

Assessing the Oral Proficiency of Adult Learners, “Heritage” and “Native” Speakers Using the *ILR Descriptions* and *ACTFL Proficiency Guidelines*: Considering the Challenges¹

Cynthia L. Martin

The *ILR Descriptions – Speaking* and the *ACTFL Proficiency Guidelines – Speaking* trace their genesis to an oral proficiency assessment system first developed at the Foreign Service Institute in the 1950s to assess “foreign language users’ abilities according to a consistent scale.”² The original purpose was clearly aimed at assessing the abilities of non-native, non-heritage adult foreign language learners, and this genesis is still reflected in the current *ILR Descriptions* and the *ACTFL Guidelines*. However, tools based on these descriptions/guidelines are increasingly being used to assess the abilities of language users who are not what we would consider to be “learners” of the “foreign” language being tested, but rather, who have commonly been described as “native” or “heritage.”³ Because

¹ I offer comments in this essay as an individual, drawing upon my experience as an ACTFL and ACTFL/ILR OPI tester over the past 25 years, primarily in Russian (and in English for nearly a decade), as well as in training and mentoring testers in a wide variety of languages (80+) to use both ACTFL and ILR assessments. The opinions expressed here are entirely my own, and I do not intend to suggest endorsement by any organization, agency or company. The ILR Descriptors may be found at <http://www.govtilr.org>; the ACTFL Guidelines at <http://www.actfl.org>.

² Lowe and Stansfield (eds), 1998 *The unassimilated History*, in *Second Language Proficiency: Current Issues* (pp. 11-51); Englewood Cliffs, NJ, Prentice-Hall Regents, p. 15. For a brief history, see also Herzog, M. “An Overview of the History of the ILR Language Proficiency Skill Level Descriptions and Scale,” published on the official www.govtilr.org website: <http://www.govtilr.org/skills/index.htm>.

³ The last 12 years or so have seen a great deal of work in the US on the concept of “heritage” speakers, primarily with a focus toward learning, but less so toward systematic evaluation of a speaker’s functional ability, especially for the professional workplace. For overviews of the notion of “heritage” language, primarily in the US, the reader is directed to the following two volumes:

‘300-plus years of heritage language education in the United States,’ in J.K. Peyton, D.A. Ranard, and S. McGinnis (eds.). *Heritage Languages in America: Preserving a National Resource*, Center for Applied Linguistics, Washington, DC, and Delta Systems, McHenry, IL, 81- 98, 2001. See in particular, contributions by J. Fishman, G. Valdes, T. Wiley.)

Encyclopedia of Language and Education, Part 4. N. Hornberger (ed). See in particular: O. Kagan, K. Dillon. “Issues in Heritage Language Learning in the United States,” 143-156, Springer, 2008.

Very little research has been done to date on the use of ACTFL/ILR assessment tools to assess the non-foreign-language learner. As stated in the essay, Olga Kagan at UCLA with the help of ACTFL is currently conducting such research on assessing heritage profiles in Russian and Spanish using the ACTFL OPI.

both sets of guidelines, ILR and ACTFL, focus on the speaker’s *functional ability to perform certain tasks*, there is no reason why assessment tools based on them should not be used to assess the proficiency of *any* language user, regardless of his or her profile and of how or where the language being assessed was learned or acquired. This focus on function, I would argue, is responsible in large part for the resiliency and flexibility of these guidelines, allowing them to endure for decades.

It is not the purpose of this essay to offer new theoretical conceptualizations or definitions of these terms, but rather to discuss some typical *profiles* of ACTFL and ILR OPI candidates who are clearly not traditional foreign language learners of the languages being assessed; to summarize current issues being faced in the ILR/ACTFL OPI testing community; and to encourage fruitful discussions about future directions. I shall discuss the following questions currently facing the ACTFL/ILR testing community:

1. What kinds of speakers who are NOT foreign language learners of the language being assessed are currently encountered in the context of ACTFL/ILR OPI testing?
2. What do the guidelines say about “native” speakers?
3. How do the profiles of speakers achieving the same overall ratings differ in terms of the rating criteria depending on whether they are learners, or what we might refer to as “native,” or “heritage speakers”?
4. Should there be a different set of assessments or criteria for “native” speakers, especially at the higher levels?
5. Might a given speaker’s profile matter to the end-user of the ratings? If so, what mechanisms are currently used and/or may be developed to provide this information?

