

## The Impact of Culture on Listening Assessment

Jonnie Hill

**Abstract:** Culture permeates our understanding of the world and things around us, affecting our psyche, our cognition, and our social interaction. To what degree does it affect our comprehension of utterances in a language other than our mother tongue? To what degree is the assessment of that comprehension shaped by our cultural worldview? This study synthesizes and pulls together research from Geert Hofstede's (2024) dimensions of culture, TED Talks about communication and listening, and studies in assessment to propose a method of instruction and qualitative assessment then applies it for a descriptive analysis of seven students' performance on a listening test. Implications are considered that would boost English language listening proficiency among Hungarian students in SZTE's Institute of English and American Studies.

**Keywords:** assessment, culture, evidence-centered design, motivation to achievement, uncertainty avoidance

Jonnie Hill is an American instructor in the English Department at the University of Szeged, Hungary. She is an applied linguist specializing in language learning, particularly listening. She has published a number of articles about the use of games in language teaching and was co-guest editor of the 40th anniversary issue of the Sage interdisciplinary journal, *Simulation and Gaming*.

jiaje@yahoo.com

# The Impact of Culture on Listening Assessment

Jonnie Hill

## 1. Introduction

Two years ago, the English Department at the University of Szeged undertook a revision of the Academic English Exam that is given to all English majors at the end of their second or third semester. It was piloted in secondary schools before being implemented at the end of the fall semester last year. It was not apparent to all teachers why students' listening comprehension skills had to be tested, but since it was, it was necessary to incorporate listening assessments as part of the curriculum for the courses about communication skills and listening and speaking. These assessments each contained ten fill in the blank with two word phrases and ten short answer questions, just as the practice test appeared that the students could access from online. After preparing her students for the listening assessment so thoroughly, an American teacher was surprised by the number of students from her class who had to resit the listening exam. What was most disconcerting was that many of those resitting the test had made the top scores on the ten listening quizzes in the course. What could have caused such dismal failure on a test that should have posed no difficulty for them?

## 2. Review of Literature

In listening, the most important skill is paying attention. Once a listener stops paying attention, the hope of meaningful communication is lost. Who is responsible for inspiring that attention? In proposing a new typology in contrastive rhetoric, John Hinds (2001) has suggested that in communicating in English, the onus is on the speakers to make themselves understood (65). Does this mean that it is also the responsibility of the speaker to capture and maintain the attention of the listener? In their Metacognitive Awareness Listening Questionnaire, Larry Vandergrift and Marzieh Tafaghodtari (2010) imply that the responsibility of paying attention and monitoring that attention belongs to the listener. Hinds (2001) stresses that his speaker-responsibility is a tendency; perhaps in the dance of intercultural listening, and particularly in assessment, good listeners need to be given the tools and training to be attentive in every situation. That way, when the speaker is incredibly boring, the student will have the strategies and skills to listen. It would then behoove intercultural listeners and their coaches to identify sources of distraction and devise ways to minimize them. They might do this through building both cultural and linguistic knowledge.

### 2.1. Building Cultural Knowledge

While there are many studies on contrastive rhetoric that consider the ways that culture influences comprehension, there is only one organization that describes the role of culture in comprehension: The American Council for the Teaching of Foreign Languages. The rubric for ACTFL's (2020) Listening Proficiency Test describes four levels of cultural understanding. At the novice level, listeners are familiar with cultural practices. At the middle levels, intermediate and advanced, listeners become aware of cultural patterns, knowing a few as an intermediate listener and acquiring a knowledge of most as an advanced listener. Superior listeners are sensitive to the aesthetic properties of cultural references.

Practices are easily envisioned. A plethora of lessons explore the holidays and daily life of the cultures that we study, but cultural patterns are rarely studied and require some defining. When one seeks a quantitative comparison of cultures, Geert Jan Hofstede's (2024) Country Comparison Tool has no equal. He began his research in the 60's of how culture is manifested in business, but as the "most comprehensive studies of culture values on leadership" (Weinland 2023, 1.3), it provides a useful means of exploring the effect of culture on educational leadership. It can also provide a

framework that can be adapted to explore the different approaches and values that American and Hungarian cultures take in assessing listening comprehension.

Hofstede (2024) identifies six dimensions of culture: power distance, individualism, motivation toward achievement, uncertainty avoidance, long-term orientation, and indulgence. As shown in Figure 2.1, in two of these dimensions, power distance and long-term orientation, the United States and Hungary are very similar. In terms of power distance, both cultures are moderately egalitarian, which means that power is something that is earned and not taken as a given. Both cultures are moderately short-term in orientation, which means they tend to focus on the present or the future rather than the traditions of the past. They also focus on relationships more than tasks (Weinland 2023, 1.3).

In terms of listening, Heather Greenhalgh-Spencer (2022) demonstrates how the dynamics of power and relationships affect attention. She advocates deep listening in order to build relationships. For this to occur, one must have a desire to listen rather than to speak, must care for the speaker, and, if in a position of power, such as in teacher to student interactions, must seek to create a space in which the power dynamic does not discourage the student from speaking. In short, create a space in which interlocutors are equal.

In two other dimensions, Hungary and the United States fall on the same side of the scale, but whereas the United States is moderately individualistic and motivated toward achievement, Hungary is strongly so. The scores in individualism indicate that while an American likes to see oneself as being able ‘to pull myself up by my bootstraps,’ Hungarians are more likely to do so. It also means that Americans are more likely to exhibit more teamlike behaviors and to consider what is best for the group. The opposite sides of the spectrum in motivation toward achievement are decisive and consensus (Hofstede 2024). Decisive cultures value competition and assertiveness; consensus cultures value cooperation and modesty (Weinland 2023). What this seems to indicate is that while Americans certainly understand competition and assertiveness, they temper it with a caring, nurturing nature.



Figure 1. Screenshot of Hofstede’s (2024) Country Comparison Tool

According to Hofstede’s (2024) Country Comparison Tool, Hungarians and Americans are at opposite ends of the spectrum in two dimensions: Uncertainty Avoidance and Indulgence. Whereas Hungarian culture strongly values structure and certainty, American culture values flexibility and tolerates a

certain level of ambiguity (Weinland, 2023). Whereas Hungarian culture values restraint in the gratification of impulses, American culture is more open to play and having fun (Hofstede, 2024).

