GSI AMTAS: Implementation and Interpretation Case Studies

Recorded Oct 28, 2020

Presenter: Laura Prigge, AuD

AudiologyOnline.com Course #35912
Partner: GSI

- [Laura] Good morning, everyone. Welcome to today's webinar, which is called GSI, AMTAS: Implementation and Interpretation Case Studies. My name is Laura Prigge, and I'm one of the audiologists at Grayson Stadler. We have a team of three audiologists, and so I like to just highlight that before we get going. I'm in the middle. Again, my name is Laura Prigge. I have worked with Grayson Stadler for almost a little over 10 years. Before that I did a bunch of different employment places. I worked at ENTs, I worked at private practices. I worked in some hospitals. I worked at some other manufacturers and I really do enjoy my job as a clinical application specialist at Grayson Stadler. To the left, we have Tony Lombardo, also an audiologist who is a clinical application specialist with Grayson Stadler. We do a lot of training, a lot of R&D testing. And then to the right is Karen Morris. And she is an audiologist in product management. And she also assists with a lot of training.

You probably may have seen some audiology online presentations from both Tony and Karen and also me. So today we are going to talk about AMTAS, Implementation and Interpretation. And so your learning outcomes are as follows. After this course, you will be able to use the quality indicators to interpret the quality of the automated audiogram. After this course, you will be able to use the speech results to interpret the quality of the automated speech testing. And after this course, you will be able to determine the most appropriate next steps for your patients based on the results of automated audiometry. So we'd like to just dive right in and find out for sure, find out the truth. Are you on the fence still when it comes to automated audiometry? It's really interesting because I think that now with our pandemic and with some of the current things that are happening in audiology, people are talking a little bit more about automation, but GSI has been talking about automation for a little bit longer. In fact, in 2017, we brought AMTAS, our Automated Method for Testing Auditory Sensitivity. We brought this automated audiometry to all of the big shows. We brought it to AAA. We brought it to American Academy of Audiology. We brought it to ADA and ASHA, and we were just curious to see what audiologists thought. It's a pretty disruptive concept. And we saw the reactions that we got from these audiologists and they weren't always favorable and they were at best sometimes on the fence, but after we started to explore a little deeper some of the benefits of using automated audiometry, people's minds started shifting a little bit and there are very clear benefits that we're going to discuss today. So today we're gonna talk about the history of automation. We're going to review the history of pure tone testing, and then we're gonna discuss the implementation of automated audiometry and review some real life examples of results that I've received from our automated system. So first let's talk about automated audiometry. This is not a new concept. In fact, in 1947, Bekesy Audiometry was released commercially as the first automated audiometry. And I don't know if you remember how this works, but it was a very subjective test where the patient controlled the loudness of the signal that was presented with a patient response button.

So the machine would do a frequency sweep. And during that frequency sweep, the patient would push the button when they heard the tone. And when the tone went away, when it was too quiet, they would release the button until the tone got loud enough, and then they'd push the button again. So what you ended up with, with Bekesy Audiometry was this jagged line that went across a frequency sweep of the patient's subjective thresholds. And it was pretty cool, but it was really considered a special test. And it wasn't really used routinely in clinical practice. It took a long time, the lines were jagged, and people found that it just wasn't clinically that feasible. But automated audiometry has been used since then in a lot of different applications. For example, we use automated audiometry in military screenings, annual military screenings, automated audiometry thresholds are used for hearing conservation programs. And a lot of research programs use automated audiometry to test their subjects before, during, after whatever experiment it is that they're studying. The interesting thing about automated audiometry is historically, it's been pretty much screening. Even if it is a threshold test, it's a pretty screening threshold test. So we're trying to confirm normal or refer for more testing. It has been air conduction only. We're looking for maybe a shift in the thresholds or a shift in the baseline, and it has been quite frankly, little diagnostic relevance to us, the audiologists who are practicing in the clinic. With the current status of automation, it's here and it's available and it's been proven. Automation is not a new concept in medicine either. It's been embraced by a lot of other medical disciplines to help with some efficiency and to help with accuracy. For example, we have automated blood pressure testing, there's automated vision testing, and there is even some pure tone audiometry automation in other areas of medicine that are being used. The cool thing about this automation that is currently available or that is a new technology, is it is going to provide cont... And then it's gonna analyze and quantify some of the behavioral responses, particularly in pure tone audiometry and correlates all those responses to give whoever's doing the interpretation the best chance for a good evaluation.

And if technology, the advancements in technology are not enough to maybe get you over the fence of automated audiometry, I think that really one of the most compelling arguments for automated audiometry is access to hearing healthcare. When we talk about access to hearing healthcare, I mean, it's interesting. We know that there are thousands and thousands of people that have hearing loss in the United States and across the world, but in 2019, there was a study done that really wanted to explain and illustrate the landscape of audiology in the United States. And so this map was, I really enjoyed this map. It was one of the first ones that we saw on the actual, on the paper that was published. And it just shows where audiologists are, where they're practicing per 100,000 people or population in the United States. I mean, it's comforting to know that in no state on this map, there are zero per 100,000 people of audiologists, but it's interesting to see the layout and the ratio of audiologists to population. And the study took it one step further and actually went out and did a survey to find where hearing loss is in the United States. So what we can see from this part of the study on the left here is this is a map of self-reported hearing difficulty in older adults. So 65 or older. And so you can see there are a lot of little pockets where there is some significant hearing loss reported from the public. The interesting thing about that audiology landscape is this is where the audiologists are. If you look at this map on the right side, we're looking at audiologists per 1000, but it's broken it down into a much, much finer view. And we can see that there is a big discrepancy between where people with hearing loss are and where the audiologists are. And so now we start to get this idea that maybe it is harder to access hearing health care than we maybe like to think or that we maybe do think because again, we're in our own little practices and we maybe are super busy, maybe super not busy. There might be a lot of competition down the road, but when you look at these maps of where hearing loss is and where audiologists are, you can see that there's definitely a gap in the services. Dr. Margolis broke this down a little bit further for us just to really paint the picture of how desperate or how... There's articles called the Crisis in Audiology, and it really discusses how there's going to be a gap in service. The services needed and the services that we're able to provide. And so I think that this is a really good illustration of how much there is an increasing need for hearing tests and how it's gonna be difficult for us as audiologists to keep up with that demand.