1. What kinds of speakers who are NOT *foreign language learners* of the language being assessed are currently encountered in the context of ACTFL/ILR OPI testing?

The definition or categorization of different types of speakers has become increasingly complex in recent years, and as stated above, I do not presume here to offer new definitions. Sweeping generalizations, of course, can be quite dangerous and misleading, yet currently, there seem to be three distinct *profiles* of adult speakers/test-takers who are being assessed using ACTFL/ILR oral proficiency scales 1) adult foreign language learners; 2) what I will refer to as “heritage” speakers and 3) what I will refer to as “native” speakers. Since the profiles of speakers who fall into the two latter categories differ significantly

from the profile of a “typical” American adult foreign language learner for whom the scales were originally developed, the application of either the ILR or ACTFL scales *in practice* has raised a number of interesting issues for the ACTFL/ILR testing community. In the subsequent discussion of questions no. 2-5 below, I shall use the following “working definitions” of these three diverse profiles.

a. A major segment of the test-taking population continues to be composed of traditional adult learners of the foreign language being tested (i.e., through explicit study outside the target culture, such as at a university or in a special language school, such as the Foreign Service Institute or Defense Language Institute here in the U.S.). The majority, but not all, of these candidates have English as their “native” language. It was for this kind of speaker that both scales were originally designed, so applying the assessment criteria to these profiles is generally straightforward.

b. A second, growing group of test-takers is composed of those I will call “native” speakers who acquired the language as a first language while living inside the target culture, and who completed most or even all of their formal education (usually at least through adolescence, or 14-15 years old) in the target culture in institutions where the primary language of *all* instruction was the target language being tested. For many languages (but not all), this means the speaker is fully literate in the language and has reasonably high proficiency in all the skills: speaking, reading, listening, and writing. Many of these speakers grew up in bilingual or multi-lingual environments, but usually consider the testing language to be, if not literally their first or “mother” tongue, then at least their strongest and most sophisticated language.

The term “native speaker” has become increasingly problematic as the diversity of linguistic profiles encountered in the ACTFL/ILR testing community grows. I offer here two very important cautions: 1) we must not conflate “native” speaker with ethnicity; 2) we cannot hold the “native” speakers to one uniform standard of pronunciation. Take, for example, the case of English. Speakers growing up in America, Great Britain, Australia, New Zealand, and even in countries such as India, may have vastly different profiles and certainly different pronunciations, though all may consider themselves to be “native” speakers of English. Consider also the case of speakers who grew up in the former Soviet republics, such as residents of Central Asia or the Caucasus region. Regardless of the actual ethnic origins of speakers from these regions, they essentially grew up in a bilingual environment. For many, regional languages may have been spoken in the home, but Russian (not Kazakh, Uzbek, Tadzhik, Georgian, etc.) was the

primary language of education, as well as of professional and public life. These speakers may have regional pronunciations that differ from that of speakers from Moscow, but many may consider their “native” language to be Russian (and their regional languages to be “heritage”) and would certainly score higher on our proficiency scales in Russian than in their “home” languages. Despite the presence of regional accents, would not these speakers be considered “native” speakers of Russian? These are just a few examples, but we could cite many others of the difficulty of linking a “normative pronunciation” to the definition of “native,” as in the case, for example, of the vast diversity of “native” speakers of Spanish, Arabic, Portuguese, and the list goes on. The sheer number of variants among those speakers whom we might consider “native” means that pronunciation alone cannot be the ultimate criterion by which someone’s linguistic identity, especially in terms of “native,” is determined.

c. “Heritage” speakers are generally considered to be those who grew up outside the target culture (or left at a very young age, usually before becoming fully literate in the language) in an environment where the language was spoken in the home, by friends and relatives, but not in the broader surrounding community. Generally, speakers with this profile who are being encountered in the ACTFL/ILR testing context differ from the first group, in that they did not receive their formal education inside the target culture or in the target language, and themselves often consider their “heritage” language to be a second language, or at least not as strong as their “main” language, or language of the society and educational system where they primarily grew up and received formal education.

I wish to reiterate here that the “native” and “heritage” profiles as described above are narrow, and I am using them here exclusively in the context of common profiles currently encountered within the ACTFL/ILR testing context.