Table 1 explores the differences of Hungarian and American culture in these six dimensions, listing them in the order of most alike to most different. Hypothetically, those that have the smallest gap for business cultures, namely time and power, would indicate similar behaviors and values in listening and similar attitudes toward testing. Conversely, the dimensions in which Hungary and the US are on opposite sides of the spectrum, namely uncertainty avoidance and indulgence may spark the greatest conflict. As barriers to listening, these contrastive patterns of culture present both a challenge and asset for Hungarian listeners. On the one hand, their propensity toward not taking risks could be a barrier. Listening takes courage in the sense that there is no guarantee that one can be successful in building relationships and deeply understanding the one speaking to you (Greenhalgh-Spencer, 2022). On the other hand, their ability to delay gratification and their leaning toward restraint allows them to divest themselves of their desire to be heard, which creates a better environment for deep listening.

Dimension	Descriptor, < 50	Descriptor, > 50	Gap
Time	short-term/ present	long-term / past	-5; both moderately short, US longer
Power	egalitarian	hierarchal	6; both egalitarian, Hungary less
Identity	collectivist	Individualist	11; both individualistic, Hungary more
Achievement & Success	Cooperative Consensus	Competitive decisive	26; both competitive, Hungary stronger
Uncertainty avoidance	Tolerates ambiguity	Desires structure	36; Hungary structured, US ambiguity tolerant
Indulgence	restrained	Indulgent	-37; Hungary restrained, US, indulgent

Table 1. Hofstede (2024) Cultural Comparisons of Hungary and the US

To level up in cultural knowledge after becoming familiar with most of its patterns, ACTFL (2020) proposes that listeners must be familiar with the aesthetics. They need to learn to appreciate how these cultural dimensions are promoted and tweaked for efficiency and excellence. For example, Americans tend to prioritize tasks over relationships, but a number of TED Talks featuring Americans highlight the importance of understanding speakers rather than messages. These can be mapped onto three of Hofstede’s (2024) dimensions of culture to reveal three core values in listening to understand: time (long-term relationships, not short-term tasks), identity (focus on speaker as a contributor to the group, not the message as edification of the individual), and motivation for achievement and success (consensus, not competition).

In terms of time orientation, long-term cultures focus on relationships while short-term cultures such as the US and Hungary focus on the task at present. The task of comprehending what is said is undoubtedly important, but it is not enough. In her TED Talk about winning arguments, Lauren Schifferdecker (2017) illustrates how in an argument with her partner about the need to replace a phone, she lost sight of what was really important. When listening to understand, it is the relationships that gain precedence over the task of being correctly comprehending. Amber Wright (2020) also emphasizes the importance of relationships in her TED talk and illustrates through several anecdotes how asking better questions led to better understanding. Celeste Headlee (2016)

also suggests that it is not the details of the speech or one's power of persuasion that listener's are concerned about, but the speaker. Although the American and Hungarian cultures' short-term orientation may favor task-focused listening, the existence of Headlee, Schifferdecker and Wright's talks bears witness to the Americans' assent that one should develop long-term listening skills for understanding in order to preserve their relationships.

Regarding identity, listening to comprehend fosters an individualistic attitude whereas listening to understand focuses on thinking of what is best for the group. Sumner (2019), Schifferdecker (2017), and Headlee (2016) all mention the importance of stepping outside yourself and your predetermined beliefs to understand what the speaker is saying. In listening to understand, the primary value is the speakers themselves and their contribution to the knowledge and psyche of their listeners.

In the dichotomy of competitive listening versus cooperative listening, Dave Sumner (2019) highlights the importance of finding new metaphors to describe argument. He suggests that rather than thinking of argument as a fight in which one wins and the other loses, one must think of it as a dance. In so doing, one listens to understand why one takes a particular stance rather than being offended and turned off because it runs contrary to one's belief. Seen in this light, listening to understand is a cooperative activity. Although listening to comprehend is not the same as Sumner's contrast of the fight, because comprehension is more easily quantifiable, it is more likely to foster competition. As with long-term orientation, the group mindset of listening to understand is counterintuitive to competitive cultures, but, as can be understood from Dave Sumner's Ted Talk, it would be of great benefit to the listeners to focus on the depths of understanding rather the superficial comprehension attained when applying one's own misguided filters and prejudices.

When considering the side of the spectrum on which American and Hungarian cultures fall regarding the issue of listening to understand the speaker and listening to comprehend the message: Hungarians and Americans have the same tendency toward short-term orientation, individualist thinking, and a competitive approach to success and achievement. All of these run counter to the skills that must be cultivated for listening to understand. However, the selected TED talks suggest that a change in one's listening approach to focus on long-term, others-oriented, consensus-building listening skills would not only benefit the students in the classroom today, but also in their workplaces and their communities in the future. When speakers feel like they have been understood beyond the decoding of their words, they can trust their listeners to do more important things as well.

## 2.2. Building Linguistic Knowledge

Another possible hurdle to comprehending speakers is the listener's linguistic prowess. In listening, one must be able to decode the words and the grammar to comprehend what is said, but the more difficult task, which listeners encounter even when listening to their own language, is listening to understand the intent and motives of the speaker. ACTFL's LPT rubric describes various levels of linguistic knowledge as well. At the novice level, listeners are most successful with formulaic and predictable language and texts that are no longer than a few words. At the intermediate stage, listeners recognize simple sentence structures with high frequency vocabulary. At the advanced level, listeners can handle longer texts with simple functions (e.g. description, narration, explanation), broad vocabulary, and time or tense. At the superior level, text types are more complex (argumentation, hypothesis, vocabulary more specialized, and grammatical structures more complex).

The challenge of assessing listening for understanding is doing it in a way that honors its core values of preserving relationships, focusing on others, and consensus. Traditionally, listening instruction, particularly in English language learning, has focused on the task of comprehending what is said: listening for the gist, listening for details, listening to determine the purpose, and recognizing vocabulary. Other metacognitive strategies for listening have been introduced by Vandergrift and Tafaghodtari (2010, 470) which include focusing, monitoring attention, predicting, and problem solving. While these are helpful when describing discrete elements of the "listening skill," they only scratch the surface of listening to understand. In order to gain a deeper understanding of the listeners'

ability to understand, one must adopt qualitative assessment measures that requires a high degree of tolerance for ambiguity, a desire to nurture growth, and sneaky playfulness.

