



A Guide to Measuring Listening

According to the ILR Skill Level

Descriptions - Listening

A Report on the DLI Listening Summits
ACTFL/DLI Contract #W9124N-10-C-0095

Prepared by
The American Council on the Teaching of Foreign Languages
December 2011

TABLE OF CONTENTS

Executive Summary	1
Section 1: Summary of the Summits	2
First Summit	2
Second Summit	5
Third Summit	8
Section 2: State of Current SLDs-Listening	10
Features to be added to the SLDs-L	13
Section 3: Usage/Application of the SLDs-Listening by the major stakeholders.....	19
Section 4: Findings	23
Section 5: Outcomes	32

Executive Summary

The ILR Summits on Listening are a U.S. government-wide initiative, sponsored by the Defense Language Institute Foreign Language Center (DLI-FLC) in collaboration with the American Council on the Teaching of Foreign Languages (ACTFL). The primary objective of these Summits is to create a community of understanding related to the skill of listening comprehension as described in the ILR. Three Summits were held: the first in September 2009 (funded under a different contract), the second in February 2011, and the third in June 2011.

This *Comprehensive Guide to Measuring Listening According to the ILR Skill Level Description – Listening (SLDs-L)* is a compilation of Summit goals, activities, discussions, and recommendation and is divided into five sections. Section 1 provides a summary of the presentations, activities, and products of the three Summits. Section 2 reports the current state of the ILR SLDs-L, including information on what currently appears in the descriptions and what is missing. For section 3, major U.S. government agency stakeholders were interviewed with respect to their specific uses and applications for the SLDs-L.

Section 4 consolidates the findings from the Summits, by organizing information by ILR level, according to the salient features for each level (task, accuracy, context, and content) and the linguistic and extra-linguistic factors that impact the ability of listeners to comprehend what they hear. Based on these salient features and influencing factors, a number of issues encountered in implementing the SLDs-L for the development of assessments are identified.

The Guide concludes with recommendations for interpreting the findings from the Summits, recommendations for the revision of the SLDs-L and suggestions for future areas of research.

Section 1: Summary of the Summits

Summit I

The first ILR Summit on Listening was held on September 16-17, 2009. The Summit was attended by 75 representatives of government agencies, language services industry contractors to the government and academia. The goals of this Summit were:

- Identify agency-wide issues involved in describing listening proficiency
- Discuss what works, what does not work, what is missing from the ILR Skill Level Descriptions for Listening
- Identify and clarify features of listening passages at various levels of the ILR Scale
- Determine topics to be explored at future Listening Summits

The Summit opened with remarks presented by U.S. Army Col. Sue Ann Sandusky, former commandant of the Defense Language Institute. The Summit featured three plenary presentations, given by the following individuals:

- Ms. Beth Mackey, U.S. Department of Defense: *“A Report on the Pre-Summit Survey”*
- Dr. Pardee Lowe, Jr. U.S. Department of Defense: *“Listen Up: Is Listening the Step-child or a Distant Cousin of the ILR.”*
- Mr. James Dirgin, U.S. Defense Language Institute: *“What is Different About Listening”*

Beth Mackey of the Department of Defense gave the opening presentation. Her presentation summarized the results of a survey that attendees were asked to complete before attending the Summit. The 11 question survey focused on how attendees use the ILR Skill Level Descriptions for Listening (SLDs-L) and their thoughts regarding the present form of the document. The survey was completed by 36 attendees. Among the findings was that 56% of respondents reported not using the ILR SLDs-L to assess participative listening, and 67% of respondents reported not using the ILR SLDs to assess overhead listening. However, 61% of respondents reported using them to assess non-participative listening. Other questions on the survey asked respondents their opinions about what they find useful about the ILR SLDs, what makes the ILR SLDs difficult to use, what features of language appear at some levels in the ILR SLDs but do not appear at other levels despite being relevant to that level, what features appear in the SLDs for other skills but not for listening despite being relevant, and what terms in the ILR SLDs-L are unclear or need to be defined.

Dr. Pardee Lowe's presentation discussed how to make the SLDs-L more useful while maintaining the ILR framework. His suggestions included cross-walking the SLDs-L with those from other skills, considering task versus the overall question, recognizing performance versus proficiency, recognizing productive skills versus receptive skills, and focusing initially on the big picture before "winnowing" the smaller details over time. His final point was to consider "product" versus "process". Dr. Lowe encouraged attendees to realize that the SLDs-L were written primarily from a product point of view and that current work on listening is investigating process(es).

The Summit featured four breakout sessions. These sessions ranged in length from 1 to 1.5 hours. Each attendee was randomly assigned to one of four groups for these sessions. Attendees were not in the same groups for each breakout session. There were two breakout sessions on the first day of the Summit and two on the second day. Each group was assigned a scribe and a facilitator to gather and record the group's findings. The topics of the breakout sessions were:

- Breakout Session 1: *"Cross-walking the Skill Modalities"*
- Breakout Session 2: *"Listening to Passages"*
- Breakout Session 3: *"Listening to Passages: Gut Feelings vs. What the ILR says"*
- Breakout Session 4: *"Where do We Agree, What is Missing, and What Still Needs to be Defined"*

The first task was to identify language features that appear in the SLDs-L for certain levels that could be relevant to other levels. The second task was to identify features of language that appear in SLDs for modalities other than listening and to indicate if that feature should also be mentioned in the SLDs-L. In the second breakout session, groups were asked to listen to audio samples and to discuss the speaker's purpose, the content and context, and the salient features affecting the comprehensibility of the sample. For each sample, the group was then asked to determine the type of communication in the sample (participative, non-participative, or overhead) and then to determine what a listener at each level of proficiency (L1-L5) could understand from the sample. In the third breakout session, groups listened to another set of four audio samples and discussed what listeners at each level might be able to understand from each sample and what the ILR SLDs-L indicate that listeners at each level should be able to understand. In the fourth breakout session, participants completed an activity sheet that addressed the following questions:

- Where is there consensus on what is already in the ILR SLDs-L
- What is missing from the ILR SLDs-L
- What terms/issues still need to be defined/determined

Each breakout session was followed by a whole group “reporting back” session. In these sessions a representative of the group presented the group’s finding to all attendees. There was an assigned moderator for each breakout session to facilitate the reporting back of information. A group discussion followed each presentation.

On the morning of the second day of the Summit there was a panel discussion among representatives of U.S. Government agencies. The topic of the discussion was “*How Agencies Perceive and Use (or do not use) the ILR SLDs-L.*” Agencies participating in the panel included the Federal Bureau of Investigation, Central Intelligence Agency, Defense Language Institute, Defense Intelligence Agency, U.S. Department of State and the U.S. Department of Defense.

Closing remarks focusing on future listening summits were delivered during the afternoon of the second day of the Summit by Dr. Donald Fischer Provost, Defense Language Institute.