2. What do the guidelines say about “native” speakers?

ACTFL Proficiency Guidelines – Speaking

Throughout the *ACTFL Guidelines - Speaking*, references to “native” speakers are limited to comments about the “native” speaker as the receiver of the message. The *ACTFL Guidelines* divide “native” speakers into those “accustomed to dealing with non-natives” and those “unaccustomed to dealing with non-natives.” As criteria-referenced guidelines, the ACTFL descriptions are very careful *not* to have as a criterion any comparison whatsoever of a speaker’s *performance* or features required to perform certain functions with an idealized or normative performance of an imagined “native” speaker. Nowhere in the ACTFL

guidelines are there suggestions that a speaker must perform a given function “the way a native speaker would” in order to be communicatively successful; quite the contrary, in fact, and the ACTFL guidelines clearly avoid drawing such parallels, focusing instead on how well a given speaker is able to accomplish a linguistic task using the language inside the target culture, communicating with native speakers. Of course, as we move up the ACTFL scale, the kinds of functions required necessarily mean that the speaker is likely increasing his/her ability to function in a target-language environment in more linguistically and culturally appropriate ways, employing the linguistic and cultural norms of that language and society.

ILR Skill Level Descriptions Speaking

In *ILR Speaking Proficiency Descriptions* for levels 0+ through 3, references to native speakers refer either to *how the native speaker would understand* the person being assessed, or *how the native speaker might have to adjust his or her own speech to be understood* by the speaker. It is only in the descriptions for 3+ (*General Professional Proficiency Plus*) and above that we find references comparing a speaker’s *performance* or *linguistic features* to that of a “native” speaker:

ILR 3+ says that the speaker “has discourse competence in a wide range of contexts and tasks, often matching a native speaker’s strategic and organizational abilities and expectations.”

ILR 4: “Organizes discourse well, employing functional rhetorical speech devices, native cultural references, and understanding. Language ability only rarely hinders him/her in performing any task requiring language; yet, the individual would seldom be perceived as a native.”

ILR 4+: “Speaking proficiency is regularly superior in all respects, usually equivalent to that of a well-educated, highly articulate native speaker... However, the individual would not necessarily be perceived as culturally native.”

ILR 5: “Speaking proficiency is functionally equivalent to a highly articulate, well-educated native speaker and reflects the cultural standards where the language is natively spoken. The individual uses the language with complete flexibility and intuition, so that the speech on all levels is fully accepted by well-educated native speakers in all of its features, including breath of vocabulary and idiom, colloquialisms, and pertinent cultural references. Pronunciation is typically consistent with that of well-educated native speakers of a non-stigmatized dialect.”

Hence, at the high end of the ILR scale, the descriptions do contain comparisons between the speech of a test-taker and a “native” speaker, raising the kinds of questions already mentioned about how we determine the standards for “native” performance. In question no. 3 below, I shall discuss the concern raised by how these descriptions are perceived by test-takers who self-identify as “native” speakers.

3. How do the profiles of speakers achieving the same overall ratings differ in terms of the rating criteria depending on whether they are learners, or what we might refer to as “native,” or “heritage speakers”?

Since both sets of oral proficiency guidelines are primarily focused on *function*, it is obviously possible for learners, native or heritage speakers to receive the same global rating while commanding very different skills and having very diverse profiles. Empirical linguistic analysis is required to confirm these observations and others, and there are some investigations into various aspects of different speech profiles currently underway at the Center for the Advanced Study of Language at the University of Maryland, for example, or a study on heritage speakers of Spanish and Russian currently being conducted by UCLA Heritage Center under the direction of Olga Kagan with the participation of ACTFL that may shed light on these differences (see footnote 3). My intention here is to highlight some of the most common differences that have been observed in some testing languages among adult foreign language learners, “heritage,” and “native” speakers in terms of the four general broad categories that form the core of the assessment criteria for the oral interview protocols currently in widespread use: *Global Tasks and Functions*, *Content/Context*, *Comprehensibility* (including *Accuracy*), *Text Type*.

Global Tasks and Functions

The key component in this assessment category is the ability to *accomplish a given task* linguistically. All speakers, regardless of profile, who receive the same rating have one thing in common: they can all perform the required tasks for the level assigned and cannot sustain performance of the required tasks at the next higher levels. The other rating criteria serve to support the functions of a given level. Hence, for this particular criterion, the same standard must be met by all speakers achieving a given rating, i.e. they must be able to sustain performance of the functions associated with that level. *How* they do so may differ dramatically from speaker to speaker. These qualitative differences are manifested in the other rating criteria categories.