So if you look at this slide, you can see there is a population from 2020, so today, to 2060. So we're looking at about a 40 year projection of things. So the first thing is just population. Right now in the United States there are 333 million people. And in 40 years, it is going to increase to about 417 million people. So the population is growing. The prevalence of hearing impairment is not growing quite as quickly. You can see now in 2020, that's about 17% of the population has some sort of hearing impairment. That's about 56 million people. And in 2060, it only increases about 1%, but the number of people that it increases to is 75 million people. So in 40 years, we're gonna have 75 million people that have a hearing impairment. If we break down, how many of those people that get tests or how many maybe seek out hearing health care? For this exercise, we're gonna assume that these people, these 56 million, these 75 million, we're gonna assume that it's probably maybe a biannual hearing evaluation. Wouldn't that be nice if everybody came in every other year to just get a hearing test to see where they stood? And then we're gonna have to subtract about 25% of those tests because they're normal hearing. They can go home for another couple years, but what we can do is we can break down the number of total hearing tests that we need to perform every year starting this year, 2020, is about 35 million audiograms. And that increases in 40 years to about 47 million audiograms annually. What does that mean for us as audiologists? Well, this is pretty interesting. If you look at the audiology profession as a whole, and there have been a lot of papers and a lot of studies done on this. The number of full-time audiologists in the United States here in 2020 is around 12,000, 12,019, but we are also working in what we would consider to be a flat profession. It is not growing exponentially. It's growing very, very, very slightly if it's growing at all, but it is a very flat projection for the growth of our profession at this time.

So in 40 years, we're still gonna have about 12,000 full-time audiologists. If we add the other people that are doing audiograms, that is maybe hearing instrument specialist, hearing aid dispensers, there's probably, we estimate to be about 6,000 in the United States working full-time now and their profession projection is also the same, it's flat. And so again, in 40 years, it'll probably be around 6,000. So that's like 18,000 people in the United States doing hearing tests. So if we break down how many actual audiograms we can do, if there are 18,000 of us and we do three per work day, and I know that many times we do more than three, but we also do hearing aid evaluations and ABRs and BNGs, and all the other things that we do as audiologists and hearing healthcare professionals, three a day is a pretty solid estimation. So that's 920 audiograms per year. So that means we can do about 17 million audiograms every single year, which is impressive. And in 40 years, we're still going to be doing the same amount, 17 million audiograms. When you compare that to what we need, what the demand is for audiograms, we are clearly falling short. And there are actual graphical representations here that indicate, look, here we are with our 17 million audiograms as the years go by, and the need for audiograms is increasing substantially. So we really do struggle with figuring out how to meet the need in the United States for these hearing tests. And the other interesting thing is this is not just a United States problem. Access to hearing healthcare is a problem globally. And in 2013, the World Health Organization put together these maps and these charts. And on the left here, you can see that this is where, and it's self-reported, disabling hearing loss occurs. And much like the other graph, the darker the color, the more disabling hearing loss there is. And if you look over here on the right, the number of audiologists per million and the number of ENTs per million reported, it's sometimes not equal where the hearing healthcare providers are to where the hearing loss is. And so this is, again, a global situation where access to hearing healthcare is tough and is tricky. I have an audiologist friend in Thailand and we always laugh. And she says, "There's 70 million people that live in Thailand and there's about 70 audiologists. So as far as I'm concerned, I am one in a million." And we giggle about it and we laugh a little bit, but then when you think about it, that's a lot of pressure. One audiologist per million people is staggering. How do you meet the needs for your citizens?

So the question then becomes, how do we address this increasing need? Well, there are a couple of options that we have. We can increase the workforce, but we know from the projections and from the illustrations that I just showed you that that's not probably the most, we shouldn't rely on that one the most heavily, because it is projected to be a flat profession. The next thing that we could do is we could you utilize telemedicine and teleaudiology more. And I will say that now, currently, we are seeing a huge increase in teleaudiology, and that's fantastic, but a lot of times, instead of increasing access, we're using the face-to-face or the remote live video teleaudiology, specifically for programming hearing aids, programming cochlear implants. If there is any testing done remotely, it is utilizing this synchronous teleaudiology, this face-to-face, which significantly reduces the geography challenges, but it does not reduce the access challenges 'cause we, the professional audiologists, still has to be using our time on that call. Another option that we have is to use our workforce more efficiently. So I know that lots of places do use technicians and a lot of places do use audiology assistants for some clean in-checks, for some counseling, for some things like that. But sometimes the things that we need help with to save time and to see more patients is not within the scope of practice of some of those technicians or audiology assistants. And I think it's very state dependent, but I think that it's for sure a fact. So another way that we could address this increasing need for hearing healthcare and access to hearing healthcare is automation. It is realistic and it's achievable with the current state of technology and with the GSI AMTAS. We can automate with confidence, pure tone, air conduction, and bone conduction testing, and we can automate with confidence the speech testing. And so what we're doing is taking that basic diagnostic test, one of the more basic tasks that occupies a great deal of our time, and we're using that to help us address the issue of access to help us address the crisis in audiology. So it really does make sense after seeing all of the compelling arguments of access and time and geography to consider some sort of automation, to assist us in gaining the time back from doing tests. So let's talk about how it is possible to automate this really important test. I mean, pure tone, air conduction, and bone conduction and speech testing is the foundation and the starting point in our test battery many times, and this is the test that we use a lot of times to base other decisions on.

