

Assessment: How to Know What Your Employees Know

What are your employees' skills? How do you know? Assessment provides key information about what your employees know and can do. In this episode, Traci Lepicki (Associate Director of Operations & Strategic Initiatives) and Dr. Bridget McHugh (Research Specialist & Psychometrician) discuss the science of assessment and how measuring knowledge, skills, and performance can inform your training decisions.

Materials contained within this podcast are the copyrighted property of The Ohio State University.

Fara Allen: Welcome to CETE Works, a podcast produced by the Curriculum and Training Team at the Center on Education and Training for Employment—a translational research center on Ohio State's campus. We work where research meets reality. I'm your host, Fara Allen, a Program Coordinator at the Center. Our first series of podcasts focuses on workforce development and will feature discussions about preparing an organization for implementation or modification of a training program. To learn more about our work, you can visit our website CETE.osu.edu, that's C-E-T-E dot O-S-U dot E-D-U. If you've listened to our previous episodes, you might be familiar with Traci Lepicki, the Associate Director of Operations and Strategic Initiatives and also the Program Director for the Curriculum and Training Team at the center. Here, you'll hear Traci talking with, Bridget McHugh, who serves as a Psychometrician for the Assessment Services program at the center, and whose past work includes creating assessments for Fortune 500 companies, and current work includes creating valid and reliable assessments for employees and students in the skilled trades.

Traci Lepicki: Hi, Bridget, thank you so much for joining me today to talk about assessment to get us started. Would you give us a general sense of what assessment is?

Bridget McHugh: So, assessment sounds like a big fancy word but assessment is actually all around us. Say you're opening up an old copy of Cosmo or you're online and you're at a place like BuzzFeed when you take a personality quiz or one of those quizzes that tells you what drink would I be. That's technically an assessment because it's measuring something. It's assessing something. So, when we talk about assessment. We're talking about a measure.

Bridget McHugh: Typically, we're talking about a measure of what's called a latent construct. So, a construct is like a concept or idea that you defined. And when I say latent. That means that there's not a direct scale or a direct measure for that construct. For example, when you can take temperature, you can take someone's height. Some of the weight those all have direct measures latent constructs do not. There is no temperature for personality. For example, if you've ever had like an injury, a pain scale, say you're recovering from an injury or something and you go see a physical therapist or you go see a doctor, they're probably going to rate, have you rate things on a scale of one to five and we refer to that as a scale because it's going to find that latent construct of pain.

Bridget McHugh: The other key thing when you're dealing with a latent construct is, you're trying to quantify the levels on that link and construct through indicator, going back to the pain scale example, you might be asked not just how much pain you feel, you know you've seen the smiley face scale. At the zero, you have a big smiley face all the way at the 10 you have someone who is doing the complete opposite of smiling. They're like, they're in incredible pain. We're measuring knowledge and skills, we're applying the same kind of model, we need to look at the

knowledge and skills needed to do a job and then we're trying to find a way to quantify levels on those knowledge and skills.

Bridget McHugh: And usually through a written test where you get a score. Now, just as a note some companies do look at personality or they look at people doing specific behaviors we typically look at knowledge and skills. But that is something that some companies do. For example, with customer service roles, they might measure your personality traits because certain personality traits are really important for customer service roles.

Traci Lepicki: So, Bridget, based on what you just described, specifically as it relates to our work at the center in education and training for employment, where we are focused on developing tests and assessments relative to Job knowledge and skills. How do we know we have identified the correct job knowledge and skills when getting ready to create a test? What does that verification look like?