Summit II

The second ILR Summit on Listening was held on February 14-15, 2011. The Summit was attended by 84 representatives of government agencies, language services industry contractors to the government, and academia. The goals of this Summit were to:

- Raise awareness on the topic of ILR Skill Level Descriptions for Listening (SLDs-L)
- Examine the adequacy of these Descriptions
- Build a community of interest in improving our understanding of listening
- Develop a common understanding of the construct of listening comprehension

Opening and closing remarks were presented by Dr. Donald Fischer, Provost, Defense Language Institute. The Summit featured an active interchange by ILR stakeholders through activities that included presentations and plenary sessions of general interest, breakout working groups, and panel discussions. Highlighted topics and discussion items included examining the adequacy of the current ILR SLDs for listening and current research findings about listening comprehension.

The Summit featured six presentations, given by the following individuals:

- Dr. Allison Blodgett, Dr. Amber Bloomfield, Dr. Jared Linck and Dr. Sarah Wayland, University of Maryland/CASL: *“Review of Current Literature on Listening”*
- Dr. Gary Buck, Lidget Green, Inc.: *“What listening comprehension is, how we should test it, and how we could do a better job of interpreting the ILR Skill Level Descriptions”*
- Dr. Ray Clifford, Brigham Young University: *“Defining Listening Proficiency: Can We Construct a Construct?”*
- Dr. Pardee Lowe, Jr., Department of Defense: *“Listening: Simplifying the Complications.”*
- Ms. Beth Mackey, Department of Defense: *“What the ILR Skill Level Descriptions Say About Listening – And What They Do Not Say”*
- Dr. Elvira Swender, ACTFL: *“Recapping the September, 2009 ILR Listening Summit”*

Following the opening remarks, Summit activities began with two brief presentations by Dr. Elvira Swender, ACTFL, and Ms. Beth Mackey, Department of Defense. These presentations recapped the events and discussion items of the September, 2009 Listening Summit. Dr. Swender identified the goals of the 2009 Listening Summit and reviewed responses of the 2009 Listening Summit attendees regarding the usefulness of the ILR-SLDs-L. Examples of these comments included providing a common metric for test development and curriculum that is criterion-referenced and describes what listeners can do, and the difficulties inherent in working with the ILR-SLDs-L (e.g., too connected to speaking, too derivative of reading, terminology lacks

clarity, descriptions are not clearly developmental and not clearly functional). Ms. Mackey's presentation consolidated the issues and questions that were raised during the first Summit. Among these issues and questions were: how to define different types of listeners (native, heritage, learner), how to account for cognitive load and memory, how to better describe the skill, and how to more clearly define the hierarchies.

On the second day of the Summit, Dr. Clifford delivered a plenary presentation entitled "*Defining Listening Proficiency: Can We Construct a Construct?*" that reviewed some of the complex challenges associated with describing language and suggested some ways to reduce that complexity. Dr. Clifford also proposed a refined ILR model by which listening comprehension might be reliably measured. The other keynote presentation on day 2 was delivered by Dr. Gary Buck, President of Lidget Green, Inc. addressing three topics: "What is listening comprehension", "How should it best tested", and "How could we do a better job of interpreting the ILR SLDs".

This Summit also included a panel discussion in which government agency representatives discussed the challenges faced by the U.S. Government when creating listening tests, how different agencies assess listening, and real world applications, i.e., how and why listening skills are critically important in the work environment. Members of the panel were: Ms. Rachel Brooks, Federal Bureau of Investigation, Dr. Donald Fischer, Defense Language Institute, Mr. Tom Haines, Department of Defense, Ms. Christina Hoffman, Foreign Service Institute, Dr. Pardee Lowe Jr., Department of Defense, and Mr. Dan Scott, Department of Defense. Panel members were asked to share how their agencies assess listening and to reflect upon the implications of the Summit activities and outcomes vis-à-vis the needs of their agency.

The Summit featured three 45-60 minute breakout sessions. The first two occurred on the first day of the Summit and the third occurred on the second day. Participants were randomly assigned to one of four small groups for each session. Each of the groups was composed of approximately 20 participants and participants were assigned to different groups for each session.

In the first breakout session participants were asked to identify the tasks and topic domains that are critically important to the work tasks performed by personnel of their agency and to discuss whether or not these task/topic domains were already mentioned in the ILR SLDs-L. If these did not appear in the descriptions, group members were asked to discuss if they should appear. Each small group was asked to address level-specific tasks and topic domains in a different to ensure the through treatment of each level.

The afternoon of the first day began with a presentation by a team of researchers from the University of Maryland/CASL (titled “*Review of Current Literature on Listening*”) that detailed the current research findings on the factors that impact listening comprehension. A panel of four, Dr. Allison Blodgett, Dr. Amber Bloomfield, Dr. Jared Linck and Dr. Sarah Wayland, presented the findings from the literature review they conducted on listening. Their presentation focused on the characteristics of listening passages that influence comprehension in a second language, examining these from four perspectives: amount of information, complexity, organization, and auditory features. Length, density, and redundancy were all mentioned as variables – redundancy not necessarily facilitating the task of listening if the redundancy merely adds information overload rather than providing clarification for the listener. How information is presented was also cited as a significant variable, i.e., the topic, concreteness, directness, vocabulary and idiom use. Also considered were auditory features such as accent, hesitation, rate of speech and pauses. Dr. Linck shared his findings on why working memory is important in L2 listening and how memory interplays with listening factors. The presenters reminded the Summit participants that reading is not the same as listening, that listening in one’s first language is not the same as listening in a second language, and that there is still a great deal of research to be undertaken on this topic.

The task of the second breakout session was to identify factors that influence listening comprehension. Session participants were tasked with discussing the adequacy of the ILR SLDs-L when dealing with the linguistic factors cited in the University of Maryland/CASL presentation. Each small group was asked to evaluate 5-6 factors that influence listening comprehension.

The chart below shows the factors that were assigned to each group.

<u>Group A</u>	<u>Group B</u>	<u>Group C</u>	<u>Group D</u>
Speaker accent	Degree of Redundancy	Syntactic complexity: Sentence structure	Passage Topic
Distortion and Noise	Type of Redundancy	Infrequent words	Discourse Markers
Clarity	Information Density	Culturally specific vocabulary and idioms	Rhetorical Structure
Hesitation and Pauses	Overall Length	Directness of text	Overall Coherence
Speech Rate	Syntactic complexity: Negatives	Concreteness	Local Coherence
Passage Type: Orality	Passage Type: Overhead dialogue vs. Monologue		

The focus of the third working group was for small groups to discuss the questions: “Where are we a community” and “Where do we go next.”

Summit III

The third ILR Listening Summit was held on June 16-17, 2011. The Summit was attended by 70 representatives of government agencies, language services industry government contractors, and representatives from academia. The goal of this Summit was to produce by consensus a written document that expands, extends, defines, and annotates the factors that influence listening comprehension along and across ILR levels 0 through 5.