So we need to figure out how to increase confidence in the procedure and in the results of an automated evaluation. So for a minute, let's go ahead and just talk about pure tone audiometry. Pure tone audiometry, the first public, or I'm sorry, the first published methods like the pure tone methods were published in 1924 in a journal called the Laryngoscope. And it was really interesting because if you read that, it actually is very similar to some of the other subsequent articles that describe how to perform pure tone audiometry. In 1944, the original Hughson-Westlake method article was published, and in 1959, the Carhart and Jerger method was published. We have really relied on those pretty heavily ever since. In the '70s, there were several papers about an adaptive psychophysical procedure for audiometry that was used mainly in research. It never really caught on clinically, but there was another a little blip in our consistent pure tone testing. And then in 2004, the ANSI Standards released some guidance on pure tone audiometry. But it's really interesting to consider that manual audiometry has really not changed much in the past almost 90 years, and each and every one of these methods leaves a lot of the procedural variables to the discretion of the tester. So basically they're saying, okay, we're gonna give you the very, very loose guidelines, and then you, the tester and your expertise, needs to figure out the methods going forward from there. So then we have to ask ourselves how consistent are pure tone tests from clinician to clinician? How much variability is there in actual pure tone testing? Well, from my perspective, as an audiologist, if you asked me this, I would say, I mean, it's pretty consistent, we all start at thousand. I mean, some of us go from a thousand to 500 and 250, and some of us go from a thousand to 2000, but we're going down 10, we're going up five. It's pretty consistent. I mean, right, I think so. But then here's the next question: what if a patient brings in a recent audiogram from another clinic? What are you gonna do? I mean, we're thinking, oh, what if it looks like this? I take look at this as an audiologist. If a patient brought me this audiogram, I would be like, man, that audiogram is gorgeous. Look at that, but wait a second. Is it really possible that all of the bone conduction scores matched up exactly to the air conduction scores at all of those different frequencies? 'Cause I'm pretty sure there's is some research that indicates there's a lot of variability in bone conduction testing, and to get an audiogram this pretty is about a one in 264,000 roughly chance. So now I'm trying to second guessing it.

So what we found is it's actually very common for the audiogram to be repeated because of lack of competence in the original results. And it's funny because I've asked several audiologists this very question. So if somebody brings in an audio, what do you do? And I have one audiologist friend who's like, "Oh, I tell him for sure. Don't show me the test. Don't show it to me at all. I wanna do the test and see if I'm close." And it's a joke there because of course they're gonna be close to, the test retest between clinicians is five DB or so. So that's fun, but then also there are other audiologists that are like, "Oh no, if I get a reference from Susie down the street, I always trust that one. I know her, I know her methods and I trust her audiograms. But if I get one from Joanne down the street, I'm for sure gonna retest that one. Also, if I get a test from someone I don't know, I'm gonna retest it." And I think that it's interesting because I don't think that the thresholds are gonna change that much, but what is it exactly that we're missing? Like what is it that makes us not trust the audiograms that someone brings us? Is it because they're too perfect? Is it because maybe they're not quite as... It's actually the human factor. It's because when we look at these results and we do our own tests, we're having observation. We have time to observe what the patient is doing in that booth. We have developed over time. We know that a lot of the testing is left to the discretion of the tester. And when we think about it, that's really what's happening. We weren't there. We don't know how that patient was responding. Why in the world would we have such beautiful responses? I mean, was the patient just that good or was the patient... Did they have to coax some of those responses in? We don't know.

So the reason that we struggle to trust some of the audiograms that we didn't do ourselves is what I like to call the human factor. It's really called tester bias. And so when we think about this tester bias, or we think about the human factor, it's really practicing the training versus reality. We know that as we gain more experience as audiologists, we learn certain tricks that ensure accuracy and it's all the way that the patient is interacting with the tones or with the speech. And we're observing things like their body language. And how fast are they responding when we're super threshold or when we're closer to threshold? I know that when I'm in a booth and I'm getting my hearing tested, my eyes are closed. I'm holding my breath 'cause I'm listening so hard, but these are the cues that we pick up on so that we can say, this is a good and reliable report or good and reliable audiogram. I trust these results because I was there watching the patient take it. We don't know what other testers are doing. We don't know how the patient is responding and what tricks and tips the other audiologists are using to get the results that they're getting and the varying levels of the skills of other audiologists and their testing methods can result in a lot of widely varying accuracy. We know that we have little tips and tricks and we have confidence that we might, and I don't wanna say we're cheating, but we might cut some corners without affecting the integrity of the results. How many times do every single patient test and retest a thousand Hertz? How many times do you give them just one more beep, 2000 bone conduction to see if you can get that bone to match up a little bit better? So we're very sensitive to different cues that the patients are giving us. And so that's what we're basing our confidence on. So we're not really concerned that the pure tones might be off by five decibels. We're really looking at the patient behaviors. And so what we're trying to figure out is how can we combine some sort of indication of how the patient was acting with those pure tones that they've brought in from another clinic? And if you ask me again and say, well, is it like a standardized method for pure tones? Even though I just said all that stuff, I still say, yeah, I think that is a pretty standardized method, but there's so much variability. It may not be as consistent as we thought. And this is why it is scary to trust an audiogram from somebody else much less from an automated system. And this is why I think that we all are a little on the fence about... We could be on the fence about automated audiometry.