Bridget McHugh: The reason we're measuring knowledge and skills or some people who do things like personality is because those are needed in order to complete actual tasks within the job. So, one of the first steps you do is you do something called a job analysis and that's a whole other technique that I've done before, but it basically gets you a list of everything that every task that you do that the job requires and then a list of related knowledge areas and skill sets. And sometimes abilities. So, once you have this list. You then have to figure out the relative criticality to the job of each task. That's accomplished through a verification survey and the point of that survey is to get, again, we're trying to quantify something that's latent, so that latent construct of criticality we quantify that by getting ratings from people who fill out the survey and combine them into criticality index based on things like importance. For example, say you are in the skilled trades. It's very important that you understand how to do things in important safety context. For example, if a fire occurs, that's something that you need everyone to know. It's not very frequent. So, then we also have the other scale, where we look at frequency. For example, knowing how to close out a service request, you might have procedures for closing out the service requests or calling ahead of time. Some of the customer service stuff. You're going to be doing that every day or several times a day. That's more frequent so that also goes into criticality. And knowing the criticality of each task is very important because that allows you to weigh how much content is on the test. Say criticality scores for each task range from one to 10. You want for something that's a nine. You want to have more items than something that's a six and then something that's a two. So, they are creating more weight that's proportionate to how critical something is, so we're also making sure that we don't just have these are the important tasks or their critical tasks. These are the noncritical tasks. And then we just do a bunch of items for the critical tasks. You don't want things to be dichotomous, you want to have a very graded scale. That's how we verify what the job actually entails, but then also what really is more important for the job.

Traci Lepicki: So, following the verification where you've identified the criticality of the tasks or aspects of the work, how does that information feed into developing a test blueprint?

Bridget McHugh: The test blueprint is basically the document that details how many items you can have where it's directly proportional to the criticality. It's just translating that criticality score for each task into the number of items you want to put in your item bank and ideally approximately how many would be on the final assessment.

Traci Lepicki: So, your test blueprint drives how many questions, for example, how many questions you would generate in order to develop a test that would get out or assess the body of knowledge that's been defined earlier in the process.

Bridget McHugh: The reason we call it a blueprint, oftentimes when you're looking at a blueprint for house you're interested, how much of the house is taken up by the living room versus the master bedroom versus the bathroom. You can think of a blueprint, the same way, you're kind of getting an idea of the percentage of items that will come from each task. And that's good, too, because having the percentage you can translate that into the number of items, if you're going to have a 200-item test that works, but it's also helpful.

Bridget McHugh: As you're going for shorter forums, you have multiple test forums where something might be a little shorter or longer. So, you're more concerned about the actual proportion of items or their percentage of items, not just the number of items.

Traci Lepicki: So, I'm going to turn the attention slightly. We've talked about what knowledge and skill assessment is how do you verify; you have the right information that you're using to create your test. We've talked about what the function of the test blueprint is. So, let's say, now you've built the test. How do you know it's valid? How do you know what measures what it's supposed to measure?

Bridget McHugh: The first thing you have to understand when you go into validation is that how much validation work you want to do might depend on how high stakes of the assessment is, but you also have to understand that validation is also a spectrum. It's almost like a measure of the quality and how good the measure is you're trying to collect evidence of the degree to which the measure is valid. Now, there's three different forms of validation and the more evidence you have for each one, or the more you focused on at least one, the more valid. You can assume the measure is oftentimes you can do just one type of validation, but the higher the stakes and the more resources you have, the more types of validation. You can tap into. So, there's three main ones. And sometimes it's hard to distinguish them without getting too theoretical, but I'll take give you a little overview. Content validity, we're thinking about our lists of tasks knowledge and skills that we're trying to tap into. Content validity is the degree to which our evaluation technique basically measures, all of those tasks to measuring what we intend to measure and nothing we don't want to measure.

Bridget McHugh: And then also, are we not measuring something that we don't want to measure. A big challenge is with a written test where you're answering multiple choice questions. You're going to probably be measuring a little bit of reading ability as well. Ideally, with content validity, you're not getting any other skill set that isn't there, the required knowledge and skills to complete the tasks that you're concerned with. You're just getting at the knowledge and skills needed to complete those tasks. So, when you do this, it sounds more complicated than it is. A good way to get some content validity evidence is to basically have subject matter experts work through all the questions and give a rating. So, remember quantifying a latent construct. We're trying to get an idea of how much each test question actually measures that task. So, you asked, like your sometimes you might be working with content standards instead. So, this is what we want this test question to measure. To what degree doesn't measure it. And that gives you some content validity evidence. That's one form of validation. Another type of validity that's more highly related is construct validity. So, this is related to content validity and it's like, are we actually getting at the theoretical construct of interest, not just how much of it, not just how much of the content domain, but are we actually measuring what we want to measure and not some other concept. One thing you can think of is

there are certain constructs that are pretty closely related. Say you have a new assessment and there's an underlying need for mechanical ability that goes into your scores for that assessment, you might have people who are taking the first batch of the assessment then take another assessment that isn't measuring directly the same construct but it's also kind of a more mechanically minded construct you're really testing.