The specific objectives were to:

- Identify the salient, critical, and distinctive features of each level
- Incorporate conditions, settings, and/or factors that influence listening comprehension across the levels
- Provide examples of how the features, conditions, and factors influence listening comprehension at each level

Opening remarks including the Summit goals were presented by Dr. Elvira Swender, Director of ACTFL Professional Programs. Pardee Lowe, Jr., of the Department of Defense, gave a brief presentation entitled: “*Speak Up! Listen Up! Write It Up!*” to orient attendees to the issues and activities of the Summit. James Dirgin of the Defense Language Institute provided instructions to the attendees for tasking during the remainder of the Summit. The Summit featured three breakout sessions, two reporting back sessions, and panel and audience discussions. Each level

of the ILR SLDs-L was explored in depth in order to identify the salient tasks, contexts and content areas, accuracy features, linguistic and extralinguistic conditions expected of listeners at that level.

The breakout sessions of this Summit were organized in a manner similar to the previous summits with attendees assigned to one of four small groups of 15-20 people for each session. Attendees were in the same groups for the first two working group sessions. The topic of the first breakout session was “Focus on Salient Tasks, Accuracy, and Context/Content by Level.” The task of this session was for the group to complete a matrix that focused on the salient tasks, accuracy, contexts, content, significant conditions and other factors at each level. Each group was responsible for one level. Levels 4 and 5 were combined into a single level. The matrix sheet for each level had five columns:

- What can listeners at this level do that those at levels below cannot do?
- How well can listeners at this level sustain this task?
- In what contexts can they do this?
- On what content areas can they do this?
- Other Comments

The task of the second breakout session was the same as the first except that groups looked at the factors across levels. The groups completed a matrix with seven columns. In addition to the five questions listed above, the group also discussed:

- What linguistic factors enable user to sustain this task at this level?
- What are some non-linguistic factors that can be handled by a user at this level that might be a problem at lower levels?

The topic of the third breakout session was “A View of the Scale by Criteria”. Attendees were broken into four groups with each group being responsible for discussing one listening criteria across all levels. The four criteria were accuracy, task, content/context, and other conditions. They groups reported on each criteria and how it is treated at all levels of proficiency.

Section 2: State of Current SLDs-Listening

Users of the SLDs-L agree that the present version of the document is useful for several purposes including:

- Providing a common metric for test development and curriculum
- Making the distinction between participatory, non-participatory and overheard listening
- Setting expectations for what a listener can and cannot do

There is consensus among SLDs-L users in several areas:

- Mention of speed is helpful
- Psychologisms should be removed (e.g., interference from native language)
- References to native speaker should be removed (“native speaker” is a psychologism).
- References to noise are helpful (but there should be more of them, and they should be more nuanced).
- Differentiation of participative vs. non-participative

Users also credit the present document for recognizing the influence of context and situation and making distinctions by level. Likewise, they appreciate that it is criterion referenced.

However, users indicate that there are many areas in which SLDs-L prove difficult. These include:

- They are too connected to speaking yet too derivative of reading
- They lack examples
- Understanding of key terms needs to be tightened in order to more clearly differential levels
- They lack empiricism
- Relationship between authentic texts and noise
- Can-do versus cannot
- Interaction of text length and level

Users identified several features that need to be clarified in the SLDs-L, including:

- Effect of “adverse conditions” and their impact on comprehension (at L2 and L4)
- Increasing ability to “detect emotional overtones” (L2+)
- Participatory, non-participatory, and overheard listening throughout the scale
- Definition/Concept of “interference” from native language (L1+)
- L2+ listeners are portrayed in terms of stress: “but under tension or pressure may break down”
- Discourse levels understood at each level (L4: “unpredictable turns of thought”)
- Length of spoken language

- References to understanding of native speech (L1 and L3)
- Incomplete references to comprehension of grammatical forms

Users also had many suggestions of what is missing from the SLDs-L:

- Non-standard variants
- A way to talk about factor interaction (e.g. nonstandard dialect plus background noise, or familiarity with format helping alleviate background noise)
- Length needs to be addressed more thoroughly (and at levels above 1+)
- Research on cognition and memory (automaticity) needs to be taken into account
- Descriptors should stress more global nature of listening comprehension, as opposed to reading comprehension. But: depending on purpose and level, details may be all that can be understood.
- Purpose for listening
- Role of aids to listening (PowerPoint, repeated playing of recordings, dictionaries, etc.)
- Non-native speakers providing the input to listening
- Accent
- References to modern audio sources
- Non-participative listening below L3 (i.e. newscasts)
- Recognition of the role of technology
- Participative, non-participative, overheard need to be fleshed out
- No hierarchy of tasks at each level. Everything is comprehension. What is your understanding at each level?
- Operationalize concepts like “used to speaking to non-native listeners”
- Regionalism vs. dialect vs. Accent
- How different dialects in the same passage interact with each other – does the presence of more than one dialect affect the level of the passage?
- Learner vs. native and heritage speakers – seems from the scale now that a native speaker must be an L5
- functional vs. development scale
- L1 should be able to get elemental words or concepts when text is not basic “survival” text. Really L0+ level.
- Guidelines use speaking as foundation, should be separate skills with separate levels.
- ILR descriptions are guidelines for radio and TV only – need to be supplemented for new listening modalities (e.g., new visual support from things on internet. Skype, VoIP)
- Need to define guidelines for both participatory and non-participatory tests. What is expected from each?
- Take cultural references into account in determining level.
- Are terms limiters? Or descriptors?
- [SUBJECT OF DEBATE - unresolved] Should remove descriptors that describe mastery at other levels? Only have relevant descriptors.... “Usable” skills vs. skills that can be used with “general utility”. OPI uses negative statements to establish ceiling level.
- Continuum from level to level regarding certain features and what is present at each level.
- Checklist / series of statements that describe the level (a good checklist)

- Multiple chances on test to demonstrate cultural knowledge, or to explain when you don't get a cultural reference?
- Should there be time frame on cultural references? Or how well-known?
- Some tests *do* use video for assessments, so we need descriptors for how video should influence listening
- Explicit mention of how listeners at different levels deal with metaphor and idioms, figures of speech, irony.
- How different level listeners deal with humor and definition of types of humor they can process at each level
- Age of listener? Does that condition what they can comprehend?
- Culture
- Knowledge of language
- Knowledge of world
- What educational level defines an L3? L4?
Listening is a challenge – notions of different types of listening. Need to review the Guidelines from the point of view of the various types of listening beyond participative? Do we need more than one set of Guidelines for listening? Is the integration of these different types of listening into one set of Guidelines a better route?
- There is lacking a greater range of listening contexts – difficult for test developers. So there is a need for fuller descriptions and concrete examples similar to speaking.
- The notion of length of passage is important. In training, there should be a great variety of speeds, of voices, of accents, etc. so that it can then be built into the testing.
- The non-linguistic elements need to be incorporated into the Guidelines.
- Notions of authenticity in listening passages.
- 0 and 0+ are really achievement levels. Level 1 is generally acultural. However, L2 and above allow for more authentic materials. How should the Guidelines address this question of authenticity of listening texts? Can authenticity be incorporated into the lower levels?
- How to address the free flowing aspect of a text (which causes multi-level issues of the text) used for listening comprehension? The text in reading has one focal level, but in real-world conversation this is not the case.
- Issues of content variability at lower levels beyond survival needs.
- The “standard dialect” issue.
- On Levels 4 and 5, the term “density” should be included as characteristic of the listening text.
- On Levels 3 and above, the listener needs to be able to handle a variety of voices within the same listening text.
- Where does the idea of background knowledge figure in the Guidelines, or in a Preface?
- Should there be a training set of Descriptive Guidelines to help us consider the various ways in which an individual learns to listen and comprehend?