At its core, the approach to quantify listening assessment comprehension is flawed. The question, “How much did you understand?” is misleading. One could conceivably understand all the words, but misinterpret the message. This is illustrated rather humorously by Peggy Parish’s (1963) character, Amelia Bedelia, who, in responding to the request to “Draw the drapes,” finds pencils and paper and draws a picture. She has understood 100% of the vocabulary, but she has not interpreted perfectly that her employer wanted her to close the curtains to protect the furniture. A better question concerning Amelia Bedelia’s understanding, as well as that of any listener, native or foreign, is “What did you understand?” (Hill 2000). In doing so, language coaches can determine whether it is a contrasting linguistic, textual, or cultural context that leads to alternative interpretations.

This, of course, requires a great deal of flexibility and tolerance from the assessor, and the acknowledgement from assessors that they, too, are fallible. The lack of a structured black and white answer key would be extremely uncomfortable for cultures that avoid uncertainty. A descriptive approach to listening assessment is fraught with challenges. Considering multiple interpretations requires a moderate level of tolerance for ambiguity. Drawing drapes in art class is a different kind of drawing than a housekeeper would perform, but which would be asked of an intern in interior design? For a test, devoid of all the contexts, one would have to consider both interpretations correct. However, much could be learned from a descriptive approach to assessment. A rich description of the listener’s understanding could flag details and grand topics, the linchpins of comprehension assessment, as well as responses to the speakers’ viewpoints, attitudes, and implicature. By analyzing the listener’s responses, the assessor can pinpoint nuances and concepts that the listener identified that are beyond the scope of the answer key.

A descriptive approach to assessment forms a foundation for fostering a stronger growth mindset in the listeners. The description of what they heard compared to the descriptions of what their peers heard can help listeners and their assessors create an action plan for cultivating better listening habits. The analyses would give listeners and assessors insight concerning the listener’s interpretations and to discuss how those interpretations could be better aligned with the speaker’s intention.

The role of the assessor as a nurturer is embodied in the Black and Wiliam’s (1998) approach of ‘assessment for learning.’ In their study of assessment practices in England, Wiliam et al. (2010) found that the students of teachers who used the data from formal and informal assessments to inform their instructional content performed better than those whose teachers simply used assessment to score a performance. One might infer that an approach to assessment for learning to listen in which the instructors see themselves as listening coaches could have the same impact on student listeners in the Institute of English and American Studies at the University of Szeged.

In three dimensions of culture–ambiguity tolerance, consensual motivation to achievement and success, and indulgence, Hungarians in their propensity for structure, competition, and restraint are most different from Americans. Thus the traditional approach to listening assessment would require some long, possibly unheeded, conversations to affect any change. However, within the classroom, where non-Hungarian instructors have the freedom to design a curriculum that fosters coaching and to devise stealthy tests to help listeners advance toward appreciating the aesthetics of cultural differences and comprehend texts that are linguistically complex, a different culture of assessment may help students prepare more comfortably for the harsh realities of the AE Exam.

### 2.3. Guiding Principles for Exploring Intercultural Listening

Three principles underpin this exploratory analysis of students’ processing of spoken texts.

#### 2.3.1 *Attention is key.*

Although Americans strongly appreciate speaker-responsible communication in which every effort is made by the speaker to gain and maintain the attention of the listeners, listeners need to make the

effort to give that attention. In intercultural listening, when the text complexity is higher than the linguistic abilities of the listener, it is of particular importance.

2.3.2 *Understanding is not about how much of a text the listener understands, but about what is understood.*

Listeners process texts according to their stages of linguistic development and cultural awareness. Thus, a descriptive approach to assessing listeners’ understanding would provide richer data in understanding the intercultural listening process than a quantitative approach.

2.3.3 *Cultural Understanding necessitates the knowledge of practices and patterns reflected in the speech.*

A different world view concerning power dynamics, courage, and the priority of relationships influences the way that listeners process texts. Listeners’ ability to recognize and appreciate these even when they are left unsaid, will influence the level of processing they can do of the speech.

3. Methods

To understand the influence of cultural and linguistic knowledge on students’ listening comprehension, this study analyzed the responses of seven students to questions about two different videos. For the first video (Lanz, 2017), a five-minute exposition on focus groups by TED, students were presented with ten questions in a Google Form in which they were to fill in ten gaps with two words. For the second video, a debate over the role of slang in the classroom sponsored by *The Guardian* (2013), students were presented with eight open-ended questions. The responses of seven students were pulled to use in this content analysis. Four students, Yi, Er, San, and Si, not their real names, made the four highest scores in the class. Three students, Wu, Liu, and Qi, also not their real names, made the lowest scores in the class. Figure 2 maps the result of these tests. The gray boxes indicate responses that do not match the answer key and were not awarded any marks. The green boxes indicate responses that do not match the answer key, but were awarded marks.

	Yi	Er	San	Si	Wu	Liu	Qi
1-1							
1-2							
1-3							
1-4							
1-5							
1-6							
1-7							
1-8							
1-9							
1-10							
	Yi	Er	San	Si	Wu	Liu	Qi
2-1							
2-2							
2-3							
2-4							
2-5							
2-6							
2-7							
2-8							

Figure 2. Map of students’ responses



To determine the difficulty level and pinpoint some possible cognitive challenges, transcripts of the videos were assessed using online tools such as the Readability Scoring System (readabilityformulas.com) and the Text Checker (Oxford). The Fry readability formula offers a way to describe the complexity of a text as it correlates the average number of sentences and syllables in 100 words, then estimates a grade level from 1 (easiest) to 15 (most difficult). Another readability scoring system, which is also useful in estimating comprehensibility, is that of lexical density and diversity. Lexical density is the proportion of content words in a text. Lexical diversity is the number of different words in a text. Less complex texts have lower scores in lexical density and lexical diversity. Oxford’s Text Checker color codes the words in the text according to the CEFR for the 5000 most common words in the English language. This tool is especially helpful in predicting the specific words that students may not recognize; it also flags the proper nouns, which is especially helpful in determining cultural artifacts that appear in the text. Table 1 summarizes the results of these tests. Hector Lanz narrated the video about focus groups. Lindsay Johns and Michael Rosen were the debaters in the second video. Table 2 summarizes the results of these tests.

	Hector Lanz	Lindsay Johns	Michael Rosen
Number of words	637	405	551
Length	4 min, 32 seconds	2 min, 22 sec	2 min, 38 sec
Words per minute	141	171	209
Syllables per 100 words	163	149	134
Sentences per 100 words	5	3	7
Fry Grade Level	14	11	9
Lexical Density	53.6 % (above)	57.3% (above)	48.5% (average)
Lexical Diversity	54.6% (above)	47.7% (average)	38.3% (below)
Oxford 5000 C2*	12	8	1
Not in Oxford 5000*	21	24	13
Proper nouns*	7	7	7

\*The exact words are listed in the analysis of what the students understood.