All speakers achieving an ACTFL Novice or ILR 0+ must demonstrate the ability to produce memorized, rehearsed language in highly contextualized situations, and can often produce lists of related items. ACTFL Intermediate and ILR 1/1+ ratings are awarded to speakers who can demonstrate the ability to create with the language in order to engage in simple conversation, handle a simple social transaction, and ask questions.

Speakers at ACTFL Advanced and ILR 2/2+ may be learners, heritage or native speakers, but the required tasks, such as narrations, descriptions, handling situations in everyday social transactions, must be met by all. For learners and “heritage” speakers alike, there may be emerging awareness of the culturally appropriate ways to perform various functions, but the required tasks can clearly be accomplished successfully in ways that are not entirely “native” or even “native-like.” Native speakers achieving these ratings are most likely able to perform the Advanced/L2 functions required in natural, native or native-like ways across all the tasks, though as we have established, *this is not a criterion specified either in the ACTFL or ILR scales at these levels.*

The highest rating on the ACTFL scale is Superior, but as in the case of Advanced-level ratings, this does not mean that there is a convergence of linguistic profiles at this level; rather a non-native learner, a heritage speaker, and a native speaker who are all rated Superior may have very different profiles, indeed, while fully meeting the criteria to be awarded a Superior rating. The same may be said of the ILR ratings of 3/3+. The functions that a speaker must demonstrate to achieve ratings of ACTFL Superior or ILR 3/3+ are: discussing an issue at the abstract level, supporting opinion, hypothesizing, and handling “linguistically unfamiliar situations” (usually indicated by the ability to circumlocute around low-frequency/unrehearsed material).

All three groups of speakers under discussion here are expected to command the language to such a degree that culturally appropriate ways of performing the required tasks are now fully operational. This does not mean that the learner, heritage, and native profiles converge entirely, but all three types of speakers demonstrate that they can be *equally effective* in performing the required functions successfully in the target culture. Native speakers receiving these ratings obviously perform the functions in a native or native-like manner. If the speaker is a learner or a heritage speaker, however, the way in which he/she performs these functions may range from culturally *acceptable* (i.e., effective and would not disturb the native interlocutors) to native or native-like. This will depend in large part on the speaker’s experiential contact with the target culture. For some learners, this experience may or may not be greater than it is for a

heritage speaker; in other words, learners may, in fact, perform these functions in culturally more authentic ways than heritage speakers, depending on the level of exposure to the target culture and its norms. It is not unusual for heritage speakers with limited experience inside the target culture to perform the functions in ways deemed successful, but that are clearly influenced by his/her dominant, non-target culture language and environment. In short, a single profile is impossible to establish at this level for natives, learners, or heritage speakers.

To be rated at the very top of the ILR scale (4/4+/5), in addition to being able to discuss issues abstractly, support opinions, and hypothesize – all in highly sophisticated ways – the speaker must also perform the function of spontaneously tailoring the language in two different contexts: formally (often in a presentational mode) and informally (usually in a familiar, interpersonal mode). This function is generally elicited using role plays. Lack of ability to demonstrate this function is probably the single greatest obstacle to attaining levels 4/5 for all speakers.

One interesting observation has emerged: most native speakers do not have difficulty with informal role plays, but many *do* have difficulty producing spontaneous formal tailoring. On the other hand, some high-level learners who have used the language of the test in professional capacities that require formal tailoring, instead have greater difficulty tailoring language informally. In both instances, performance of these functions depends in large part on a speaker’s experience using the language in the target culture in situations requiring tailoring. Given that the assessment is non-compensatory and all criteria must be met in order to achieve a particular rating, failure to tailor *either* formally or informally will result in a maximum rating of ILR 3+ for all three types of speakers, regardless of profile.

Attempting to identify indicators that the speaker has a broad knowledge of and ability to use appropriately “native cultural references” (ILR 4/4+, ILR 5 uses “cultural references”) has also become more difficult in an age where access to education has been increasingly democratized and the content of most educational systems continues to diversify, not to mention how to account for differences based on factors such as age, region, background, etc. Defining and then designing assessments that could “test” for such things is extraordinarily challenging in a world where “native” speakers are becoming less and less defined by their belonging to a homogeneous group of speakers who share the same cultural, linguistic and even physical or territorial reality. Currently, the best that we can do when testing at the very top of the ILR scale is to look for

spontaneous production of speech that shows the speaker's ability to use cultural references in rich and robust ways, rather than attempt to elicit production or evidence of "knowledge" of specific references.