As part of the survey administered before the first Summit, attendees were asked to review the SLDs for other modalities (reading, speaking, writing) and identify factors, contexts, features,

etc. that are mentioned in the descriptions for that skill but not for listening but would be useful in the SLDs-L.

Factors/Features from Reading:

- Ability to draw inferences directly from the linguistic aspects of the text (L3 SLDs Reading)
- Ability to get to the gist of main and subsidiary ideas in texts (L2+ & L3 SLDs Reading)
- Increased strength of performance when the topic is familiar (L2, L2+ & L3 SLDs Reading)
- Can guess at unfamiliar vocabulary if highly contextualized (L1+ SLDs Reading)

Factor/Features from Speaking:

- “Comprehension of normal native speech is typically nearly complete” (L2+ SLDs Speaking)
- “Structural precision with sophisticated features that are readily, accurately and appropriately controlled” (L3+ SLDs Speaking)
- “Cultural references, proverbs, and the implications if nuances and idioms may not be fully understood” (L3 SLDs Speaking)
- SLDs Speaking offer examples
- Users of the SLDs who attended the first summit had many suggestions of features to be added to the SLDs-L. They are listed in the table below.

Features to be added to the SLDs-L:

Participatory, Non-participatory and overheard listening throughout the scale	More consistent descriptions of text type, length and grammar complexity at all levels
Clarity on dealing with noise	More acknowledgement of the modern conditions of speech
Can often detect emotional overtones needs to be included at L3 and above	Can and cannot do statements and context domains at each level
Ability to deal with abstract speech needs to woven in	Ability to handle cultural references, slang, idioms, standard vs. non-standard language and dialects
Definition of “interference” from native language at L1+	Effect of vocabulary
Clarify some of the relative terms such as “increased ability”	How stress affects listening ability at levels other than L2+
L4 should be able to recognize what is the point of a spoken argument and what points constitute an aside	Comprehension of grammatical forms
Dealing with humor and sarcasm at the upper levels	Concept of main idea needs to be more clearly referenced at the higher levels

Users had the following suggestions of terms/ideas that still need to be defined:

- Dialect and standard language
- Qualifiers such as “most”, “often”, “simple”
- Noise/noise levels

- Words common to SLDs such as fact, concrete, well-known and control
- Ideas/terms related to audio such as “length of utterance” and “unfavorable conditions.”
- “Professional needs” in “understand....speech pertinent to professional needs”
- “implications and inferences”
- Hierarchies such as:
 - Everyday vocabulary versus general vocabulary
 - Can discern relationships among sophisticated listening materials in the context of broad experience
 - Tentative state of passive grammar
 - Emotional overtones

The activity from the first breakout session of the first Summit asked participants to review the SLDs-L and identify features that currently appear at specific levels but should (or should also) appear at other levels. Additionally, attendees were asked to identify features that appear in the SLDs for other skills but do not appear in the SLDs-L despite being relevant.

The table below shows each feature identified and the comments suggested by the attendees:

Noise	<ul style="list-style-type: none"> • Currently appear at L2 and L4. Also relevant for all L1, L3, L5 • Important enough to be mentioned in the preface • From Level 2 and up there should be overt mention of the effect of noise • Difference between noise and background sound needs to be defined
Participative vs. Non-Participative	<ul style="list-style-type: none"> • Participative is assumed at L0-L2 • Preface should include definitions of participative and non-participative • Non-participative listening should appear at L0+ - L2+. It currently appears only at L3 • Participative listening appears at L1-L5. It is relevant at L0 as well.
Overheard	<ul style="list-style-type: none"> • Must be defined: is it real time or asynchronous • Intent or pragmatic issues: Should appear at L3 and higher
Adverse Listening Conditions	<ul style="list-style-type: none"> • Measure/describe how these impact listening at all levels (ex. L1 even a small amount of adverse conditions will prohibit comprehension but the effect will be less pronounced at L2)
Effect of length of spoken language	<ul style="list-style-type: none"> • Currently appears only at L1 but is relevant for L1-L5 • Note-taking is not mentioned in the SLDs but perhaps it should be • Role of repetition and redundancy
Rate of Speech	<ul style="list-style-type: none"> • What is “normal” rate and how should it be measured? • Definition must include more than fast/slow; it must address variable speed • What is meant by “shows spontaneity in understanding by speed?” • What is the effect of the interaction of speed with factors such as idioms, distortion, mumbling, slurring • Should there be a function scale for dealing with speed? • Relevant for all levels
Utterance Length	<ul style="list-style-type: none"> • Only mentioned directly in lower levels (L0, L0+, L1) and implied at other levels but should be mentioned directly at all levels.
Main Idea	<ul style="list-style-type: none"> • Needs to be more specifically defined and tied to level of spoken text • Should appear at all levels of the SLDs-L
Interference from Native Language	<ul style="list-style-type: none"> • Implies that the scale can only be used for 2nd language learners • Is the real issue lack of knowledge of the target language