Bridget McHugh: Is this measuring the construct by seeing if it's related scores on this assessment are related to scores on other assessments that it should theoretically be related to. So, you're not necessarily trying to measure the same thing twice, but you're trying to measure that your assessment has a theoretical connection and supposed to have a. Then the third type of fluidity I'd like to discuss is criterion related validity. This one's very important, but also very hard to measure. Even for high stakes assessment, you might not necessarily do it when you first introduce the assessment, but it is good to follow up, if you can. The criterion is basically the end goal of what we consider success or the performance. When you're measuring knowledge and skills needed to be a good employee, you want your knowledge and skills that you're measuring that do this by if you have a measure of actual performance. You can use that and correlate it so performance appraisals work sometimes and sometimes you might have another measure of the same construct. If you're measuring performance more directly and then the key is that you want to see. Does my measure of the knowledge and skills needed to do the job predict performance in the future? So, you might hire some employees using the new assessment and then you look at how well they're doing on the job. A year later, or you can do something where you look at current employees and you give them all the assessment that you're using to hire people now. And you see are the people who tend to do poorly on their performance appraisals. Do they tend to score a little lower on the assessment? Are there people that are the low performers? For example, the people who may have not passed the assessment if they'd had to take it before getting hired. And then also, if you don't have a good performance appraisal system. Sometimes you can use more direct measures of performance. So, those are three broad areas you can focus. So, that's how validation works.

Traci Lepicki: Thank you, Bridget. I want to turn back to you mentioned the stakes involved with assessment and I believe you mentioned high stakes. Could you touch on what that topic is? What are the stakes in assessment? How does that impact maybe the type of assessment you choose or your approach to developing assessment?

Bridget McHugh: I'm thinking back to our Cosmo and BuzzFeed quizzes that are technically assessments, that's a good example of the lowest stakes assessment you can think of. High stakes is when there are negative consequences for not doing well on the assessment or positive consequences for doing well. So, you're applying for a job, and you have to take an assessment in order to get hired, that's one example, sometimes during the same position and you have to take an assessment to be placed in a new series or be promoted and then one, we do have quite a bit are credentialing exams. Credentialing exams are licensure exams. A certification exams and certificate programs you have some kind of credential that allows you to be recognized for your skill set. And so, there's a high motivation for you to do well on it. This changes things. Again, the motivation is higher. You take a cosmo quiz was you take a BuzzFeed quiz. If you don't want the answer. You don't have to share it to Facebook or tell anyone about it and then that changes how much effort you put into the assessment. The other thing though is when your motivation is higher, you are more likely to cheat or try to unduly influence your score. This means that for high stakes assessment you are very, very concerned with test security because A, that means that you kind of have to look out for people while they're taking the test to make sure that they aren't cheating. Usually high stakes assessments are proctored, someone is watching you take the assessment. The other thing that

impacts validation is that the standards of how well that assessment measures the latent construct or higher. And one of the reasons is that when there are more negative consequences for that score on the assessment, not being accurate. There are worse consequences for false positive or false negative. The false positive is someone doesn't have the knowledge and skills needed to complete the tasks for the job and the measure your assessment is not precise enough and it makes it seem as if they actually do have the knowledge and skills to successfully complete the job. The false negative is that someone actually is qualified, they have those knowledge and skills, but then the assessment makes it seem as if they are not qualified. So, the more precise the test is the more meaningful that scores and the less likely you are to get a false positive and a false negative.