So what is the missing link? What is the missing link in our comfort level of standardization? And I think that the missing link is standardization. So when we're doing a hearing test, we know that the basic pure tone test doesn't require the skillset of an expert audiologist. They've been using automated audiometry for screening evaluations for decades. But what those things don't tell us is the story of the patient behavior. We need to be confident that the patient and the reliability and the responses are something that we've observed and something that we can evaluate. And so with the GSI AMTAS, automated audiometry system, we are actually... in your interpretation. So when we talked about standardized audiology, we're not actually talking about everybody needs to do the test in the same exact way. We're talking about, AMTAS, the automated audiometry. That is a data-driven program that is going to provide data not only on our pure tones, but on the behavioral cues of the patient. With AMTAS, the test is repeated the same way every time. They do test and retest a thousand Hertz, and the level of consistency is thorough but efficient and not taking shortcuts. And then we can be confident in the method. And the automation lends itself to this consistency of standardization. The way that we actually are standardizing or displaying our patient cues is a patented method called Qualind. And this is a way that we determine the accuracy of the tests. It's giving us from the second that you start an AMTAS automated audiogram, we start taking measurements. And so we, at the end of our evaluation, have a data-driven set of quality indicators that provide this information on the cues we need to interpret these audiograms with confidence. So the more familiar you become with these quality indicators, the more meaningful the test is that has been done by an automated system. And it's easier for you to make decisions confidently about the most appropriate step for your patient in their hearing health care journey. When we do automated audiometry, again, when we start AMTAS, GSI AMTAS, we're looking at not only thresholds, but these quality indicators that I'm going to review individually.

And then we're going to jump into some case studies to try and really figure out how they're working with the thresholds to give us the cues necessary. There are up to nine quality indicators that come with every report of AMTAS, the automated audiometry. Sometimes there will be fewer, depending on which tests you've done. And sometimes there will be more, but we always have some quality indicators that will give us data about the behavior that the patient was exhibiting during the evaluation. So the first one that we like to talk about is the predicted accuracy. So this is your judgment. This is in a manual audiometry situation. This is the judgment that you give to your patient based on how they were acting in the booth. It's the oral reliability of the test. And just like most audiologists, the Qualind or the quality indicator will give you either good, fair, or poor reliability based on the patient responses and based on the other quality indicators and then some advanced statistical modeling that will predict the quality of the evaluation that you're reviewing. The second quality indicator is the predicted average absolute difference between automated and manual thresholds. So basically what happened is we needed to figure out how we can display and how we can show that these are accurate, even when compared to someone who had this same test done manually from an expert audiologist. And the way that this data was derived was from two studies that were done by Dr. Margolis. And the first study was of course, to see, and to verify and to confirm the inter-test differences in manual audiometry. So they took two expert audiologists that had at least 10 years of experience, and they went to three different sites and they each had six different patients, and they just tested their hearing. And thank goodness much like we would expect and much like we were trained, the inter-tester differences were very small, within five DB. There were a couple outliers, maybe about six, but with air conduction thresholds of the test-retest from two different testers was very consistent and very similar. For bone conduction thresholds, there was a little bit more variability, but we will refer to the fact that there is a lot of bone conduction variability just by the nature of the test. And there are actually several articles and several, even a webinar on audiology online about bone conduction testing that I would encourage you to take a look at.

So we realized or we figured out and we verified that tester differences manual audiometry are pretty stable. Then the studies went to the second stage and they were actually comparing the AMTAS automated audiometry to the manual threshold comparison. And so in this particular case, they had a couple of audiologists, expert audiologists. They had 30 subjects and these subjects went through automated audiometry and then manual audiometry. And they did the comparison and found that it was very, very comparable. The air conduction differences were well within the tolerances between three and 4.5 decibels and bone conduction again was a little bit more variable, but we can be confident in our quality indicator that is predicted average absolute decibel difference. So we can tell and we can interpret how much different would it be with this patient if we did manual audiometry versus automated audiometry? The third quality indicator that we look at is a masker alert. So with the AMTAS software, every presentation has masking in it. And so when we start reviewing the case studies, you'll notice that there are masked symbols for all of the pure tone, air conduction, and bone conduction tests. And so there is masking presented. When the test has been completed, there is an analysis performed of all the masking levels of all the thresholds. And if there are areas that were maybe too high of masking or too low of masking, so there was under masking or over masking, it will indicate that on the quality indicator so that you can review that. And then you can make the determination because you are the audiologist, the expert clinician, whether what you wanna do about that. If you want to either retest with a different masking, or if you want to just continue on with your other evaluations for the patient. The next thing that we look at in our quality indicators is the time per trial. And this is the average response time from when they say, listen for a tone to when the patient responds, whether or not they heard the tone. And what we do is we average from the whole test, because we know from our clinical testing, that when you're presenting a tone that is super oral, they respond pretty quickly. But once you start getting down toward the threshold, it might take a little bit longer for that patient to respond.