	<ul style="list-style-type: none"> • Interference of the native language is only an assumption of why comprehension breaks down Only at L1, L1+, L2 • Word order from L1 can interfere with perception of L2; may be a psychologism, you don't know why someone isn't understanding...unless the person is given a task specifically targeting the difference between L1 and L2; or not, if you don't have information about their L1, or you're trying to design a more general task, in a more functional situation, you don't know that it's interference. • Maybe it should be stricken from Level 1 as well. But what does "interference" tell us about their proficiency level?
Mastery	<ul style="list-style-type: none"> • There is a conflict in terms of the language used and how it related to mastery • What are the tasks at each level that have to be accomplished • Certain factors (such as auditory interference) should be mentioned at every level
Emotional Overtones	<ul style="list-style-type: none"> • Appears only at L2+ • Somehow ILR needs to capture body language, gestures, and emotional cues • "Sometimes" for 2+, "often" for 3, but if not understanding words, low level listener may rely on overtones more than lexical content
News Items	<ul style="list-style-type: none"> • Mentioned at L2 and L3 • Used a lot for testing but not enough description of what is in the news • Needs definition of typology
Dialect	<ul style="list-style-type: none"> • German (standard, dialects everywhere used in daily life) • Pashtu (no standard dialect, a lot of dialects that are spoken in different areas) • Russian (no dialects heard in the course of daily life, small features that are easily picked up) • Considerations of dialect should be removed in light of these differences across languages • Chinese is notorious – wait until L4 to discuss ability to understand dialects • The general picture of L3 in German has to include some dialects, but dialects may not enter into Russian proficiency at all • Arabic – ability to detect dialect must be entered into the concept of proficiency. What exactly is "a" standard dialect? This needs to be better defined. MSA may not be the standard dialect for certain situations. • Order of dialectical learning depends on where you learn – formal training will result in learning MSA first, while learning on the street will result in learning a dialect first. So dialect may need to be considered at all ILR levels • Even at 0+, your basic greetings may be delivered in a dialect. Which Levantine will we use? On one test, you see all the dialects. Seems too ambitious. • Should the ILR be a learner-centered scale or an ability-centered scale? Individual differences come much more into play when you consider it as a learner-centered scale. Makes it non-functional. • Speed interacting with other factors like idioms, distortion, mumbling, slur. • At what point can speed be introduced in the lower levels? Can people cope? Is there a clear boundary, or should it just start at 2+ and above? This is authentic, so you go by the natural speed. What you can control is the measure of the level of comprehension. This also interacts with participatory vs. non-participatory/overheard listening. Participatory you can ask for clarification – if it's non-participatory, does it have to be artificially slow for it to work? Functional scale for dealing with speed? • Define the term vis-à-vis regionalism, slang, dialect—should be in preface
Tension and pressure	<ul style="list-style-type: none"> • Currently appears at Level: L2+/should appear at Level:L3 • What breaks down under pressure? At other levels, this information should be specified, and how much do you "break down?" Is your mastery lost? • No one is going to do as well under pressure. • How does it relate to non-participatory/participatory distinction? How does it relate to the speed of the speech, cognitive load? Hard to operationalize?

	<ul style="list-style-type: none"> As a social psychologist, the problem is the lack of operationalization in the terms in the ILR. How can you make an observable behavior out of “comprehends” or “understands?” This may be for the test construct, test design? It should be somewhere. It would be nice if the ILR had it – it provides a framework. It would be useful for testing and teaching, and those applications of the ILR. Good English text called “Eavesdrop” that teaches one about every level of the society. Students learn a lot about the society associated with the language. Record conversations heard while traveling to use as stimuli.
Socio-Linguistic/Cultural References	<ul style="list-style-type: none"> Currently appears at Level:3+,4+/should appear at Levels 3,4
Unscripted Speech	<ul style="list-style-type: none"> With visual vs. no visual support, taking speed and idioms into account Level of predictability scripted at lower levels Less scripted at lower levels
Repetition	<ul style="list-style-type: none"> Be mentioned until controlled Rarely needs to at higher level A participative, overheard and... (not finished)
Context of Communication	<ul style="list-style-type: none"> “Face-to-face” change to “participative”? Due to Skype, web conferencing – socio-linguistic terms and definitions apply? Distinction more effective than “face-to-face”: mediated, in-person; who is ENGAGED in communicative act? (impact of non-verbal language)
Listening Comprehension vs. Global Communication Skills	<ul style="list-style-type: none"> Seems as if “real” communicative listening doesn’t happen (as described) until level 2+; Below that seems to be a “learner” environment; the entire scale should be non-participatory? A participant who doesn’t ask questions
Understanding Abstract vs. Concrete speech	<ul style="list-style-type: none"> Currently appears at level: L4+ and above/should appear at Level: 3, 3+, 4 Associated with L3 tasks, but not mentioned there. Implied by mention of listener using context at lower levels, and not mentioning it later. Is there a disconnect in the descriptions across modalities? Is speaking ability (use of abstract language) considered higher than listening ability, or should abstract be par to level 3 listening abilities (conversations between educated native speakers on abstract topics?)
Understanding of basic cohesive features	<ul style="list-style-type: none"> Currently appears at Level: L1+/should also appear at Level: should not appear at any level (psychologism) Grammar features – should we describe specific features at each level? This assesses misunderstanding by assuming nature of misunderstanding (psychologism) – <i>only possible with participatory assessment.</i> In non-participatory it is hard to verify
Content	<ul style="list-style-type: none"> Travel requirements, basic survival needs at 1 such meal, lodging, transportation Should also be included at other levels
Following unpredictable turns of thought	<ul style="list-style-type: none"> 4 – sometimes 5 – always How do we quantify the comprehension?
Relative understanding of concrete/abstract functions	<ul style="list-style-type: none"> Currently appears at Level:L2 Concrete first mentioned in L2+ assessment (“some discussions on concrete topics”) what is the role of concrete vs. abstract in listening comprehension

Purpose of Text	<ul style="list-style-type: none"> • Currently appears at Level 3+ but should appear throughout • All levels should indicate WHO the audience is • What the level of the test is in the authentic
Context	<ul style="list-style-type: none"> • High and low context languages (especially important in overheard conversations) • Useful for learning not just “spook” purposes
Other factors	<ul style="list-style-type: none"> • Familiarity with topic: an asset in helping to complete/carry out the task • Authenticity: General level of authentic native speech, not modified by speaker for a non-native; interlocutor responsibilities • Multiple speakers in a fast-paced conversation: All levels need to address different listening content • Topic Shift: Currently appears only at L4 • Beyond the Lines: Currently appears at L4 • Text types: Needs to be delineated overtly? • Quality of Speaker Delivery • Cognitive Load: It’s implied in the higher levels. Should it be explicit? • Non-verbal communication is a huge part of face-to-face listening comprehension • Repetition and Redundancy • Interlocutor – how much sympathy? • Memory load • Will listener make every effort to understand (listener’s purpose) • Listen to Foreigners Speaking the Language: Where in the SLDs should this go? • Cultural Context: Background knowledge based on culture • Accuracy Statements: Need more or more precise accuracy statements • Sympathetic Listener needs to be defined; where in the SLDs should it go?

Following this breakout session, there was a whole group “reporting back” where all attendees could hear everyone’s comments and suggestions. Col. Sandusky suggested to “impose a structure. At each level there should be specific tasks, some kind of discussion of accuracy, and context, as well as explicit acknowledgement of the domains you would expect the person to be able to operate in. Proficiency according to DLI is that it doesn’t matter where you learned it, and it will apply across a wide variety of domains. But at L1, travel language, survival needs are specified – does that need to be included?” She went on to say “the way the ILR descriptions are written, it appears that participatory is given more weight or credence, but a lot of agencies deal more centrally with non-participatory listening. Because of this, the way the ILR descriptions are presented don’t help for the development of the DLPT. Overheard is a subset of non-participatory. We have to figure out how to separate the levels for non-participatory proficiency.” Dr. Fischer suggested that “the uses are different now – emphasize non-participatory; there’s a lot more media that is presented in various forms than we did when ILR was first conceptualized. We have to think about what is really being done with the DLPT – it is

more global than it is given credit for. Presents content that will be useful to a wide variety of professions, even the participatory tasks are still giving professionals skills that will be useful to them. You may need to distinguish at the performance level but not the proficiency level, the ILR descriptors are very useful as is.” Another summit attendee suggested that “There is a lack of attention to culture. It comes in at higher levels of listening, not lower levels. Cultural awareness should be emphasized at all levels. At L1, appropriate greetings may be important. In some languages, knowledge of proverbs is important at 0+. “Social formulae” vs. allusions to proverbs? In Arabic, proverbs are omnipresent in speech.”