Content/Context

This broad criterion is applied uniformly for all three categories of speakers. Speakers who receive ratings of ACTFL Novice/ILR 0+ are limited essentially to basic courtesy requirements and immediate survival needs that have likely been rehearsed; ACTFL Intermediate/ILR 1/1+ (typically learners or low-level heritage speakers with very little actual experience inside the target culture) are limited in content to topics related to autobiographical information and everyday routines, and contexts that are restricted to everyday interpersonal interactions related to these familiar content areas. The content and context areas manifested in the performance of speakers achieving ACTFL Advanced/ILR 2/2+ increase beyond the familiar and personal to include more high frequency concrete topics such as work and current events. At ACTFL Superior/ILR 3/3+, this criteria expands to include most practical, social and professional (not highly specialized) topics, as well as settings where these topics might be discussed. At ILR 4/4+/5, the speaker is expected to operate effectively in virtually all (4/4+) and all (5) settings and contexts that may be encountered in the target culture.

Comprehensibility/Accuracy

This category includes linguistic features such as delivery – pronunciation, intonational patterns, tones for tonal languages, pace or "fluency," as well structural control including grammar and syntax. These features are assessed in two primary ways: 1) *in relation to the level of accuracy required to perform the function*: are these linguistic features adequate to perform successfully the communicative task at the intended level? and 2) *in relation to what kind of listener would be able to understand the message*: very sympathetic, accustomed to dealing with non-natives, or a neutral, non-sympathetic listener? It is only at the highest level of the ILR scale (ILR 5) that pronunciation is explicitly mentioned. The ILR descriptions of 4 and 4+ say that the speaker, despite high-level language use, would "seldom be perceived as culturally native" (ILR 4) or "not necessarily be perceived as native" (ILR 4+), which are often interpreted as including marked foreign pronunciation, among other things. We discussed the complexity of defining "native" pronunciation in no. 1 above, and I shall address the consternation caused in test-takers that self-identify as "native" in question no. 3 below.

Text Type

At ACTFL Novice/ILR 0+, the text type is essentially the same for all speakers: isolated words and phrases and perhaps some sentences or questions that can be used only in highly contextualized everyday interactions, such as greetings and social courtesies. For ACTFL Intermediate and ILR 1/1+, the primary text type consists of simple sentences and strings of sentences. In order to be rated ACTFL Advanced or ILR2/2+, the speaker must control paragraph discourse in order to perform the functions (narrations, descriptions, for example); often heritage and certainly native speakers rated at this level have more consistent control over this text type than many learners. At ACTFL Superior/ILR 3-5, extended discourse must be controlled in order to perform the functions across a range of topics and tasks.

In terms of text-type/structural control, I would like to offer the following observation. For learners and often for heritage speakers, syntactic patterns—even at high levels of proficiency in the language being tested—tend to be influenced by the dominant language of the speaker. In significant testing of native speakers of English who have also achieved high-levels of proficiency Russian, for example, those who score at Advanced-Mid to Superior in ACTFL and 2+ or above on the ILR scale, dominant English syntactic patterns often stubbornly remain reflected in spoken Russian. Many other features, such as lexical control, grammatical accuracy, content/context domains, socio-cultural appropriateness of language use, and even pronunciation and intonation may approach native-like Russian usage, but target-language syntax appears to be among the most difficult thing for adult foreign language learners to acquire.

For example, English syntactic structure is quite different than Russian, and because of morphology, Russian word order is generally thought of as much more “flexible” than English. Furthermore, impersonal, “subjectless” sentences are quite common in Russian, something that English structure does not tolerate. I must repeat here that this observation is not based on empirical studies, but rather, on experience over nearly 25 years of testing, teaching, and interacting with thousands of learners of Russian as well as “native” speakers, I have come to think that this may be the most difficult part of the language to “acquire” and fully internalize. More recent experience with heritage speakers whose dominant language and language of formal education is English, suggests that the same may be true for many of them, as well.