Bridget McHugh: There's also legal standards and usually accreditation standards as well that you have to follow. You want to make sure that your assessment isn't advantages or disadvantages and certain groups so adverse impact you indirectly might be hiring fewer people or promoting fewer people are giving a certification to fewer people within a group. So, this is when you can look at kind of two broad areas of the possibility of adverse impact. The more common one I've seen and dealt with is the decision we talked about false positives and false negatives. Do you have more false negatives for a certain group, are there more women who look like they actually are qualified and are actually not be hired, for example. And this is where thinking about content validation and construct validation those two types of validity are also really important if you have some kind of exercise and you're measuring something that's not really job relevant. Or you're not measuring certain skill sets that are important for the job. That impacts the preciseness in the measure and increases the chances of getting that false negative or false positive. That can really impact how much your assessment holds up in court. If a group is disadvantaged, the other level you can look at to look at issues where your assessment is disadvantaged union group is actually at the exercise or test question level. So, you're not just looking again at that decision. You might be looking at specific questions and you can get a little better idea of what content you might be measuring. For example, physical ability tests, do you have an exercise on your assessment that is disadvantaging people were shorter, and is it really necessary to have that exercise that focuses more on height than the actual job relevant tasks.

Traci Lepicki: Bridget, with the complexities that you described, both within the high stakes aspects of assessment as well as different components for different types of assessments, whether they're for certification or accreditation or some other purpose. It has me thinking about two things specifically, stakeholders, the folks who are involved in different levels relative to the testing and then communication. I'm assuming communication to the different stakeholder groups, both in the test development process as well as the test administration process as well as use of the results.

Bridget McHugh: Communication is important because it puts everyone on an even playing field. I guess probably one of the most extreme examples is coming into an assessment that you didn't realize you're going to have to take because no one told you. You can see knowing more about the assessment puts people on an even playing field. So, it's incredibly important to get everyone to take the assessment in the correct manner and the correct timeline, making sure that everyone knows about the assessment, so some someone doesn't, for example, get a leg up like they know the assessment content better than someone else. That impacts fairness. This means that you really also want to tell them the timeline to complete it, and also the consequences of doing poorly. You want everyone to have the same level of motivation, approximately, you don't want some people to not realize that the assessment is going to impact their job, or that they will be able to get a certification or even something like

they know they won't be able to get their certifications. They don't pass, but they don't realize, for example, that they can only take it once a year. So, you need to tell people about the consequences of doing poorly.

Bridget McHugh: It's also good idea for people to have a good understanding of the grading scheme because it increases these perceptions of fairness and also the purpose. Communication is really important in part because of buy in. You want people to believe the assessment works and also believe that the assessment is fair and most importantly believe the assessment is valid and that its purpose makes sense for what you're using it for. Communication and getting more of that buy in is really helpful for getting people motivated to take the assessment. Seriously. And then once they see the consequences of it, not feeling like the consequences are unfair because they think that the assessment isn't really worthwhile. A couple other things in terms of communication is you want to help people understand the process, so they think it's fair and also explain the results to them in a way that's meaningful and take your audience into account. So, someone who doesn't have an assessment background, you want to explain where this assessment came from and where they did poorly, if you can, and why, and not make it too technical. One of the best ways to get buy in, is to try to get some people who are stakeholders to be involved in actually developing the assessment, I talked a lot about getting validation evidence, where you have people write the test questions. That's one way you can get managers to come in and then they understand where this assessment is coming from.