Section 3: Usage/Application of the SLDs-Listening by the Major Stakeholders

Several U.S. Government agencies provided insight on how the SLDs-L are used by their agency. The Department of Defense /NSA reports that the SLD-Ls are used for rating listening passages to be used in testing and training:

“Our listening tests (used when there is no DLPT) are selected to be at either Level 2 (HIGH) or Level 3 (THRESHOLD/SOLID) and the test taker provides a translation on the Level 2 tests or a transcript for Level 3 tests. Passage rating also informs all of our training design - sequencing the difficulty of materials, for example, in all of our classes, whether for general proficiency or for work related training”

The agency goes on to report that the SLDs-L are used by test developers and classroom teachers (both government and contract personnel). Most of the agency’s instructors and test developers have taken a Passage Rating class (an instructor-led version of the class as well as the NFLC online course are available). The SLDs-L are operationalized by classroom teachers, contractors, and test developers when selecting classroom and test materials. The Department of Defense tests listening and finds the SLD-L’s usable, but not specific enough. The agency also reports the following:

- There is no taxonomy of listening texts (in the SLDs-L)
- The accuracy statements are for the most part minimal.
- There is little guidance on use of authentic materials.
- There is little of no guidance on the distinctions between participative, non-participative, and overheard listening.

The agency reports that when specificity is lacking within the SLD-L’s, crosswalks are made to the SLD’s in the other skill modalities (speaking, reading, & writing). They suggest that “it would be helpful to build out the ILR SLDs so that content/context, task and accuracy statements are consistently addressed across all levels. Outside research is sometimes invoked, but it often is aimed at the lower levels of the ILR scale, and is highly analytic where the ILR scale is holistic.”

The Department of Defense/DIA reports the following about the SLDs-L

Question:

For what purposes are the SLDs-L used?

Answer:

As part of the greater understanding of language competencies across the four modalities: listening, reading, speaking, and writing, listening comprehension is used to base-line an individual's capability to understand foreign voice/speech. SLDs-L is used in the context of proficiency and to a lesser extent as an evaluative/diagnostic assessment of an individual's existent capability and what may be the potential barriers to improved and increased listening comprehension capability.

Question:

By whom are they used?

Answer:

For the most part managers and foreign language officers use the SLDs-L to help determine relative proficiency levels for foreign-language enabled professionals in the context of their work and job skills as they are exposed to foreign language material and sources.

Question:

How are they operationalized?

Answer:

Operationally, listening skills are used in both passive/non-participative as well as participative environments. Some personnel focus on the passive spectrum of listening: observe speech and audio material without having to respond in a discourse/dialog environment. Others find themselves in a participative listening environment where communicative skills and production skills: speaking and reacting are essential parts of the listening continuum. As well, there is a combination of both passive and participative listening environments used operationally.

Question:
Does your agency test listening? If yes, what issues does your agency encounter when interpreting and implementing the ILR –SLD’s for test development?

Answer:
We test listening as part of the DLPT family of tests: listening comprehension DLPT as a test of record. We also test for speaking using the Oral Proficiency Interview, however we do not recognize a listening score as part of the OPI evaluation. Although, listening comprehension is a direct consequence of speaking competence. With several exceptions, there is a ‘dual-modality’ OPI available for use, however these tests and languages may only be used when a DLPT 5 or DLPT IV do not exist. Additionally, it is understood that the evaluation of listening as part of the dual-modality OPI may only be accurate as an assessment up to ILR1+, perhaps 2. It is not a very robust or consistent assessment.

Question:
What additional information, supplements, compendiums to the SLDs-L would be beneficial to your agency’s work with listening comprehension?

Answer:
A complete distinction, if possible, between passive and participative listening competence. We are not sure that the ILR listening comprehension has the appropriate fidelity to discern the differences. Perhaps a comparable scale between passive and participative listening may be more appropriate as an assessment/evaluative tool. Given in our community, the exclusivity of passive listening as a job skill rarely requiring participative skills, there may need to be separate evaluations. However the research and test development to discern the two as being exclusive with marginal overlap may be lacking.

The Department of State/FSI reports not using the SLDs-L despite offering listening testing. For this agency, interactive L/C is embedded in its Speaking (S) test. As a result, the SLDs for Speaking (SLDs-S) is used to rate the sample. The FSI S test has been developed in mid 1980s and amended several times to better suit Foreign Service needs and obtain a ratable sample. All amendments have been made in accordance with SLDs-S. Listening has been cited as a much needed skill among Foreign Service personnel. A discrete Listening test may become part of the future testing. The agency finds that examples included in the SLDs are extremely helpful in

understanding the applicability of the scale. Raters relate to the example part of the scale easier than to the scale itself when they need to apply it to a testing sample.

The Federal Bureau of Investigation reports having limited if any use of the SLDs-L since it uses the Defense Language Proficiency Test (DLPT) to test listening.

Section 4: Findings from the Summits

This section identifies and organizes the salient features for listening at each of the ILR skill levels. These features are categorized in terms of the tasks, expectations of accuracy, contexts, and content areas that are most characteristic of the level and that distinguish the level from other levels. Following the salient features’ chart, is a chart that contains a listing of linguistic and extra-linguistic factors that may influence a listener’s ability to understand at that level. These factors appear at all the levels for which they should be considered when assessing listening comprehension. The individual, level-specific features and factors are followed by a summary of the issues that are encountered when using the ILR SLDs-L to assess listening at specific levels and across levels, and several suggestions for addressing these issues.

Salient features for Listening at ILR Level 1

Tasks	Accuracy	Context	Content Areas
Understand short and simple utterances (e.g., statements, questions) that contain high frequency vocabulary and structural patterns	Most accurate with high frequency, predictable speech presented in simple sentences with topical cohesion	In participative context: requires a sympathetic interlocutor for routine exchanges and simple transactions	Self and immediate world Basic needs, travel, and courtesy requirements
Understand redundant and repetitive language of literal and factual information that is purpose specific	Accuracy improves when meaning is supported by context Ability to understand may be inconsistent depending on content/context factors	In non-participative context: announcements, radio or TV broadcast, eavesdropping	Survival related areas: food, shelter, health, travel, security, weather Basic high frequency words, base nouns, action verbs, basic personal information, kinship terms
Understand a standard dialect and accent			