In other words, learners and heritage speakers who have received little or no formal education in Russian and who consider English to be their “stronger” or “primary” language, often share this feature. Certainly, empirical linguistic

analysis is required to confirm these observations, and perhaps the abovementioned study being conducted by UCLA's National Heritage Language Resource Center with ACTFL participation (looking at ACTFL-rated heritage speakers of Spanish and Russian) may shed light on this phenomenon. Interestingly, the one most common thing that learners and heritage speakers share is the lack of formal education in the target culture; in other words, speakers of these two profiles become fully literate in English while being educated formally in a primarily English-speaking environment. I would expect interesting findings to emerge from a study examining the relationship between the syntactic structure of the written language of a speaker's formal education and the syntactic patterns of the target language. Is it possible that in this regard, the syntactic patterns of one's strongest language (in which one has received formal education, especially post-elementary school and in which one has the highest degree of literacy) has a greater influence on one's acquisition of target-language syntactic patterns of oral speech than when and where and how the language was learned or acquired?

4. Should there be a different set of assessments or criteria for "native" speakers, especially at the higher levels?

This question continues to be discussed in the ACTFL/ILR testing community, but so far, there is not a significant consensus or even strong sentiment arguing for a new set of assessments designed specifically for "native" speakers. However, there has been some discussion of designing a separate assessment just for the upper ends of the ILR scale that would distinguish between native and non-native performance only in specific cases where there would be a need to distinguish between these two profiles. I suggest that if the focus of proficiency assessments is to remain the *functional ability of a speaker to perform certain communicative tasks*, then there is no reason to develop a separate set of assessments. An entirely new set of guidelines and assessment tools would need to be designed if our goal were to be the ability to assess how well a given speaker's performance *matches* that of a "native" speaker. Such an assessment would have to compare a speaker's performance with some idealized "native" standard, which begs the questions: How would we define "native" speaker for these purposes? On what basis would we attempt to characterize an "ideal" native-speaker performance of specific tasks? As discussed earlier, the complexities involved in defining "native" are increasing in the 21st century, not diminishing. Hence, it seems that designing *valid* and *reliable* assessments that would assess whether a speaker is a "native" or how closely s/he approximates

one, would prove virtually impossible, since setting standards would be impossible.

Since the vast majority of organizations seeking individuals with language proficiency for the workplace usually look for ratings in the ACTFL Advanced-Superior ranges and ILR 2-3+ ranges, the need for such an instrument does not appear to be urgent. However, a reasonable suggestion continues to be discussed that, for cases in which a rating of ILR 4/4+/5 is desirable or required for a specific purpose, it may be both prudent and possible to design a separate assessment for *any* speaker who has already scored an ACTFL Superior or ILR 3 in order to determine whether s/he can perform the *functions* of 4/5 as described by the ILR Descriptions. In practical terms, the design and administration of such a test would resemble more the official *Test of Russian Language* (administered in Russia), or the tests administered by the University of Cambridge ESOL (based on the Common European Framework). These tests (at all levels) are targeted to a specific narrow range of levels; that is, a test-taker applies to take an assessment designed specifically to generate a rating at a given level or at least within a limited range of levels, not across the entire scale. Such an assessment would be used only for speakers who had already achieved a rating of Superior/3+ and only in cases where an ILR 4/4+/5 is the desired target. The potential certainly exists for developing new assessments exclusively aimed at ILR 4/4+/5, but to date, I am unaware of any specific plans to do so. This approach, however, may serve to address concerns of high-level test-takers, particularly those who self-identify as native speakers, about the criteria that are actually being used to assign ratings at these levels currently, an issue I shall briefly address here before moving on to question no. 5.

Test-takers’ understanding of the guidelines and assessment criteria: While it is commonly agreed in the testing community that not all native speakers automatically score Superior (ACTFL) or Levels 3-5 (ILR), pressure is being exerted on the community by those taking the test who consider themselves to be “native” speakers to articulate more clearly what the criteria are for these ratings. When educated native speakers read the guidelines/descriptions, they understandably self-assess at the top of each scale. This is less problematic for the *ACTFL Guidelines*, since it is not uncommon for native speakers educated in the target culture to score Superior in an ACTFL OPI (where there is no distinction after crossing the Superior border, meaning that an ILR 3-5 would all receive the rating of Superior). But when native speakers read the ILR descriptions for Levels 3-5, they also self-assess at the very top of the scale, and are often

surprised to be awarded a rating of “only” 3 or 3+. While this rating may be based on the speaker’s inability to perform certain functions at L4/5 as tested *in practice* by looking for the ability to tailor language both formally and informally (as discussed above), as well as the lack of spontaneous production of highly sophisticated language when discussing conceptually complex issues and the use of cultural references with facility, these expectations are not explicit in the ILR 5 description. It is not surprising that educated native speakers find ILR 5 the best match for their abilities, when ILR 4 clearly says “would seldom be perceived as a native”, and 4+ says “the individual would seldom be perceived as culturally native.”

