at the GUGD. As such, we will continue to collect cross-sectional and longitudinal data for the sort of oral proficiency measured by the GST and the ACTFL Guidelines. Together with other areas of investigation, such as the regular assessment of development in second language writing complexity and accuracy, we will enable further understanding of how aspects of the Developing

Multiple Literacies curriculum function, how students best learn within its learning context, and how to adjust pedagogical approaches accordingly. Such data promise to provide a rich context for future research into the development of oral proficiency among instructed students of German.

References

American Council on the Teaching of Foreign Languages. (1986). ACTFL proficiency guidelines. Yonkers, NY: ACTFL.

Barnes, B., Klee, C., & Wakefield, R. (1990). A funny thing happened on the way to the language requirement. ADFL Bulletin, 22(1), 35–39.

Breiner-Sanders, K., Lowe, Jr., P., Miles, J., & Swender, E. (2000). ACTFL proficiency guidelines—speaking (revised, 1999). *Foreign Language Annals, 33*, 13–18.

Brod, R., & Huber, B. J. (1996). The MLA survey of foreign language entrance and degree requirements, 1994-95. ADFL Bulletin, *28*(1), 36–43.

Brown, J. D. (1996). Testing in language programs. Upper Saddle River, NJ: Prentice Hall Regents.

Brown, J. D., & Hudson, T. D. (1998). The alternatives in language assessment: Advantages and disadvantages. TESOL Quarterly, *32*, 653–75.

Byrnes, H. (1998). Constructing curricula in collegiate foreign language departments. In H. Byrnes (ed.), Learning foreign and second languages: Perspectives in research and scholarship (pp. 262–95). New York: The Modern Language Association.

Byrnes, H. (2002). The role of task and task-based assessment in a content-oriented collegiate foreign language curriculum. Language Testing, *19*, 419–37.

Byrnes, H. & Kord, S. (2001) Developing literacy and literary competence: Challenges for FL departments. In V. Scott & H. Tucker (eds.), SLA and the Literature Classroom: Fostering Dialogues (pp. 31–69). Boston: Heinle & Heinle.

California State University Monterey Bay. (2003). University learning requirements. Retrieved April 01, 2003 from http://csumb.edu/academic/ulr/ulr/language.html.

Center for Advanced Research on Language Acquisition. (2003). Minnesota language proficiency assessments. Retrieved July 24, 2003 from http://carla.acad.umn.edu/MLPA.html.

Center for Applied Linguistics. (1995). German speaking test rater training kit. Washington, DC: Author.

Center for Applied Linguistics. (1996). German speaking test. Washington, DC: Author.

Cizek, G. (ed.) (2001). Setting performance standards: Concepts, methods, and perspectives. Mahwah, NJ: Lawrence Erlbaum.

Clark, J. L. D. (1988). Validation of a tape-mediated ACTFL/ILR-scale based test of Chinese speaking proficiency. Language Testing, 5, 197–205.

Erwin, T., & Wise, S. (2001). Standard setting. New Directions for Institutional Research, *110*, 55–64.

Fleck, J. (1970). The foreign language requirement: A panel

discussion, 1969 AATG Meeting, New Orleans. Die Unterrichtspraxis, 3(2), 101–4.

Foreign Language Annals. (2002). Interview with Robert Slater, Director, National Security Education Program (NSEP). Foreign Language Annals, 35(2), 258.

Freed, B. F. (1981). Establishing proficiency-based language requirements. ADFL Bulletin, 13(2), 6–12.

Freed, B. F. (1984). Proficiency in context: The Pennsylvania experience. In S.J. Savignon & M.S. Berns (eds.), Initiatives in communicative language teaching: A book of readings (pp. 211–40). Reading, MA: Addison-Wesley.

Freed, B. F. (1987). Preliminary impressions of the effects of a proficiency-based language requirement. Foreign Language Annals, 20, 139–46.

Georgetown University German Department. (2003) Developing multiple literacies. Retrieved April 01, 2003 from http://www.georgetown.edu/departments/german/programs/curriculum/.

Grosse, C. U. (1992, November). Testing the limits of proficiency: The ACTFL OPI and FL departments. Paper presented at the 26th Annual Meeting of the American Council on the Teaching of Foreign Languages, Rosemont, IL.