Factors that may influence Listening at ILR Level 1

Factors that influence Listening at L1	Comment
Automaticity	Automatic processing is at the word level and is idiosyncratic in that this will depend on what the individual listener has been taught and has learned
Clarity of speech	Speech must be delivered more clearly than normal
Cohesion	Basic cohesive features may be unreliably understood; reliance is on topical and not linguistic coherence
Cultural load	Survival, high frequency cultural norms. Low cultural load will ease the task of understanding.
Directness of text	Best understood when context supports meaning
Familiarity of topic; previous knowledge of topic	Topics must be highly familiar. Can be influenced by where and how language is learned.
Frequency of vocabulary	Only vocabulary related to the most elementary needs. Basic, high frequency vocabulary. Cognates facilitate comprehension.
Hesitation and pause	May be required for understanding
Information density	Research required
Length of passage and listener fatigue	Comprehension diminishes as discourse gets longer
Number of speakers	Can tolerate up to two speakers
Purpose of listening text	Comprehension may be task or purpose specific; listener needs to know what he/she is listening for
Rate of speech	Requires slower than normal rate of speech
Redundancy	Requires frequent repetitions or paraphrasing
Sentence structure	Simple questions and answers, statements. Basic unmarked structures, straightforward discourse, logical order. Coherent but not cohesive.
Syntactic complexity	Misunderstandings arise due to overlooked or misunderstood syntax and other grammatical clues
Unfavorable listening conditions	Needs for the entire message to be audible
Visual support	Less important if topic is highly familiar but can add meaning and content when attempting listening tasks at a higher level

Salient features for Listening at ILR Level 2

Tasks	Accuracy	Context	Content Areas
<p>Understand the main idea plus details</p> <p>Able to follow a thread and understand the essential elements of information collectively from the core (intent of the speaker)</p> <p>Able to follow the sequence of facts; and to understand the connection between the events being reported</p> <p>Able to gather factual information and understand the relationship among the facts</p> <p>Understand the mutually intelligible form of the standard language</p> <p>Understand cause and effect</p> <p>Able to distinguish between time frames</p>	<p>Most accurate in understanding temporal, factual and special details and relationships between different messages in a paragraph length narration, description, or instructional text</p> <p>All of the essential elements of information</p>	<p>Participative listening</p> <p>Overheard – monologue</p> <p>Explicit information</p> <p>Factual information</p>	<p>Every day, common, personal and general topics</p> <p>Well-known current event topics</p> <p>Assumed common knowledge</p> <p>Routine workplace topics, job requirements and topics related to one’s own specialized, professional field.</p>

Factors that may influence Listening at ILR Level 2

Factors that influence Listening at L2	Comment
Automaticity	Processing is at the sentence level with systematic errors when language becomes more complex
Competing voices	Can deal with overlapping speakers until the overlap interferes with the core message
Cultural knowledge	Would not be expected to understand culturally loaded descriptions (e.g., weddings, holidays, ceremonies).
Density of information	Research required
Directness of texts	Understand the facts but not “between the lines”
Factual nature of the content; concreteness of topics	Most accurate in understanding temporal, factual and special details and relationships between different messages
Frequency of vocabulary	Every day, common, general words and phrases; high frequency media vocabulary
Information density	Should be at word or prepositional level
Language variety	Standard language
Length of passage	Research required
Linguistic knowledge	Familiarity with linguistic factors will facilitate comprehension
Previous knowledge of the content	Facilitates comprehension
Rate of speech	Can handle normal rate of speech without abnormal pauses
Redundancy	Some repetition or rewording may be necessary for comprehension
Sentence structure	Complex sentences (e.g., dependent clauses) are not always understood
Syntactic complexity	May display weakness or deficiency due to less secure knowledge of grammar and syntax
Unfavorable listening conditions	Needs the core of the message to be audible. Authentic noise should not impede comprehension; may occasionally understand words and phrases produced in unfavorable conditions; cannot sustain comprehension when core elements are obscured by unfavorable conditions.
Visual support	More of a redundancy in terms of meaning at level; can aid in comprehension of tasks at a higher level

Salient features for Listening at ILR Level 3

Tasks	Accuracy	Context	Content Areas
<p>Understand readily spoken language on concrete and abstract topics related to general, professional contexts and specialized fields of interest or expertise</p> <p>Can monitor and derive meaning from discussions of abstract issues, argumentation, professional transactions and hypotheses in extended discourse</p> <p>Can detect emotional overtones, implications</p> <p>Understand linguistically communicated inferences that do not depend on extra linguistic competence</p> <p>Understand commonly used slang, idiomatic expressions and cultural references</p> <p>Understand a variety of speech acts that express requests, apologies, and suggestions within professional contexts</p>	<p>Able to fully understand message and intention of the speaker</p> <p>Able to fully understand the essential details of the discourse; follow accurately the essentials of conversations between educated native speakers</p> <p>Breadth and depth of listening competence (vocabulary, grammar, syntax, etc.) is sufficient to bridge gaps in comprehension of unfamiliar vocabulary, structure, tone, etc.</p>	<p>Professional and social contexts encountered living and working in the target environment</p> <p>Listening proficiency is sustained in both participative and non-participative contexts</p>	<p>Professional topics; range of topics of general, social, and special interest</p> <p>Interviews, webcasts, new reports, editorials</p>

Factors that may influence Listening at ILR Level 3

Factors that influence Listening at L3	Comment
Accent, regional pronunciations	May understand most common regional pronunciations
Automaticity	Non-systematic; present for familiar topics but not in novel contexts
Clarity	Normal clarity and reasonably clear
Coherence	Can follow only some unpredictable turns of thought
Complexity	Increased ability to comprehend unusually complex structures
Directness of text	Can detect emotional overtones, implications, supported opinions and hypothesis but may miss some subtleties
Length of passage	Research required
Multiple speakers	Able to understand even speakers who may talk over each other
Rate of speech	Normal speech but may not understand native speakers if they speak quickly
Redundancy	May be required for cultural purposes
Rhetorical structures	Will miss distinctions in language tailored to different audiences
Unfavorable listening conditions	Able to tolerate and compensate for limited distortions in predictable situations; can deal with gaps that relate to main content because of the overall strength of their language
Vocabulary	May have difficulty with culturally specific vocabulary, slang, colloquialisms

Salient features for Listening at ILR Levels 4/5

Tasks	Accuracy	Context	Content Areas
<p>Understand all forms of speech</p> <p>Able to analyze the functional intent of messages</p> <p>Understand tailored speech, register, tone</p> <p>Able to interpret most cultural cues precisely</p> <p>Understand “beyond the lines”</p> <p>Understand language that is deeply embedded in culture in which all the syntax, vocabulary, grammar, literature, and history emerge automatically</p> <p>Able to distinguish between own perspective and speaker’s perspective</p> <p>Automaticity extends to implication and deception</p> <p>Comprehend native speakers using a non-standard dialect</p> <p>At L5 understand highly colloquial speech</p>	<p>“First time –final”</p> <p>Able to process non-straightforward discourse</p> <p>Able to anticipate and fill in culturally; understands omissions, code-mixing and code-switching</p> <p>Can handle discontinuity, fragmented coherence, interrupted line of thought</p> <p>Understands fine discrimination of phonological features; parsing variable syntax, honorifics, power and politeness; idiomatic collocations; implications</p>	<p>All of the contexts of a highly educated listener of the language</p> <p>Participative, non-participative, overheard</p>	<p>Wide professional domains</p> <p>Round table discussions; both sides of a debate</p> <p>“Hot topics; ” topics of current, historic and cultural literacy</p> <p>Topics, issues, and concepts typically associated with being well read</p>