Bridget McHugh: Another element is the job analysis, and then also even writing the test questions. Sometimes when you're writing a completely new assessment instead of taking off the shelf, assessment and customizing it to the organization, you can get subject matter experts that are stakeholders to come in. A couple notes about what a stakeholder is, is basically anyone who gains or loses something based on the results of the assessment. The clearest example, these would be the test takers because those are usually the people that are most impacted. But remember how I was talking about managers and supervisors. What's going to happen to the supervisor or manager or an employee who does poorly on an assessment or someone who really need someone in the job and that person that they were expecting to get hired does poorly on the assessment, that's a stakeholder. Another thing is I talked about certifications and licensure exams. Oftentimes, there are certain functions within a business or other type of organization where someone needs that credential that certification or licensure in order to perform their job and if that employee fails that certification exam, the business is impacted. So, people within that business are also stakeholders. People who run businesses where people are certified and then also within a business. They also may have consequences if everyone within a department or a unit tends to do poorly on an assessment. So maybe you have an assessment that isn't as high stakes, but it is partly developmental. So, for example, you can have a low stakes assessment that's at the end of training. And a manager of a business unit finds out that all of his employees did really poorly on the post assessment for sexual harassment training. That might reflect when you poorly on that unit. So, there are some kind of indirect consequences and less severe consequences to sometimes that means that sometimes another stakeholder is in play.

Bridget McHugh: Another example for anyone who has had a child in school, since the know Left Behind Act or No Child Left Behind Act started or someone who isn't school in the last 20 years since the NOLA behind act really brought testing to the forefront of education. The student is the test taker in a school when they take assessment that look at their learning. But that's really supposed to be a measure of the teacher skill set, how well do they teach to the student did they learn what they needed to learn. So, they're the teachers are the

stakeholders. What this means is that there were a lot of people are impacted by an assessment and when you're planning your assessment, you're developing your assessment and figuring out who to communicate things to, you really want to make sure that everyone has a seat at the table, or at least someone who represents each one of those groups of stakeholders. When you're thinking about assessments of employees, you can think about assessments that are internal to the organization external to the organization and this helps you understand where certifications come in. So, internal assessments. We talked a little bit about selection assessments or assessments for training and this is where the assessment is somehow directed by leadership within the organization that the employee works for. This is where you have selection exams to prescreen job applicants.

Bridget McHugh: I talked a little bit about develop developmental assessments or another one, those are pretests at the beginning of training and posttests, for example, it helps you know how effective your training program is. Performance appraisals are technically a form of assessment, even though they aren't what we think of as a paper and pencil test. Qualifying assessments for promotions and what you can get from all these examples is that when you have an internal assessment that tends to be more customized and specific to the tasks needed for that particular job at that organization. So, an example is say you have an electrician here at Ohio State University. There are certain electrical tasks that are specific to our organization from things that come up in a dorm two things that come up for example of the Med Center. So, OSU if they have an assessment is going to be more custom tasks that come up in that context, an electrician that does house calls or the electrician that's on staff at a large corporate headquarters, they might have tasks that are more tailored to that organization needs. So, internal assessments are going to have more of those tasks. And that's often why people use internally directed assessments.

Bridget McHugh: Now, external assessments are certifications or other credentialing exams. When I talk about external assessments, think of anything that's not directed by the organization, so certificate programs. That's where you take some kind of training, and you get a certificate saying that you pass the training and it's usually managed by a third-party organization. Done certification programs and you can get specific certifications and types, different types of electrical work, for example, certain types of skilled trades. And then also licensure exams. Those tend to be managed more there's large boards or licensure entities and there's usually more of a legal own to there where you can't practice unless you have that licensure. These tend to be more at the occupational level rather than a specific job level within an organization. Licensure is for nurses or doctors, that's more at the occupation level. And then, this means that they tend to be more portable, and they may or may not be paid for by the employer. Many certifications or other external assessments, such as credentialing exams. They could be taken by an employee that is trying to get into an occupation. Or they know they have to have that certification exam or that other credential exam in order to get a job or do any functions of the job. So, that's a good distinction to make. And it's very important when you're thinking about planning the assessment is getting the purpose why are we making the assessment and what type of assessment you're really aiming towards.

Traci Lepicki: Thank you, Bridget. I really appreciate you taking the time to touch on this topic and information about the work that you do at the center and how it can be customized for organizations.

Fara Allen: Thank you for taking the time to listen to today! We hope you enjoyed it and will share with your colleagues and friends! If you'd like more information on this topic, you can

contact us at go.osu.edu/OhioState4Work. See our description for more details. Be well and bye for now.

©2021TheOhioStateUniversity