Table 2. Complexity of Texts

To examine what exactly the students understood from the videos and the depth of their answers, a content analysis was conducted in which the answers of the students and the answer key of the teacher were compared to the transcripts of the videos, paying particular attention to the words that Oxford Text checker coded as C1 and the words that were not coded at all, presumably because they were C2 level. To determine the effect of culture on the students’ comprehension, their responses were scanned to identify the uses of proper nouns.

To determine what students understood on a conceptual level, the students’ responses will be re-scored using schemes suggested by ACTFL (2020, 8). At level 0, where the response is deemed implausible, the student’s response contradicts or is unsupported by the transcript.

Level 1 corresponds to ACTFL novice or CEFR A1 (Goethe 2024). The descriptors at this level suggest that listeners have “a basic awareness of formal and informal practices” of culture and the use of memorized chunks and formulaic phrases (ACTFL 2020, 8). These are the responses in which the students presented a phrase or concept that was presented in the text but perhaps put it with the wrong question. Jennifer Rives (2024) described this level as “general understanding” (16).

Level 2 corresponds to ACTFL intermediate. An intermediate interpretation notices “a few of the most common patterns” in culture and “predictable loosely connected language” (ACTFL 2020, 8) These responses differ from Level 1 in the level of modifiers added to the text. Rives (2024) describes this as “detailed understanding” (16). ACTFL intermediate is equivalent to A2/B1 (Goethe 2024).

At Level 3, ACTFL advanced, some degree of implicature can be observed. According to ACTFL, in terms of culture, students recognize “most of the cultural patterns” and longer passages “with clear



predictable structures” (ACTFL 2020, 8) on topics of general and current interest; “listeners are able to comprehend the facts presented in oral discourse and are often able to recognize speaker-intended inferences” (ACTFL 2020, 20). Responses in which students paraphrase the concepts show that they have processed the text at an advanced or B2/C1 level.

At Level 4, which corresponds to ACTFL’s superior level, the student’s level of implicature is deeper: “Superior listeners can understand not only what is said, but sometimes what is left unsaid; that is, they can make inferences” (ACTFL 2020, 19). In terms of culture, students can understand the “aesthetic properties” (ACTFL 2020, 8) of lengthy texts on both familiar and unfamiliar topics. At this level, students notice subtleties in the message and their responses will reflect critical thinking. The procedure for this content analysis is as follows:

1. Analyze the most difficult prompts, those in which most of the responses diverged from the answer key (e.g. Prompt 1-2 and Question 2-6). Compare the listeners’ responses with the script to determine which part of the script the students used to formulate their answers. Code responses regarding level of understanding.
2. Examine lexical resource and code remaining prompts for level of understanding
3. Analyze prompts with cultural references (e.g. Prompt 1-6)
4. Interpret the coding to describe the performance of the listeners, comparing the descriptive results to the numerical results.

Whereas it is too late at this point to give formative feedback, the insights gleaned from this analysis can shape the curriculum and provide help to future students at a time when feedback can aid learning.

#### 4. Results

The main task in providing a descriptive analysis of the students’ responses was to code them using the rubric adapted according to ACTFL guidelines. The most difficult prompts were coded first, then those only related to vocabulary, and, finally, those that shed light on cultural patterns.

##### 4.1. Coding Difficult Prompts

The difficulty of prompts was discerned by the number of listeners who gave a response that was marked as incorrect. In the case of Prompt 1-2, however, this does not mean that the students had a Level 0 in understanding. The disagreement with the answer key for Prompt 1-2, “*Rather than asking straightforward questions with quantifiable answers, the researchers (Merton and Lazarsfeld) engaged people in \_\_\_ \_\_\_ discussion*” was that five of the students only gave one word answers, “open,” when they should have given two, “more open.” The answer “open” demonstrates a general understanding of the text but a disregard for the instructions. The open cloze question format does not allow for much implicature, so the best level that one could attain in the prompts from Part 1 is a Level 2, which is shown in Wu’s response: “*more opened.*” Si’s response of “*in-person interviews,*” while it is a part of the context of Prompt 1-2, it does not reflect any sensitivity to the collocation with ‘discussion,’ thus rendering a Level 1 interpretation.

Question 2-6, “*What example does the teacher use to refute the linguists’ point that slang should be studied, not banned?*” is a tricky question because the teacher does not refute the linguist’s point. The first time the linguist suggests studying slang, “Now my view would be no, you study it” (Guardian 2013, 2:34), the teacher responded, “I have a zero tolerance policy with my young mentees down in Peckham and I try and correct the way they speak.” (Guardian, 2013, 2:37) The second time, after the linguist gave an example of how to study the use of slang in Shakespeare, the teacher responded,

I think it's very easy for liberal academics and writers who are often, not always, but often cocooned in a very safe, closeted world, an ivory tower in my opinion, I think, is actually very hypocritical because I think you'll find that the majority, not all, but the majority of those liberal academics and writers– they themselves enjoyed the benefits of a Rolls-Royce

humanities education and I'll wager with no slang whatsoever. I deeply struggle with the notion that, for example, in when, when we hear young people speaking street slang we're hearing the authentic rhythms of Africa. (Guardian 2013, 3:33-4:09)

Only one student, Yi, recognized the teacher's subtle attack on the character of those who did not support banning slang, "*He claims that linguists are hypocrites.*" This response demonstrates an aesthetic appreciation for an underhanded rhetorical technique, showing a Level 4 understanding of the text.

Qi's response does not reflect any subtle understanding, but does include some details. His response "*that slang should be kept outside the classroom*" comes with an example, "*bravo'--zero tolerance policy.*" Thus, this response can be coded as Level 2. Er's response also contains no implicature, but some modification, which raises the answer to a Level 2: "*People in power don't use slang and he wants to help students be empowered.*" Unfortunately, the first part of the response is false, neither debater claimed that people don't use slang.

If one overlooks the indecipherable handwriting and convoluted grammar, Wu's response could be considered plausible at a literal level: "*the teacher cannot his students regarding let to use slangs.*" The idea of the teacher not wanting his student's to use slang is a fairly accurate summary of the teacher's first response to the linguist. However, there is no observation of the subtleties and no details, so this response simply reflects a general understanding of the text. The ungrammaticality of the response also limits it to a Level 2.