And so this average time per trial, when we average that out, it really gives us a good idea of how responsive and how dedicated and how much attention was being paid during the evaluation. The next thing we look at is the false alarm rate. Now the false alarm rate, we know when we're doing manual pure tone audiometry, we can't present the tone every second and a half over and over because the patients will fall into a rhythm. And so we don't have the luxury of maybe spacing out the time and things like that. When it is a closed set, make a decision automated system. So what AMTAS does instead is it presents catch trials. So that means it will say, listen for a tone. And then it will say, did you hear a tone? But they didn't present a tone. So we have these catch trials randomly presented throughout the test and then that lets us quantify the number of false alarms. And this is an interesting one because you if you're concerned about a malingerer or you're concerned that the patient may be getting into a rhythm, this is something that you can check because there is a note or a page that is displayed on the software that says, "Hey, you just responded to a tone when there wasn't a tone, so please listen." And that is the equivalent of us refocusing our patient after they get into a rhythm in manual audiometry. So this is a really interesting quality indicator to review and to look at when we're interpreting actual responses from patients. The next quality indicator, it's an easy one. It's just the average test retest difference. And so they take the average of the first time they tested a thousand Hertz, and then they retest every time in both the right and left ear. And they just take the difference of that. And it should be within a decent tolerance, 5, 10 DB. And if it's not, then again that raises some red flags and some interpretation flags for us, the interpreting audiologists, to make decisions about in the patient testing. This next one is the quality control fail rate. And I really, really liked this one a lot, because again, how are we verifying that... What if a patient gets lucky? What if they're really not paying attention and they just get lucky and I get this audiogram that looks good but it's not accurate. Well, with all of the other quality indicators and especially this quality control fail rate, we really do catch a lot of the, I'm gonna, say malingerers. It's really funny, whenever GSI gives out a demo or they go in and show the AMTAS to people, particularly college students, the college students and the professors are like, oh, we're gonna cheat this. We're going to win by cheating and we're gonna show you that it may not be as accurate as you think. And they can't cheat because the quality, the data that we're collecting with these quality indicators shows you the inconsistencies.

And it shows you some of the places where they're really trying to cheat and it shows up in false alarms and it especially shows up in the quality control fail rate. So every time AMTAS establishes a threshold, it will immediately present a tone that is five DB louder. Everyone should be able to respond to a tone that is five decibels louder than their threshold. And so if you don't respond to that, then that can throw up some red flags. And then the other quality indicators that we have are regarding air/bone gaps. And this just, again, alerts you to the number of air/bone gaps that are greater than 35. And then it also alerts you to, if there are an excessive amount of reverse air/bone gaps. And so these two quality indicators just are identifying and evaluating the number of air/bone gaps and alert you to any issues that may be caused by air and bone conduction gaps. So in the AMTAS system, there are up to nine quality indicators. That's what the AMTAS pro, that's when you're testing air conduction, bone conduction in speech testing. When you use our product that is called the AMTAS Flex, it is a portable tablet-based AMTAS, you still get the threshold and it's still the same process, but there are not the same number of quality indicators because there's not bone conduction testing. So those will fall off and you won't see those in the

AMTAS Flex. But overall, those quality indicators are giving you the cues that you need to really imagine what was happening during the evaluation. In addition to quantifying the behavioral cues that the patients are giving us, I think that one other goal of AMTAS is to actually classify the audiogram so that we can have a nice and standard description of the shape of the audiogram. There was a study done on this to figure out what the best method for classification was. And it's a really, I think kind of funny, funny story. They got together five audiologists who are expert audiologists with 10 years of experience and just gave them a whole bunch of audiograms. And so these audiologists, their instruction was, classify these audiograms. And so they started writing, and they're expert audiologists, they're like, oh, cookie bite hearing loss, right. Ear precipitously, sloping, blah, blah, blah.

So they started doing all of these different, the words that they learned in placements or in college or in the field. And when they added up all the words for configuration, severity, and site of lesion, there were enough words that would create 476 billion unique combinations of classifications of audiograms, and that seems ridiculous. It seems like too many. It seems like too many to actually have some sort of standardization. So the audiologist and the people in charge of the actual experiment came to a consensus and they decided that for AMTAS, we will use just a few. So for configuration, we just use normal, flat, sloping, rising, trough, peaked, and other. Severity, we've cut that down. There's no moderately severe, there's no slight. We just use normal, mild, moderate, severe, and profound. And for site of lesion, conductive, sensory, neural and conductive, sensory, and mixed. And so we're able to, with our automated system, evaluate the shape of the threshold so that we can classify it and it will start giving a pretty standardized classification for the shapes of the hearing losses that we see in AMTAS. So then of course the next question is, all right, so pure tone, that's cool. The quality indicators, I'm really excited because I can see how those will assist me in understanding the behavioral cues of the patient. How are we gonna automate speech testing? So, first of all, it's important to note that with AMTAS, this is a one-time patient setup. So we're putting the bone conduction and the headphones on, we're pressing start. And then there's very minimal supervision that is needed for the patient to complete this test on their own independently. So everything we do is a forced choice and speech is no different. So with SRT, it's pretty easy. I mean, it really is very similar to the pure tones. They say, listen for a word, we throw up four spondees and they select the word. And of course the intensity goes down until we establish a threshold. For word recognition, it's a little bit different. Because it is forced choice, the developers of AMTAS needed to figure out what an appropriate level for presentation was. 'Cause what was found is if we presented at, I don't know, 65 or 40 SL or something like that, the task was too easy, even though we're using the phonetically balanced word lists, which are any six by the way, and we're putting up rhyming words. If the decibel level was too high, it was just too easy of a task. And we wanted the results from this automated hearing test to be more meaningful.

