The difference between CBMs and CATs (and why it matters!)

Reach All Readers Podcast #240

Anna Geiger: Welcome, Alisa.

Alisa Dorman: Hi. Thank you, Anna. It's a pleasure to be here.

Anna Geiger: So you and I connected at an event at, I think it was either the Reading League or Plain Talk, where Stephanie Stollar had one of her gatherings of people, and we got to talk a lot about Acadience and universal screeners. We had a great conversation and I thought you would be a wonderful person to have on the podcast.

We're going to talk about universal screeners and how teachers should be using them. But before we do that, could you introduce us to yourself and talk to us about your history in education?

Alisa Dorman: Absolutely. Yes, I was one of the individuals that was gathered that evening. It was such a pleasure to meet you. I had been following your work, so I'm really happy to be here.

I am an educator - was, an elementary teacher. I was certified also in early childhood education. I started my first experience in an intermediate school, working with fifth and sixth grade students teaching English language arts.

It was my first recognition that students, not all of them, were readers, and that I needed to begin differentiating supports and instruction. So, opportunity presented itself for me to enter a master's program for reading, to become a reading specialist.

So in my first year of teaching, I started that program and learned a lot about how to support students more diagnostically, but I still had many students in my general education program that were struggling to meet the demands of content reading.

So I ended up, as fate would have it, after I finished my program, in a kindergarten classroom, and I saw what a difference it made to be able to start on the front ends, by providing what I would say would be direct explicit instruction in those early foundational skills of reading.

It really changed my life, pretty significantly, to be able to see that difference between what it's like to start right versus to try to catch up for students who didn't get that support early on.

I went from teaching in the classroom into a department of education. I worked in policy initiatives that were related to early childhood, and then reading specifically. I happened to be part of the Reading First Initiative that was originated out of the No Child Left Behind reauthorization of the Elementary and Secondary Education Act.

So some may know that period of time and, again, I was put into positions to be learning from some of the greats that we all know, Sharon Vaughn and the team at the University of Texas were supports for our regional center and had great influence on sort of the direction I took when it came to expanding my knowledge of research and education, particularly reading.

And at the same time I was growing my knowledge of assessment and was really influenced by the work of Doctors Good and Kaminsky and their use of general outcome measures or curriculum-based measurement. We use that assessment in our state initiative, so I really dove into understanding how data can inform instruction and supports for students.

My career from there took me a few different places. I ended up working for Dr. Good and Kaminsky at their organization. I ended up back in education policy at another State Department of Education. I ended up in education publishing for a moment before I returned back to assessment and research and the development of screener tools to help educators make informed decisions about all students in their care. So that's a little bit about my history.

Anna Geiger: So how long have you been at your current position?

Alisa Dorman: I have been here since, 2018 in the role of Acadience leadership. I've moved through roles, but currently am the president of Acadience Learning, guiding a great team who is focused on supporting educators who use our tools.

Anna Geiger: So just a little bit about my history. When I was a classroom teacher, I didn't know anything about universal screening. I really didn't know much about assessment except for giving quizzes and tests at the end of a unit, and I didn't really know how to do data-based decision making. That was certainly a foreign term to me.

I used running records to find levels and did miscue analysis. I didn't really have any concept of these early literacy skills and how we had to screen for them to see if students were on track. These were things I learned about through a course I took at Mount St. Joseph with Stephanie Stollar and the personal study I've done over the years.

So I want to help teachers who might be new to this or might be, you know, working with different tools that might call themselves screeners, but they might not be a universal screener by definition.

Let's talk about, first of all, what is a universal screener and how did they come about? We can talk a little bit about the history and CBMs and all that.

Alisa Dorman: I think it's important. I'm glad you're exploring this topic. And I think I would just add before we dive in that I think we commonly exchange terms in education across many things. So I think defining what we mean is important.

Universal screening is really, in my mind, it is a process, and we have tools that support the process. So it really is allowing us to assess all students, in a way that allows us to examine where they are at this moment in time. Are they on track for meeting the outcomes that we expect or are they struggling and likely to not meet outcomes without support?

A universal screener is for every student much in the same way doctors screen children for health risks. Universal screeners, when it comes to reading assessment, are doing the same thing. They're screening all students to see what their risk is for later development so that we can appropriately distribute resources and supports to accelerate opportunities for those who may be lagging behind.

Anna Geiger: All right, so we've laid out what they are, and I like how you talked about how it's not just to find out where they are, but so we can do something about it. We'll get to that.

When did we start using universal screeners? How did that work?

Alisa Dorman: If we want to go way back, and look at sort of the history and the development of this particular technology, it has its roots with Deno, many refer to him as the sort of grandfather of curriculum-based measurement. He was followed by others in the field, Fuchs and Fuchs, who also did a lot of work in curriculum-based measurement. Their work then was sort of disseminated among a group of people, one being Dr. Good and Kaminsky, who picked that work up and continued to expand that work. So it was in the sixties actually when people began to talk about what was curriculum-based measurement.