The other responses were not plausible and were coded at Level 0. San's response is half true, but implausible: "*That it is hypocritical that writers use slang in their words but they are against it otherwise.*" Neither debater said that writers were against the use of slang. Si gave an incomplete answer: "*the linguists.*" Liu's answer "*empowering the young people, he heard them using African slang.*" While both elements were mentioned in the debate and can be supported by what the teacher said, the way that the listener has phrased the response contradicts the purpose and main thrust of the teacher's argument.

#### 4.2. Coding Encounters with the Vocabulary

Table 3 summarizes the lists of words that Oxford Learner Dictionary deems as C level and the proper nouns and the ways in which they were encountered in Part 1 of the listening test. The numbers in parentheses after the words indicate the question in which the word was encountered. Because of the nature of the test and its questions for Part 1, one can only observe the words that students recognized in the text. Students were not penalized for misspelled words, but correctly spelled words could indicate a deeper familiarity with the words.

	Encountered only in listening	Encountered in Question	Required in response
C1	handy, unprecedented, propaganda, seemingly, gears	straightforward (1-2) substitute (1-10) according (2-1, 2-4)	polls (1-1), reasoning (1-3), contrary (1-5), interference (1-9)
C2	advertisers, sociologist, wartime, coined, qualitative, marketers, unrelated, basics, generalize	quantitative (1-1) quantifiable (1-2) moderator (1-8)	exploratory (1-4), recruit (1-7)

Table 3. Encounters with Advanced Vocabulary in Part 1

Quite by chance, every question included a word from the list. Words from Questions 1-1,1-2,1-6, 1-8, and 1-10 were encountered by students in the text. As can be observed in Figure 3.1, all the students' responses to these questions except 2 agreed with the transcript. In terms of demonstrating their

understanding through the reproduction of the words as they encountered them in various phrases, the students often matched those of the transcript, and the use of two words in the open cloze exercises facilitated the modification needed for a Level 2 interpretation. At times, there were misspellings (e.g. Er and Si typed “*poles*” for 1.1) or typos (e.g. Yi wrote “*controary*” for 1.5 and Si typed “*comoanies recruit*” for 1.7), but there were only four instances—one each for Questions 4, 5, 7, and 9—in which the examiner had to make a judgment.

For 1-4. “*Rather than providing definite conclusions, focus groups can be used for \_\_\_* \_\_\_\_\_,” Hans Lanz said “exploratory research.” Liu wrote, “*expository research*,” demonstrating recognition of the first and last three phonemes, but clearly misinterpreting the middle syllables. Since the two words have very different meanings, this response is not a plausible response, and reflects a Level 0.

For 1-5, “*An example of how focus groups give a deeper understanding of consumer habits, the narrator, Hector Lanz cites the finding that \_\_\_ to \_\_\_ opinion, at the time, wives had more influence over their husbands when choosing what car to buy.* [Note: the answer key includes 'to'],” Liu wrote “- to their,” indicating with a dash that the first word of the series had not been understood. The transcript includes the phrase “contrary to public opinion.” Because the absence of the word “*contrary*” changes the meaning of the sentence and “*their*” is not clear, this response also reflects a Level 0 interpretation.

Yi’s response to 1-7, “*How does the group work? First, \_\_\_\_\_ six to ten participants*” requires a closer look. On the surface, “*researchers gather*” makes perfect sense and might even be considered as reflecting a deeper interpretation. However, in a broader context, one can see that it is not the researchers that are doing the recruitment: “First, companies recruit between six and ten participants according to specific criteria that meet their research objectives (Lanz, 2:25).” Thus, though “*gather*” represents a deeper understanding of the concept of ‘recruit,’ the change of agency renders the interpretation less plausible. However, because it is not contradictory to any of Hector Lanz’ script, it is an example of a Level 1 interpretation.

In responding to 1-9, “*A limitation of focus group interviews is \_\_\_\_\_, the notion that the simple act of observing something can change it.*” Si attempted to use a synonym: “*observer effect.*” Although this is implied in the text and the word ‘affected’ appears, it is not as precise as interference:

But although they can provide valuable insight, focus groups do have their limitations, and one of the main ones is that the simple act of observing something can change it. This principle is called observer interference. The answers participants give are likely to be affected by the presence of the researchers, social pressure from the rest of the group, or simply knowing that they’re taking part in a focus group. (Lanz, 3:35-3:53)

In this case, though there is a plausible general understanding of the text, a detailed understanding of ‘interference’ is not apparent. This is an example of a Level 1 interpretation.

#### 4.3. Coding Lexical Resource for Part 2

Table 4.3.1 facilitates an analysis of the students’ lexical resource in responding to the questions about the debate sponsored by *The Guardian*. In this part of the listening test, students responded to open-ended questions, so they may have chosen the word that they heard, or they may have expressed the idea using other vocabulary. The numbers under the students’ moniker designate the question to which students used the word in the response. As a whole, this sample used from five to seven advanced lexical items.

As seen in Table 4, the questions that elicited the most C-level vocabulary from the students were 2-1 (all seven students) and 2-8 (all seven students). Questions 2-3 (four students), 2-4 (four students), and 2-5 (five students) inspired a moderate use of C-level vocabulary. The most salient words were *slang*, *uneducated*, and *linguist*.

	Encountered in listening	Used in Response	Yi	Er	San	Si	Wu	Liu	Qi
C1	Mentor, advocate, mobility, liberal, whatsoever, empower, prejudice, tolerance	Empower	3	3		3		6	3
		Prejudice		4					
		Tolerance							6
		cease					8		
C2	Exceedingly, frontal, anathema, lingua franca, mentees, tangibly, closeted, ivory tower, humanities, cocooned, wager, relativism, reductive, bidialectal, desirability, pronunciation, Uneducated, lobotomy, slang, code-switching, hypocritical, dialect	Uneducated	1	1	1	1	1	1	2
		Lobotomy							1
		Slang	5	2,4,5,6,7	3,4,5,6,7,8	5	4,5,6,7	4,6	5,6,7
		Hypocritical	6		6				
		Code-switching			2				
		Persuasive*					8	8	
		Dialect		7		7,8			
		Linguist*	8	8	8	8	8	8	8
		Refute*						8	8
		Biased	8		8				
		Pros & cons			8				
		Unquestionably				8			
		Advanced usage	6	6	7	6	5	6	8