And so there was a study done and there was a poster that was presented at the American Auditory Society in 2016, that determined the best predictor of the word recognition in correlation with real life and with manual and live voice testing is if you present the any six word list, 22 decibels louder than the PTA, the pure tone average, or the SRT, you can choose. And then when we present those words at a pretty soft level, it's not as cut and dry. It's not as easy of a task for the patient who's doing the test. And so what that gives us is actually a pretty good idea of their speech understanding because if they get 88% correct on their word recognition at 22 decibels louder than their pure tone average, that's excellent. We're calling that good and excellent and then we stop the test. If they don't get 88%, we increase the decibel level about eight DB, exactly eight DB. And we present another list. And what we wanna see in that situation isn't that everybody's getting 88% or 96% or whatever the case is, we wanna see that there is a PIPB improvement when we turned up the volume, because with hearing, when you turn up the intensity level, when you turn up the decibel level, not only does the volume increase, the understanding increases. So that's how we handle the speech testing with the AMTAS software. So now what we're gonna do is we're gonna look at a number of different case studies. And these are submitted from actual audiologists around the country who are using AMTAS and they come from a variety of settings. So some of them are from a large medical practice. Some of them are from an ENT practice. Some of them are from a private practice that has a lot of medical referrals, but we're going to dive in and just take a look at what we, I, the audiologist, receives when I get an AMTAS, automated audiogram from a patient. So here's the first one that we're gonna look at. I chose an easy one. I chose one that is perfect for us to actually look at. So the elements that are on the AMTAS audiogram are as follows. You get the audiogram, which is great. Please notice the mass air conduction and bone conduction on the audiogram. There's a nice little thing that shows you the normal hearing level. You also get the masking level table. So you understand what levels were used for the masking to get those results, and then you have your quality indicator table, your Qualind. With this particular evaluation, it's an overall good quality. If you look at each one of the individual quality indicators, they all support that this was a solid test and that you can comfortably move forward knowing that this is accurate.

So the accuracy, remember comparing manual versus automated, was about 6 DB, no masker alerts, time for trial was three seconds. This patient was fast. The false alarm rate was about 3%. The test retest at a thousand was three decibels, and there were no check fails. So they were able to successfully hear all tones presented five DB louder. The one interesting thing about this audiogram is the bone conduction. There is one reversed air/bone gap here at 4,000. And again, I'm not gonna spend a lot of time talking about this because of the bone conduction variability that has been well-documented in the literature. But I would encourage you to just look at some of the resources that are available because when we get AMTAS audiograms, you'll notice, we will notice that there are some outliers, and sometimes it's the patient's true responses without any coaxing. What we have to be able to do is realize when it is just a variability or when it is something that is diagnostically significant. All right, let's take a look at a couple other good ones. So this one is another good one. You can see it's from a different location, but you can see the audiogram looks fantastic and consistent up here. You have nice normal hearing. AMCLASS calls this a sloping hearing loss in one of the ears and flatten the other one based on our predictive classification. And it's a good quality result. Accuracy between manual and automated is six DB, no masker alerts, trial time is 3.9 seconds. So that's the average time. No false alarms, test-retest was three DB and the check fail was zero. They successfully completed all those things. Now take a look down here. I guess I would really expect this when I do the word recognition scores at 27 DB, which is 22 decibels higher than the PTA. Both ears got an 88% correct. So we are able to move on. From this particular evaluation with our next appropriate next step, is that the quick SIN? Is it some speech-in-noise testing? Whatever's next on your agenda for testing is what you can do with confidence. The next couple reports we're gonna look at are interesting because much like every audiologist who considers the automation, we also just wanna make sure for our own good, just because it's scary when you start trusting a computer to do your job.

So this is kind of interesting though, I think. It's cool because you can see here on the audiogram, you definitely see the thresholds are here and when they were repeated manually, I would say for sure that they were within that six B accuracy that we saw from the statistics predictors in the studies that were performed to develop that prediction. The masker alerts, there were nine, trial time, it was three seconds, false alarms, minimal, test-retest was, I can't read it, maybe three, and there were no check fails. And so this, I would say again, is a very positive reinforcement of how accurate the AMTAS results are. Again, we see those outliers. You guys, you see these little reversed air/bone gaps here. I secretly wonder if I were doing this test, if I might have presented maybe just one more time at like 20, at 2000, just to see. I mean, I'm not saying I'm cheating, but I'm just saying that every once in a while I present just one more time and see. But if you look at the other tests that are always part of a comprehensive audiologic battery, you will be able to interpret and determine whether or not these outliers are normal variability, or if there's something that is diagnostically significant, that is contributing to these results. Here's a change. Yep, here's another one just from that same customer where they had the thresholds that were done with AMTAS and then the thresholds that were done manually and they are in very, very good agreement, good reliability. And again, I would move forward with my next testing. Here's another good test report here. And again, I'm looking at these just to get the hang of what I'm actually seeing. How are these quality indicators contributing to my comfort level in moving forward without having to repeat this test? So this one's an interesting one only because there's only one bone conduction threshold, and that's because AMTAS doesn't test bone unless the thresholds are poor than 10, because if there are conductive components, those will be picked up in otoscopy, in tympanometry. And so your battery of tests that you have plenty of time to do now because you are getting your air, bone, and speech done with AMTASs, you will catch some things, like maybe some conductive components there. And then you can do some bone conduction if needed, but otherwise, the quality and the accuracy, this test is, from my perspective and my interpretation, a great test. We'll continue on with probably seeing them next year. Depends, I guess, on what they're in for.