It has really sort of two definitions and paths that, sort of, this went. One is more in the lines of curriculum-embedded assessment that becomes more like mastery measurement, that becomes the part that is really directly connected to curriculum. I am teaching something, I'm going to have something like a unit test. It's going to measure what I taught and if students mastered directly what I taught before we move into the next sequence of, you know, teaching in the cycle and then measuring if they mastered that. That's sort of like one sort of early beginning of the way that this specific skills mastery measurement started.

But another more current way of thinking about curriculum-based measurement is really one that has evolved to a term called general outcome measurement, meaning we know the outcome, and in this case, reading, we want students to be able to read successfully at grade level. We want them to be able to read silently, comprehending for meaning. We want that as the goal or the outcome. So in the sort of curriculum-based measurement that led to what we call general outcome measurement, we are really looking at the outcome as the indicator. We're screening students to see where they are and able to predict their likelihood of achieving that outcome or their likelihood of not achieving it again without supports.

That's the area in which, I think, most people, when they come to universal screeners, are really thinking more about the latter. How do we look at this as an indicator of general outcomes? How do we screen students who are likely to achieve or not achieve that and deploy the right supports?

Anna Geiger: So for a universal screener to be effective, it has to be predictive. Therefore it has to be tested with students to see if it really does predict if they're going to be proficient readers. And it has to be testing the right skills, right? The skills that are actually predictive of future reading success.

Alisa Dorman: Absolutely. So when we think about the history of assessments and we think about the, as you said, the quality of those assessments and really what I would say is the technical adequacy of those assessments in order for them to meet the purpose of helping us sort students into categories of need so we can deploy supports - we do need them to be valid, measuring the right stuff, right?

And then we need them to also be reliable. We need to have evidence that has been studied and those proof points have been tested and found to be true, right, that this assessment does work in the way it is intended.

In the universal screener, we would want that validity and reliability to be measuring the right things and to be able to predict successfully to the outcome.

Anna Geiger: So I usually direct teachers to Acadience, just because I'm most familiar with it. I used it in my postgraduate work and also I've used it in a volunteer capacity, so I understand how the tools work,

but there are others. Can you list a few others that teachers might be familiar with that you would recommend?

Alisa Dorman: Sure. The work, as I mentioned, of Deno and others sort of penetrated, as I said, into different researchers who took that work and applied it to the development of other assessments.

Some of the commonly known assessments that are sort of in this category that have been around a really time are.. EasyCBM is a type of curriculum-based measurement that sort of fits this category. Aimsweb is another one that has been around for a long time. Even the work of Ted Christ and others out of the University of Minnesota that led to FastBridge. Aspects of FastBridge had elements of this curriculum-based measurement as well.

Anna Geiger: And also DIBELS 8 for example.

Alisa Dorman: Yes. That would be another example of a group who conducted work that would be aligned to this description of a type of assessment, curriculum-based measurement or general outcome measurement. Yes.

Anna Geiger: So we talked about why teachers should be using these. It should be so they can predict whether their students are on track to be good readers or okay readers. What grades would you say teachers should be using these universal screeners on?

Alisa Dorman: Well. Prevention is the key to many of these tools, right? Catch students early so that we can get to the business of supporting them so they have the greatest opportunity to succeed.

I would say that my answer is one, grounded in sort of the evidence of education or education practice, which is in the earliest grades, kindergarten to second, kindergarten to third grade, are certainly key sort of windows of opportunity I think. Because of that, many policy initiatives are also driving the screening of students in those grades, K-2 to K-3.

Does that mean that you can't screen in upper grades? It doesn't mean that at all. In fact, we may need brief, efficient, reliable, and predictive screeners to continue into grades 4, 5, 6, even up to middle school and high school if we have students who are still needing to be identified, and that's really the purpose of the screener.

If we can be brief and efficient and we can quickly identify those who need support from those who don't need support, we can continue this for multi grades. But the greatest opportunity to have the impact I think that a screener offers educators is in the earliest grades at the exact moment we begin to see divergent paths between those who are responding to instruction and predicted to achieve outcomes, and those who are not responding to the instruction and therefore we need to make adjustments or adaptations to give them the greatest benefit or opportunity to achieve those same outcomes.

Anna Geiger: So universal screeners typically recommend assessing beginning, middle, and end of the year. Why do you recommend those three time periods?

Alisa Dorman: I think that all of the assessments of this type have typically administered beginning, middle, and end much in the same way.

Again, back to a medical model. In the earliest years of life, if you think about infancy, the doctor has you on a schedule that's very close in proximity because so much growth and development is happening at that moment that we don't want to miss an indicator of risk.

In a similar way, applying that to the earliest grades of instruction, we want to be able to measure frequently so that we don't miss someone.

We would not want to measure just annually because we would miss a whole year to respond. We wouldn't want to maybe screen just beginning and end of year because we've missed the power of that middle of year to adjust the course of what we're doing for some students. So by measuring beginning and middle and end of year, it is again much like that well check visit model that we've talked about in health.

We bring students, all students, not because they're sick or unhealthy, but we bring them because it's a check-in to make sure that they are on track for those important milestones or markers of development in the progression of the outcome, which is reading.