Table 4. Lexical Resource in Part 2

Question 2-1, “According to the young teacher, how does slang make his students sound? “required students to listen for a detail which included a C2 word, uneducated. All but Qi wrote, “*stupid and uneducated*” going beyond a one-word answer. Qi chose a different line of text that also showed a Level 2 understanding: “*as if they had a painful lobotomy.*”

Question 2-2, may not have elicited C-level vocabulary, but it did, in a few cases, allow for a greater range of implicature, In response to “*For what two reasons does the teacher discourage his students from using slang?*” several students produced at least one reason that showed Level 3 processing. For San, both reasons were at a Level 3: “*it makes a bad image and could prevent opportunities.*” Issues in collocation and the brevity of the response limits it to Level 3, but San has captured the essence of the teacher’s reasoning. For Yi, the second reason, that their “*application may be dismissed*” and for Er, the idea of “*people in power not using slang*” show some paraphrase of a thought that the speaker definitely implied. The most frequent answer, “*to be taken more seriously*” is directly taken from the text. It is long enough and contains the modifier ‘more’ so it can be considered Level 2. Qi and Liu were the only one’s to use C-level vocabulary. Qi used the word that the others had used in Question 2.1t: “*they sound stupid and uneducated.*” Answering in this way, it still shows Level 2 understanding; however, it only counts as one reason. Liu used the word code-switching, picking up on the teacher’s utterance that he was “*not a fan of code-switching.*” This is only loosely connected to the prompt, so it shows a general understanding. Si’s answer that “*language is power*” is also picked from the text and shows a Level 1 interpretation in its loose connection to the question. Wu’s response “*to make their life’s better with decisions*” picks up a part of the text, “I want the young people that I mentor to be taken seriously by those who have the power to take decisions which can affect their lives for the better or for worse” (Guardian 2013, 0:26) but the agency is wrong. The response does not show that it is the people in power making the decisions, so it is a Level 0 for understanding.

Question 2-3, “On what point does the older linguist agree with the young teacher?” also required students to listen for information. It was clearly stated, “I agree with some of that from the point of view that we want young people to be empowered” (Guardian 2013, 0:45). Empowered is a C1 word that Yi, Er, Si and Qi quoted to produce a solid Level 2 answer. The remaining responses show a Level 0 interpretation of the text. San’s response that “*the students slang is one type of slang*” and Liu’s response concerning the “*possibility of learning standard English,*” were suggestions from the linguist,

but not a concept that both debaters agreed on. Wu's response "*for the people to be expressed*" makes no sense and is not in the text of either debater.

Question 2-4, "*According to the linguist, what is the key issue that one should consider about the students who use street slang?*" also required listeners to isolate one piece of information that was clearly marked: "so the key issue is, why don't your mates in Peckham choose to speak Standard English or maybe they can and– or maybe they know how to– and choose not to" (Guardian 2013, 1:30) Yi, Si, and Liu isolated some version of the utterance and presented a plausible, but literal answer. Er's response, a Level 0, was incorrectly attributed: "*They will use street slang and be prejudiced in important cases like job interviews and college interviews*" was one of the central points of the teacher. Some of the students complained that it was difficult to answer some of the questions because in having two male voices, they lost track of who was saying what. San kept up with the right speaker, but the response does not reflect the purpose of the speaker in any way, "*that students' slang is one type of slang and they way they use their dialect makes them use slang to an extent.*" This is a Level 1 example of choosing a phrase from the text, but putting it with the wrong question. Likewise, Wu's answer, "*that people use slang everywhere but a different type of slang*" demonstrates a general understanding of the text. Qi's response picks up a significant proportion of the text, but missing the central idea of why: "*we should consider that they can learn/know standard english, too.*" It, too, is a Level 1 response.

#### 4.4. Coding Cultural References

Prompt 1-6 and Question 2-5 challenged students to process concrete information about cultural references: one to an American toy and the other to British literature. The most obvious way of finding cultural reference is by observing how students process proper nouns. However, more subtle inferences concerning the student's process can be observed in how they dealt with the text types and, as ACTFL points out, their observation of the cultural patterns. In this regard, the students' ability to deal with the rhetorical structure of argumentative texts can also provide insight into the role of culture in listening comprehension.

Responses to 1-6 give a glimpse into the students' understanding of culture. The names encountered in the text were always accompanied by an introduction of who the person was and what he had done. As a neighbor, Austria should be well recognized by the Hungarian students. One could guess from the context that Chrysler was a car company and Mattel a toy company: "And Dr. Dichter himself conducted focus groups for Mattel to learn what girls wanted in a doll. The result was the original Barbie doll" (Lanz 2018, 2:05-2:11)." In the quiz, students encountered "*The result of a focus group interview conducted by Dr. Dichter himself for Mattel was the \_\_\_\_\_.* [Note: the first word is a proper noun and should be capitalized]." Five students wrote, "*Barbie doll*"; San and Wu gave a more detailed interpretation, "*Original Barbie.*" Both answers reflect a Level 2 interpretation. Thus, in respect to the Barbie doll, cultural encounters did not challenge the students' comprehension.

Question 2-5 elicited the most frequent references to English literature: "*Which example does the linguist use to refute the argument that educated people do not use street slang?*" In his third turn, the linguist begins, "If you go to Shakespeare, you'll find that Shakespeare uses the word, 'cuz" (Guardian 2013, 2:53-2:54). Yi, Er, and San all stated this. Yi and San also added the detail that it was street slang. Si reiterates the linguist's purpose by prefacing the response with "*We should studie slang words instead of banning it,*" before writing "*for example, the word 'cause in Romeo and Juliet.*" Liu added the inference, "*studying its history is interesting.*" This extra information provides a glimmer of a Level 3 interpretation.

Questions 2-7 and 2-8 explore the students' subtle understanding of culture: the organization of the text. Indeed, most of the questions in Part 2 of the listening text challenged students to note the patterns in the text, which, as we saw in 2-6, was not always speaker-responsible. Only two students, Wu and Qi, did not demonstrate a Level 2 understanding for Question 2-7, "*What is the linguist's primary claim about banning slang?*" Unlike many of the other questions in which listeners had to point out a specific reference that had markers like "*key issue*" or "*agree*" from the text in the prompt, Question 2-7 required listeners to think of the whole text and how all of the linguist's arguments fit together. The linguist's concluding statement was, "The obvious survival of dialect and local speech is

evidence that [banning] has never worked” (Guardian 2013, 4:56-5:02) Many of the responses show some appropriate paraphrase of this statement: “*banning will not solve the issue*” (Yi), “*slang can be in speech or just a dialectic thing, banning it is not effective*” (Er), and “*it had been attempted to ban slang, local speech even, but it has never worked before*” (San). Liu’s response began as a Level 3, “*it could never work because it will survive,*” but added a reference that was in the text, but only loosely associated, “*Newspapers would speak different than conversations.*” Si recycled the vocabulary, “*We cannot ban a dialect*” and showed a Level 2 interpretation.