So here's another example of AMTAS. It's a good quality, even though the bone scores look just a little bit off. When I look at the accuracy, the masker alerts, the trial time, the false alarms, the test-retest, and the check fail, I'm confident that those are likely not contributing to the hearing loss. This is just seems as though it is a natural variation. And again, these things will be caught when you do your other evaluation. Another good quality test. And I think the reason that I think it's important to go through these good ones is just to see some of the things that people see. When people call me after they've started using AMTAS, some of their questions are, gosh, these bone conduction scores look kind of weird. I can't tell what's going on here. So even though this is a good quality test, my first initial thought is this bone conduction score is up here. Why, is there really that much of a conductive component in the low to mid frequencies? Well, let's look at some of the other stuff. Let's look at the accuracy and the masker alert and let's look at the speech testing. I mean, if you look at the speech testing down here, it was 76%, 64%. So we of course increased the decibel level and the percentage increased as well. My initial thought, because I have interpreted so many AMTAS audiograms, is this is not a conductive component. This is actually the patient misunderstanding the task and responding to masking noise. So because it doesn't seem right, but guess what, I'm gonna do some more testing. I'm gonna actually do some temps. I'm gonna take a critical look in the ear to see if there's any indication in otoscopy that there might be some fluid behind there, and this is going to help me determine the next step. So now we look at a good reliable test, but it has a red flag. Red flags just mean that based on the studies that were developed in this patented quality indicator table, it is higher than the 80th percentile. So they have responded in a way that is higher than the 80th percentile of accepted level. And that will give me a red flag. That does not mean I need to throw out this test because a lot of times, you still have a good reliability.

See, the patient is good. I'm not sure, the false alarm rates, that means that they've pressed yes when there wasn't a tone. That is something that I need to be concerned about. But then I also need to remember that sometimes it's hard to hear. What if they're just listening real hard? What if there's a little bit of tinnitus that was sort of distracting them throughout the test? And that's, again, I look at the patient history. I look at some of the other indicators from talking to the patient or from this actual result. Accuracy is good, trial time is good, false alarms are not good, test-retest is good, their speech is fantastic and inline with a consistent test. I think that I'll take this test and move on with my next step. And then here we have just a very, very basic, I bet this is the fastest test ever, even though their trial time was 4.4 seconds, remember, were these really quiet sounds and they had really good hearing. So I bet that it took them a while to determine whether or not they heard the tone. You can also notice in this particular good quality AMTAS tests that there was no bone conduction tested. And that's because again, if the thresholds are 10 or better, we do not implement bone conduction. And if you look down at the speech testing, they did a fantastic job at 22 decibels higher than their pure tone average. So now this is an interesting one too, I think. It's just interesting because holy cow, there's a huge asymmetry in this. And initially I'm like, oh my gosh, I gotta get them into the booth and test them right away. But really, the only red flag that I'm seeing is the masker alert. And if you look here in the actual legend, it tells you that there is likely under masking at 4,000, and you can see how that's possible. There are limits to the transducers. But again, if you look at the other stuff, the accuracy, the trial time, false alarms, retest, these are all indicating that these responses are pretty accurate. And honestly, I don't know, I might retest this one and I might not... If I had just first got AMTAS recently, I probably would retest it but with more and more experience, I mean, obviously there's gonna be more testing. You're gonna be doing temps. You're gonna be doing reflexes. There's gonna be some speech testing. But yeah, I think that these thresholds, I personally would be confident in moving forward with these as my basis. Here we have another good audiogram, the major red flag in this one is the trial time. So it just shows that they were really deliberate in their pressing of the yes or no. But if you look at all the other ones, all the other quality indicators, the accuracy, false alarm, test-retest, everything looks great. And I would proceed with this one as well. Even speech testing was fantastic.

And now this is our last good one that we're gonna look at. And this is an interesting one because this one, the quality indicator that is flagged is the check fail. Remember, this is the one that you're presenting a tone, five DB louder, and they should hear it every single time. Yet somehow, the check fail has been highlighted in this particular quality indicator table. So now I instantly think, oh my gosh, this is a person who's obviously malingering, but then I look at the other cues. So now I'm gonna look at the quality. It's good, even though it's a check fail red flag, how is that possible? Well, the accuracy is within six DB, there's no masker alerts. The trial time was completely appropriate. False alarm's within normal, test-retest was dead on. So maybe they're not malingering at all. We're looking at these other things. Maybe they just really didn't hear it. And it's just something for us to look at. We look at the speech recognition as well, and in the right ear they achieved their 88% the first time. And in the left ear we turned it up and the speech remained the same. But this is one that again the reliability was good based on the quality indicators and the one red flag that I saw, the check fail, after looking at all of the other quality indicators, all the other indications of this evaluation and the patient behaviors turns out it's okay too. So now we're gonna get into some reports that are not quite as good. So now we're looking at quality that is fair instead of good. And we do see some red flags and in my audiology mind, I'm like, well, I guess I should throw out every test that's fair because it's not good. And I don't know if I can trust a test that is not good, but then I also think, you know what, in my practice, there were a number of patients that I did rate as fair. They weren't as good at the task. It doesn't mean I'm gonna get better responses or better thresholds as I retest them over and over. They're just not as good at the test. And if you looked at this one, look at these no response bone conduction scores down here. I mean, that probably contributed to the fair situation in the trial time. It took them a while to get through the evaluation.