Factors that may influence Listening at ILR Levels 4/5

Factors that influence Listening at L4/5	Comment
Accent, regional pronunciations	Only at L5 would one be expected to understand illiterate language, slang
Unfavorable listening conditions	Even the most adverse listening conditions can be overcome

Issues encountered in interpreting and implementing ILR SLDs-L in test development and suggested solutions

A number of issues arose from the Summit discussions regarding the challenges that exist in the usability of the SLDs-L for purposes of the development of assessment. Some of these are:

- The various listening modes (i.e. participative, non-participative, overheard/conversation, overheard/monologue) are not addressed specifically and across levels. If the SLDs–L are going to be useful for test development then expectations of participative, non-participative, and overheard listening need to be articulated for each level.
- Differences between proficiency, performance, achievement are not always clear. Is it possible to test listening proficiency at the lowest levels using real world listening tasks or is testing listening at the lower levels more about achievement than proficiency? Some agencies require information regarding general listening proficiency, other agencies require information about job-specific performance, and even others a combination of both.
- The definitions of terms are not unilaterally accepted. A common glossary along with samples of listening texts that are representative of what a listener at that level can comprehend, and with what degree of accuracy, would serve to address this issue.
- There is no consistent mention in the current SLDs–L of the impact of factors and conditions such as memory load, cultural load, tension and pressure, density of information, length of passage, etc. Which of these factors, if any, should be taken into consideration when creating test specifications?
- The current SLDs-L do not address the types of redundancy that are expected at different levels. What is the effect of repetition of listening items on the test taker? How should this be factored into the creation of test items? Should prompts be repeated? If so, how many times? How many opportunities for listening are necessary?
- Testing at the higher levels, where multiple choice options are not appropriate, presents unique difficulties. There are difficulties in choosing accurate texts for the higher levels.

At the higher levels in some languages, it takes much longer to “get to the point” in an argument, thus requiring a much longer listening passage.

- Is there a hierarchy of tasks for listening articulated in the SLDs-L and, if so, should text level and task level be aligned for testing purposes?

Section 5: Outcomes of the Listening Summits

Guidelines for interpreting the findings from the Listening Summits

A tremendous amount of information, multiple perspectives, questions, and opinions were generated from the Listening Summits. In general, these findings are suggestive and preliminary in nature. It can be said that there may be too much material to consider, some playing more of a role than other. An overarching recommendation for interpreting the Summit findings is to prioritize these findings as to their relevance for the multiple purposes and audiences served by the descriptions.

The findings do suggest that the existing descriptions can continue to be used in their current state to describe listening comprehension according to the ILR provided that there is a consensus regarding the definition of terms, qualifiers, and context/content domains. It is more appropriate, however, that the findings serve as a basis for decision making if and when the SLDs-L undergo revision. The first step in the revision process would then be the co-selection of features, factors and other conditions to be included followed by the alignment of what should be in the SLD's with the reality of user needs and applications.

The findings can also be interpreted in terms of their relevance for test design specifications. There is clearly a need for a common metric for test development. If the SLDs-L are to be useful for test development, there should be specific references to the expectations for participative, non-participative, and overheard listening tasks at each level. The descriptions need to take into account the difficulties encountered when testing at the higher levels where passages can be dense and lengthy, as well as the challenges of recreating real-world listening tasks at the lower levels. In addition, the descriptions must make no claims about processing information, i.e., the processes that influence listening comprehension, but rather only describe the output, the performance of listening tasks.

Recommendations for the revision of the SLDs–L

The revision of the ILR SLDs-L was not a specific goal of the Listening Summits. However, a number of Summit participants were clearly in favor of such an initiative. This section deals with suggestions and recommendations to consider should such a revision be undertaken. It must be stated first and foremost that a revision of the SLDs-L does not mean the revamping of the ILR scale. This does mean providing more of a context for listening and offering greater specificity, support, structure, and examples to assist users in interpreting and using the SLDs-L.

A plan for revision should begin with the creation of a small number of working groups, each of which is assigned a specific task involved in the revision process. One of these tasks would be to undertake the prioritizing and co-selection of all of the information generated from the Summits with the goal of determining what information is essential and should be included in a revision. For example, reference to the salient features that differentiate one level from another (i.e., the tasks, accuracy expectations, contexts and content areas that are unique to the level) as well as the conditions and factors that influence listening at each level are a critical starting point.

The various modes for listening: participative, non-participative, and overheard must be addressed in the revisions with reference to what listeners are able to comprehend (or not comprehend) depending on the listening mode type. Decisions regarding revisions should reflect the differing listening needs of individual agencies. The factors that influence listening comprehension at each level must also be considered. Examples of listening passages and other samples of speech should accompany statements that are made about listening comprehension. Specificity, support, structure and examples are essential for revised descriptions.

A revision to the Listening Descriptions might also benefit from the removal all references to the “native” listener and the inclusion of references to the “highly proficient, well-educated” listener of the language. At the same time, an effort should be made to define listening for different types of learners and learner profiles. Additional recommendations include the inclusion of cultural references in the descriptions as well as the recommendation that attention be paid to the cross-walking with other skills to ensure that there is alignment of level of difficulty across the four skill areas.

Notwithstanding the number of recommendations for inclusion in a revision of the SLD's-L, attention must be paid to keeping the size of the document manageable, useful, and no longer than what is necessary to get the job done.

Future areas for research

The overarching goal of a research agenda would be to move the ILR scale from one that is experiential to one that is empirical. With that goal in mind, the following research topics were among those most often proposed for further study:

- Research on listening that is language-specific, addressing listening for different populations and different types of learners appeared to be a top priority for Summit participants. The majority of the research data presented at the Listening Summits was based on studies that had been conducted in the field of English as a second language.
- Impact of length, density, and organization of speech on a listener's ability to understand at each level. What is the span of text that highly proficient listeners are able to understand?
- What is the effect of unfavorable or adverse listening conditions on listening comprehension? Should these be considered in test design?
- What is the relationship between proficiency level and memory load?
- What is the relationship between proficiency level and cultural load? (In some languages, one does not get to the point quickly in supporting an argument. How long can a listener be expected to follow an argument in a language where this is the norm?)
- Can the research document that there is a hierarchy of listening tasks (as there is a hierarchy of speaking)? Previous researchers have not taken fully into account a hierarchy of tasks wherein each level describes a full standard leading to the design of test specifications aligned to the standard. Can this be validated?
- What can the research tell us about the best practices for designing test specifications? What are the practical constraints when developing listening tests? What kinds of test items are most appropriate at the lower levels where is it almost impossible to recreate real-world listening tasks?
- How to test culture as a component of listening comprehension?