Huber, B. J. (1992). Characteristics of foreign language requirements at U.S.colleges and universities: Findings from the MLA's 1987-89 survey of foreign language programs. ADFL Bulletin, 24(1), 8–16.

Kenyon, D., & Tschirner, E. (2000). The rating of direct and semi-direct oral proficiency interviews: Comparing performance at lower proficiency levels. Modern Language Journal, 84, 85–101.

Klayman, N. (1978). Views on the foreign language requirement in higher education. Modern Language Journal, 62(5–6), 235–38.

Lange, D. (1990). Assessing language proficiency for credit in higher education. ERIC Digest, ED321588.

Magnan, S. (1986). Assessing speaking proficiency in the undergraduate curriculum: Data from French. Foreign Language Annals, 19, 429–38.

National Standards in Foreign Language Education Project. (1999). Standards for foreign language learning in the 21st century. Yonkers, NY: Author.

Norris, J. M. (1996). A validation study of the ACTFL Guidelines and the German Speaking Test. Unpublished master's thesis, University of Hawai'i at Manoa, Honolulu.

Norris, J. M. (1997). The German Speaking Test: Utility and caveats. Die Unterrichtspraxis, 30, 148–58.

Norris, J. M. (2000). Purposeful language assessment. English Teaching Forum, 38(1), 18–23.

Norris, J. M. (2001). Use of address terms on the German Speaking Test. In K. Rose & G. Kasper (eds.), Pragmatics in language teaching (pp. 248–82). New York: Cambridge University Press.

Norris, J. M. (2004). Validity evaluation in foreign language assessment. Unpublished doctoral dissertation. University of Hawai'i at Manoa, Honolulu.

Norris, J. M., Brown, J. D., Hudson, T. D., & Yoshioka, J. K. (1998). Developing second language performance assessment. Honolulu: University of Hawai'i Press.

Pfeiffer, P. (2002). Preparing graduate students to teach language and literature in a foreign language department. ADFL Bulletin, 34, 11–14.

Schulz, R.A. (1988). Proficiency-based foreign language requirements: A plan for action. ADFL Bulletin, 19(2), 24–28.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. Language Testing, 11, 99–123.

Stansfield, C. (1996). SOPI test development handbook. Washington, DC: Center for Applied Linguistics.

Stansfield, C., & Kenyon, D. (1992). Research on the comparability of the oral proficiency interview and the simulated oral proficiency interview. System, 20, 347–64.

Stansfield, C., & Kenyon, D. (1993). Development and validity of the Hausa Speaking Test with the ACTFL Proficiency Guidelines. Issues in Applied Linguistics, 4, 5–31.

Stokes, G. (2002). Guidelines for foreign language and literature teaching responsibilities. Retrieved April 01, 2003 from http://cohesion.rice.edu/humanities/depts.cfm?doc_id=1426.

Thompson, I. (1995). A study of interrater reliability of the ACTFL Proficiency Interview in five European languages: Data from English, French, German, Russian, and Spanish. Foreign Language Annals, 28, 407–22.

Thompson, I. (1996). Assessing foreign language skills: Data from Russian. Modern Language Journal, 80, 47–65.

Tschirner, E. (1992). Oral proficiency base lines for first and second year college German. Unterrichtspraxis, 25, 10–14.

Tschirner, E. (1996). Scope and sequence: Rethinking beginning foreign language instruction. Modern Language Journal, 80, 1–14.

Tschirner, E., & Heilenman, L. (1998). Reasonable expectations: Oral proficiency goals for intermediate-level students of German. Modern Language Journal, 82, 147–58.

Voorhees, R. (2001). Competency-based learning models: A necessary future. New Directions for Institutional Research, 110, 5–13.

Weber State University (2002). Catalog. Retrieved April 01, 2003 from http://documents.weber.edu/catalog/0203/~forlngb.htm.

Weber State University Foreign Language and Literatures Department (2003). Foreign language requirements for the bachelors of arts degree. Retrieved April 01, 2003 from http://documents.weber.edu/catalog/0203/~forlngd.htm.

Wiggins, G. (1998). Educative assessment: Designing assessments to inform and improve student performance. San Francisco: Jossey-Bass.

Wolfe, W. J. (1959). Foreign language entrance and degree requirements for the B.S. degree. Publications of the Modern Language Association, 74(4 pt. 2), 34–44.