So I don't think that these thresholds are probably off by much, but I think that this patient is, had a little bit more trouble with the task. Let's look at this one. This is another one where I'm like, huh, fair, oh, there's three red flags. So the trial time, it's longer than you would expect. Accuracy is within eight DB, masker alerts, zero, false alarms, three, test-retest, five, check fail, it's red flag but it's not the end of the world. And we have these two reversed air/bone gaps. So this is one again that perhaps I would probably just move on to my additional testing. You know that there is more tests and it needs to be done. So with the accuracy being appropriate, the test-retest being appropriate, this is the one that with confidence, I'm gonna say, you know what? The patient took a lot of time, struggled a little bit, but I'm gonna do more testing. So I'm gonna take these thresholds and move forward. Here's another fair one. And again, this is one where I have a little bit of conductive component in the actual thresholds here, in the left ear specifically. The accuracy is fair, sorry, the quality is fair and the accuracy is. All of these indicators tell me that the thresholds are probably pretty accurate. There is some under masking at 4,000 again, but the thresholds were not, this is not the only test that this patient is going to receive. So we're just gonna keep moving forward and take these thresholds along with the next step. And there are just a couple of other fair reports just to indicate and show you that just because the quality is fair does not mean that the test should be thrown out. It is a overall look at all of the behavioral cues that are indicated in our data-driven quality indicators that will help us make the determination. So then we get to this one and I'm thinking to myself, oh my gosh, what is happening here? Look at all these red flags. The quality of the test is poor. The accuracy is 14 DB. That is a 14 decibel difference between manual and automated audiometry. The test-retest is 15 decibels. So I'm thinking, gosh, this patient was really struggling with the task. So I would probably retest this one manually. I need to look at all the factors of the big picture, what other tests are happening to maybe address the air/bone gap here in this section? But every time I look at these quality indicators that have been highlighted, I really think that this patient just wasn't good at doing the pure tone testing. It was a difficult one. Honestly, the thresholds are probably close, but this is one that I for sure as an expert audiologist would retest manually. They say that about 85% of people in the study was performed in the veterans administration. 85% of veterans were able to successfully complete AMTAS.

And so our skills are certainly still valuable and needed for some of these people that just struggle with an automated task. Here's another poor one. And I gotta tell you, when I look at this and I'm like, holy cow, look at this conductive component. And then again, you're not gonna get negative 10 bone conduction with that kind of loss that looks like it's a typical loss for our population that we're testing. And after interpreting AMTAS results for long enough, I can clearly see and immediately know that this patient was responding to the masking noise for bone conduction. And that set up all of these other red flags. You can see there's over masking, there's under masking. And so it is all based on this patient who was actually just listening to the masking. So I might either reinstruct with the AMTAS and have them do it again, or just verify manually the results of this test. Here's another one, it's poor reliability. And I can see instantly that this patient was responding to the masking noise in bone conduction, and that just sets off a bunch of other red flags and the same thing. I would probably either reinstruct, have them do AMTAS again or do those things manually just to verify. Here's another poor accuracy one. This one's a tricky one. Again, I don't think that the... I don't know, I don't think that the thresholds are off by much, but it's 11 DB accuracy, false alarm rates. I would likely do this one as a manual test as well. And then finally, we're leaving you with the tough one here. This is another one that is just a poor accuracy test. There's a 12 DB predicted difference between manual and automated. There's a masker alert, there's false alarms, there's air/bone gaps, and this one would require additional testing and you can choose whether to do it via AMTAS or whether to do it via manual audiometry. I think that really going through these case studies and talking about the different configurations that we see with our quality indicators is a true testament to the fact that automated audiometry, AMTAS does not replace audiologists. The interpretation and the intervention requires the skillset of an experienced audiologist to know when to move forward, to know when to do additional testing and to know when to repeat the test manually.

And these quality indicators are what gives us the confidence in the results and that really, those qualities, that patented quality indicators are what sets AMTAS apart from a lot of other automated evaluations, because we're moving towards standardizing and quantifying not only the threshold, but the cues that the expert audiologists use to make the diagnosis of patients. And one final thought before we go is, just like most GSI products and special tests and things, we are continuing to evolve and to make changes. And lots of these come from reports from the field, from additional research. And so we do have some fund enhancements that are going to be coming in our automated audiometry. For example, we saw two poor reports that were a clear indication that the patient was responding to bone conduction masking noise. So in the future, before we do anything, we're gonna do a non-ear specific bone conduction test that will hopefully reduce the errors in responding to masking noise. Of course, if there is any asymmetry in the air conduction, we'll retest the bone with masking, but we're hoping that not only does it reduce errors and confusion but also decreases the amount of time that it takes to do this test. And then the other thing that's really exciting about AMTAS in the future is the introduction of a non-linguistic speech test. So instead of relying on the patient reading the words, there will be something that is non-linguistic. So this is good news because we're able to provide automated speech testing that does not require literacy or great vision. And it also does not depend on English as being your first language. So these are advancements that are coming in the future to again try and address some of the crises in audiology to address some of the increasing need for hearing health care that we may or may not be able to, that we statistically will not be able to address. And with that, thank you all for coming. If there are any questions, you can type them in the question and answer pane. Oh, wait, we do have questions.

So the question is if the patient has SCCD, and the air conduction is normal and the temps are normal, can you program to test bone conduction to see if there's an air/bone gap in the low frequency range? Yes, the answer is yes. You can set AMTAS to just do air conduction or just do bone conduction or just do speech testing in any combination. So yes, you are able to customize in that way. At this time, the bone conduction is just 500 to 4,000 and the air conduction is 250 to 8,000. Great, thank you. All right, and if there are not any further questions, I will end the webinar. Thank you again for attending. And if you do have specific questions, our contact information is available.