The last question of Part 2 was not focused on a particular part of the script, but called upon students to synthesize the entire debate: “*Who presented the most persuasive arguments in this debate?*” A perceptive answer might have drawn upon other things learned about persuasion and argument, including a TED Talk by Dave Sumner (2019) that students had been assigned for homework for weeks before the exam. The responses have been sorted according to the descriptors for cultural understanding in the ACTFL Listening Proficiency Rubric (2020, 8).

At the novice level, the listener has a “basic awareness of informal and formal practices.” (ACTFL 2020, 8). In regard to argumentative rhetoric, the idea of backing up arguments is a basic practice. Thus, Er’s response, “*He could always back up his arguments*” is Level 1.

An intermediate listener recognizes “a few of the most common cultural patterns” (ACTFL 2020, 8). Patterns in argument would include refuting the opponent’s claims. Liu’s response is a weak Level 2, but one must appreciate the attempt to add details about what makes someone sound persuasive: “*The teacher sounded more persuasive. He brought out more examples that he experienced and always tried to refute the linguist’s claims.*”

Wu’s response is also a little weak regarding patterns, but also provides details about the type of evidence: “*think the linguist’s evidence for persuasion is more important regarding to use slang in our daily life because it will be everywhere no matter the actions to cease to exist.*”

An advanced listener recognizes “most of the patterns” (ACTFL, 2020, 8) Si’s response shows a knowledge of the patterns of explaining, giving examples, and proving arguments and provides examples of these: “*At the end when he explained that a dialect cannot be banned and posed an example, which is already unquestionably proven, it shows that stronger forces had already tried what the teacher is trying to do, and they failed.*” Si’s response is a Level 3 interpretation.

A superior listener recognizes “Cultural references and aesthetic properties.” One of the most important yet subtle aspects of debate is the debaters ability to listen and respond to each other. Yi’s response is a Level 4 because not only did the interpretation point out the subtleties of listening, but it also provided a contrast between the most persuasive and the least persuasive debaters: “*The linguist’s approach is calmer, less biased. The teacher is not open to the linguist’s opinion.*” Qi also points out the subtleties of being open to listen to the other side: “*he obviously more fearless with the topic but he is still open to hear the teacher’s opinion and as is able to refute them or agree with them.*” San did not pick up on the aspects of listening, but did appreciate the nuances of the linguist’s argument: “*The linguist was more persuasive because he sees and explains both sides of the topic, he sees the wider range of language and its usage. He talks about the pros and contras of a slang and explains it in education, clearly without being biased*” (San).

#### 4.5. Describing Performance

Figure 3 maps the performance of the listeners as coded. The students are ordered according to the scores awarded on Part 2 of the final exam: Yi 100, Er 80, San 80, Si 80, Wu 50, Liu 50, and Qi 40. When compared to the map of the level of understanding demonstrated, a few observations can be made.

Part 1							
	Yi	Er	San	Si	Wu	Liu	Qi
1-1	2	2	2	2	2	2	2
1-2	1	1	1	1	2	1	1
1-3	2	2	2	2	2	2	2
1-4	2	2	2	2	2	0	2
1-5	2	2	2	2	2	0	2
1-6	2	2	2	2	2	2	2
1-7	1	2	2	2	2	2	2
1-8	2	2	2	2	2	2	2
1-9	2	2	2	1	2	2	2
1-10	2	2	2	2	2	2	2
Part 2							
	Yi	Er	San	Si	Wu	Liu	Qi
2-1	2	2	2	2	2	2	2
2-2	2	2	3	2	2	2	1
	3	3	3	1	0	1	0
2-3	2	2	0	2	0	0	2
2-4	2	0	1	2	1	2	1
2.5	2	2	2	3	1	2	1
	2	2	2	2	1	3	1
2.6	4	2	0	0	2	1	2
2-7	4	4	4	2	1	2	1
2-8	4	1	4	3	2	2	4

Figure 3. Map of coded responses

Yi remains at the top of the class, having shown nothing lower than a Level 1 on the whole test, and nothing less than a Level 2 on Part 2 of the test. On the questions related to culture, when given the freedom to produce texts, Yi showed a depth of insight concerning what the text was about both in overt references to culture (Romeo and Juliet) and patterns regarding argumentative discourse.

On the other end of the spectrum, Qi’s performance is quite erratic. Regarding lexical resource (See Table 4), Qi was the most prolific user, precisely using eight of the words. In the last question, this listener also showed an ability to produce nuanced interpretation of the debate. In Part 1, Qi performed at the highest level possible. In Part 2, many of the responses would have fit better with another question.

As far as the superiority of assessing what students have understood over the numerical value is concerned, there is little difference in the scores. Yi clearly had a superior understanding of the videos; this is demonstrated in both descriptive and numerical assessment. Er, San, and Si showed a very good understanding of the text in some responses, and no understanding at all in one or two others. Their performance is not that different. There is a discrepancy in the line up of Wu, Liu, and Qi. Compared to Liu, Wu had fewer responses showing a detailed understanding, twice as many showing a general understanding, and twice as many showing no understanding. Compared to Qi, Wu had more responses showing a lack of understanding (Qi’s Level 0 response was due to no response being given). Both Wu and Liu had four responses that showed at least a detailed understanding, but one of Qi’s responses demonstrated an ability to draw inferences.

While there is little difference in the overall scores, the descriptive approach to assessment shows a different profile regarding difficult questions. The numerical approach flagged 1-2 and 2-6 as being the most difficult. The descriptive approach agrees that 1-2 showed weak performance and that 2-6 drew implausible responses from two students and a general understanding from a third. However, it shows that 2-3 (*On which point did the linguist agree with the teacher?*) was more difficult to answer than 2-6 (*What example does the teacher give to refute the linguist’s argument?*). Three students

gave nonsensical responses and the other four had no better than a detailed understanding. This indicates a gap in the students' ability to identify more subtle points and examples in English debate.