Test-takers who self-identify as “native” speakers clearly need to be given more information about the expectations *in practice*, not just in the Descriptions/Guidelines, to attain a rating of 4, 4+ or 5 on the ILR scale. *Language Testing International, Inc.* (the ACTFL Testing Office) has developed supplementary explanatory material about the expectations for attaining a rating of 3+ and higher to be made available to test-takers, explaining that, while native speakers may believe that their language fits the ILR 5 description “speech on all levels is fully accepted by well-educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms and pertinent cultural references”; in practice, the assessment will only award a 3+ unless the speaker can demonstrate the ability to tailor language *both* formally and informally. In the absence of a new set of high-level assessments aimed at ratings of 4/4+/5, editing these descriptions may also help alleviate much of the distress currently caused when self-identified native speakers do not receive ratings of 4, 4+ or 5.

5. Might a given speaker’s profile matter to the end-user of the ratings? If so, what mechanisms are currently used and/or may be developed to provide this information?

When third-party independent testing is conducted, the testers and raters do not always have any information about the client requesting the test or the purpose for the assessment. In many testing practices, the only information provided to a tester is the candidate’s name (or a test ID number) and often the name of the requesting client. It is generally thought that testers and raters do not need, and in fact, perhaps should not have, any specific background information about the candidate or about the potential uses of the rating. I agree that a tester does not need this information in order to elicit and rate a speech sample using current ACTFL/ILR interview protocols. The only information provided by a tester or rater is a final rating, no matter the profile of the speaker, unless otherwise

requested by the client.

Obviously, end-users may request information about previous language acquisition and use *directly from the test-taker*, but the testers/raters are perhaps in a unique position to provide observations about the speaker’s linguistic profile, as well. In practical terms, a final rating is indicative of the tasks (limited to the ones elicited depending on the assessment tool) a given speaker can perform using his/her language, while sustaining the other criteria required for that rating. In a number of cases, clients have worked with test administrators to customize oral proficiency assessments based on specific job demands, for example, by developing role-play situations that are specifically job-related. A number of end-users have begun to develop feedback mechanisms that supplement the rating, and that may be used for a variety of purposes, including diagnostic assessment. Such feedback mechanisms may include questions about those features of a speaker’s language that prevent him or her from achieving a desired level (where is the breakdown?), for example, or that ask the tester to indicate the speaker’s acquisition profile (adult foreign language learner, heritage or native) and whether or not the breakdown from the next level was primarily functional, linguistic, or both, or what level of comprehension of the questions was exhibited by the test-taker (a kind of gauge of interpersonal listening).

I would suggest that more end-users reflect on what, besides a final rating, would be useful to know about a particular profile and that the testing organizations work with them to help provide pertinent information. However, an important word of caution is in order here: we must be cognizant of potential unintended consequences of using oral proficiency interviews to glean and record information about a speaker’s background. Official interviews are often used to make high-stakes decisions about employment, job assignments and even pay scales. In official commercial testing, for example, a tester is prohibited by law from asking questions about age, sex, race, color, religion or national origin, sexual preference, marital status, health and political viewpoint.⁴ Designing or using assessments that draw conclusions or make assumptions about the ethnic or regional origins of the speaker, for example, could lead to both the perception, and perhaps the reality, of having an individual or entire group disadvantaged. We must be careful not to create a context where discrimination, either real or perceived, based on linguistic profile might evolve.

While I support a standard set of criteria by which to assess the *functional*

⁴ Under the 1964 Civil Rights Act, prospective employers may not ask age, sex, race, color, religion or national origin. The other categories have been added subsequently by the Equal Employment Opportunity Commission.

proficiency of all speakers regardless of background, such as the ILR Descriptions and ACTFL Guidelines, we must consider the challenges inherent in such an approach now that the testing context has changed significantly from what it was when these guidelines were first written. Inside the ACTFL/ILR testing community, we need more discussion and more empirically based investigations into the complexities of testing in a context that is growing linguistically more diverse. We also need to make greater efforts to educate end-users and, perhaps most importantly, test-takers themselves, about the criteria and practical considerations involved in applying those criteria in currently used assessment tools based on the ACTFL Proficiency Guidelines – Speaking and the ILR Proficiency Descriptions – Speaking.