From this data set, the questions about cultural items such as Mattel, Barbie, Shakespeare, and *Romeo and Juliet* seem to pose no difficulty for the students. In fact, the highest levels of understanding were achieved on Questions 2-7 and 2-8, in which students examined the functions and cultural patterns in the debate.

## 5. Conclusion

Although the observations gleaned from a descriptive analysis of the students' responses to a summative listening test do not suggest a great variance, they do provide some insight into the listening process and comprehension of the students and some implications for future testing. In examining the questions and the students' level of cultural awareness and linguistic proficiency in answering them, one can see that proper nouns do not cause as much drain on the cognitive faculties as hunting for points that are weakly implied. We can also see that open-ended questions are more efficient than open-cloze prompts in eliciting responses that reflect a higher level of thinking. It is not that open-cloze questions cannot be made difficult- one-syllable adverbs can create a tricky prompt– but two-word responses that mirror concepts in a speech very rarely challenge the students to draw inferences of any kind. The ability to draw inferences is what separates lower and higher levels of understanding. When creating tests, if the goal is to push students to achieve something beyond a detailed, intermediate level of understanding, test makers need to design questions that push students to draw inferences and to listen critically to the recording.

From this data set, and in this test, there was no opportunity for listeners to consider patterns of the cultural dimensions of power, competition, relationship-building, ambiguity, individualism, or indulgence. This would require a judicious selection of audio-recordings and some focused guidance from a listening coach to help students recognize these patterns. A creative curriculum developer could perhaps include recordings and assignments to direct the attention of the students to these, but such a focus in the curriculum would only be beneficial if woven into the fabric of contrastive rhetoric.

A limitation of this study is that the listeners did not receive feedback on their performance. An assessment culture that values the process of testing to promote learning is crucial for the development of any skill. A stealthy type of assessment in which listeners receive feedback in the same way they receive feedback from their favorite games might help students, even the restrained ones, become deeper, more active, more efficient intercultural listeners. Perhaps this kind of listening instruction will reduce the number of conscientious students who have to resit the AE Listening Exam.

## Works Cited

- American Council on the Teaching of Foreign Languages (ACTFL). 2020. "Listening Proficiency Test Familiarization Guide."  
<https://www.actfl.org/uploads/files/general/Documents/2020-LPT-Familiarization-Guide-1.pdf>.
- Black, Paul and Dylan Wiliam. 1998. "Inside the Black Box: Raising Standards through Classroom Assessment." *Phi Delta Kappan* 80 (2): 139-144.
- Goethe Institut. 2024. "CEFR-ACTFL scales comparison - Concepts and materials." Goethe-Institut.  
<https://www.goethe.de/ins/us/en/spr/unt/ffd/mdg/ver.html>.
- Greenhalgh-Spencer, Heather. 2022. "Deep Listening and Bridge-Building in School-Community Partnerships." *Educational Theory* 72 (6): 793-811. <https://doi.org/10.1111/edth.12561>.
- The Guardian. 2013. "Should schools ban slang from the classroom." The Guardian. YouTube video.  
<https://www.youtube.com/watch?v=I8vD8dvjNxM>.
- Headlee, Celeste. 2016. "10 ways to have a better conversation." TED Talk.  
<https://www.youtube.com/watch?v=R1vskiVDw14>.



- Hector Lanz: *How do focus groups work*. 2017. Featuring Hector Lanz. TED-Ed. [https://www.ted.com/talks/hector\\_lanz\\_how\\_do\\_focus\\_groups\\_work/transcript?language=en&subtitle=en](https://www.ted.com/talks/hector_lanz_how_do_focus_groups_work/transcript?language=en&subtitle=en).
- Hill, Jonnie. 2000. *Exploring EFL Listening Comprehension*. Stillwater, OK: OSU - Electronic Theses and Dissertations. <https://hdl.handle.net/11244/11493>.
- Hinds, John. 2001. "Reader vs Writer Responsibility: A New Typology." In *Landmark Essays on ESL Writing*, edited by Tony J. Silva and Paul K. Matsuda, 63-74. Vol. 17. Mahwah, New Jersey: Hermagoras Press.
- Hofstede, Geert J. 2024. "Country Comparison Tool." The Culture Factor Group. <https://www.hofstede-insights.com/country-comparison-tool?countries=hungary%2Cunited+states>.
- Lanz, Hector. 2017. "Hector Lanz: How do focus groups work?" TED-Ed. video. [https://www.ted.com/talks/hector\\_lanz\\_how\\_do\\_focus\\_groups\\_work?language=en](https://www.ted.com/talks/hector_lanz_how_do_focus_groups_work?language=en).
- Oxford University Press. 2024. "Text Checker." Oxford Learner's Dictionaries. <https://www.oxfordlearnersdictionaries.com/text-checker/>.
- Parish, Peggy. 1963. *Amelia Bedelia*. New York: Harper-Collins.
- "Readability Scoring System." n.d. Readability Formulas. Accessed June 14, 2024. <https://readabilityformulas.com/readability-scoring-system.php>.
- Rives, Jennifer. 2024. "Teaching Listening Comprehension with Communicative Strategies." In *American English Live Teacher Professional Development Series*. Online Professional English Network. <https://openenglishcommunity.org/page/current-series-20>.
- Schifferdecker, Lauren. 2017. "The secret to winning arguments." TED Talk. <https://www.youtube.com/watch?v=MoUlxJR-xZM>.
- Sumner, Dave. 2019. "To win you must lose: How to argue better." TED Talk. <https://www.youtube.com/watch?v=EfiGTJ86oSs>.
- Vandergrift, Larry, and Marzieh H. Tafaghodtari. 2010. "Teaching L2 learners how to listen does make a difference: an empirical study." *Language Learning* 60 (2): 470-497. <https://doi.org/10.1111/j.1467-9922.2009.00559.x>.
- Weinland, Kathryn, (ed.) 2023. *Intercultural Communication*. Stillwater, Oklahoma: Creative Commons. <https://open.library.okstate.edu/interculturalcommunication/chapter/cultural-characteristics/>.
- William, Dylan, Clare Lee, Christine Harrison, and Paul Black. 2010. "Teachers developing assessment for learning: impact on student achievement." *Assessment in Education: Principles, Policy & Practice* 11 (1): 49-65. <https://doi.org/10.1080/0969594042000208994>.
- Wright, Amber. 2020. *Ask better questions*. N.p.: TED Talk. [https://www.youtube.com/watch?v=\\_gzfIf1w7A8](https://www.youtube.com/watch?v=_gzfIf1w7A8).