Anna Geiger: And also, we know that, you know, measuring is not just a tool for figuring out where a particular child is at in terms of risk, but it's also in, in some ways a measure of our Tier 1 instruction, right? Like if we notice that loads of kids are below benchmark, what is that telling us? Can you talk a little bit about that?

Alisa Dorman: Yes. So the focus of the last question, I was really emphasizing the reason for doing that with students, but to your point, education systems that are implementing universal screening and also continuing with an assessment that supports multiple times a year, as you described, three times a year, allows us to look at the overall health of the system as well. It allows us to be able to see how students are responding, how effective is our instruction across multiple tiers, and particularly what people call Tier 1 or core instruction. It allows us to see if what we are doing in that space that everyone comes to is meeting the needs of most of our students, or if we need to target support there.

I know you will have heard this if you've been with Dr. Stollar, but the number one line of defense is effective Tier 1 instruction. We can't make up in Tier 2 or Tier 3 what we can't bring effectively to Tier 1. And so we need to make sure that we are using screener data also to inform decisions at the system level, when it comes to what we teach, how we teach the resources that we have, the training and supports educators need, et cetera, in that whole comprehensive systems model.

Anna Geiger: So I think the thing about a universal screener, compared to what I used to do, which was the Fountas and Pinnell Benchmark Assessment, is it's much it's much faster and more efficient.

You're doing every child with these one minute measures, and the measures are different depending on where they are in their educational journey. But it typically can take like 10 minutes per student, which isn't nearly as much time as Fountas and Pinnell.

Although I'm seeing some teachers in schools now switching to computer adaptive assessments where they can do everybody at once because they're all on screens. I can see the appeal of that. It's faster. But I think that we're losing some things when we do that.

Can we talk a little bit about computer adaptive tests? And what exactly they are and maybe how they're different from CBMs, the universal screeners.

Alisa Dorman: So as we were talking about CBMs or general outcome measures, they're really effective for screening students and their educational risk. So it's a really nice alignment and placement to that purpose of sort of universal screening.

I would say computer adaptive assessments have a place. Actually though I might consider it to be differently placed than in the universal screener space. That's not maybe where I would think that it has its strongest potential for contribution.

I think that they measure a much broader sample of behavior. They are looking to see through the nature of adapting the assessment where students sort of hit the ceiling and hit the floor. So they're looking to have a rich item bank and that item bank is presented to the students based on their responses to individual items.

So, let's contrast that a little bit to curriculum-based measurement or general outcome measurement where every student in the class gets a fixed form and they see all the same items for the same amount of minutes, and we're still trusting their thresholds, right? But it's in the same information so that I have the same for the whole class.

In computer adaptive assessment, we're still testing limits, but every student's experience is slightly different in a computer adaptive assessment. They may start at the same place, but based on the way that the algorithm works, it will substitute or present new items to every individual student based on, again, their response to the previous item or items.

Like you said, a computer adaptive assessment could get a whole class into a room at one time, all on computers, testing these thresholds. And we'll get some sort of broad sample of behavior and we can sort of speak to that.

And many of those items are often aligned to what people would say are the sort of standards of what they're teaching, what the state curriculum might have in mind, but over here on the other side, we're looking at all the same students in the curriculum based measurement against the same skills that are at the same time, given the continuum of sort of development of a reader.

And we're able to see how effectively we're teaching those to students, how they are responding to the things that we are teaching, and also which ones are predicted to meet that sort of outcome that we have, which is reading at grade level, independently, silently, and for meaning.

So if in the same environment, I would say that, if both are in play in the same system, I might utilize them differently for different purposes. I might utilize the screener with a general outcome measure, like Acadience or one of the others, to see where all students are and to really look for that risk, that educational risk, for achieving the outcome and look for students who are struggling and have difficulties in reading.

I might be using the computer adaptive assessment to measure many, many more things, that we could measure, very quickly in the sense of like all students at one time, in one place. It might take more minutes, but I can get it all done and then I can see where they are in maybe a lot of different variety of skills that are not necessarily all directly related to the same sort of outcome.

So that's a little bit of the way that I would say that they are different, and how they could be in the same system but utilized for different purposes.

I would say one more thing, and that is one of the features of curriculum-based measurement or general outcome measurement that we haven't spoken about that is a real differentiator is the ability to

monitor progress of students in in a way that is brief, efficient, and can measure sensitively those changes that are tiny little increments of one more phoneme, one more letter sound, one more word read correctly, for example. I think that adaptive assessments, while they can be repeated, may not offer that same level of sensitivity that we're talking about for monitoring the progress of students, particularly those in treatment receiving support so that we can make timely adjustments to that support in a greater opportunity to accelerate those students towards that outcome.

Anna Geiger: And something else that you and I talked about before we, before we hit record, we had talked about how with those universal screeners, like Acadience, kids actually have to produce a response, right? So when you're giving a bunch of words and they give you the first sound let's say if they're in kindergarten, they actually have to produce that. Versus with a computer adaptive assessment, they're just clicking buttons. I know from my own kids who come home having done computer adaptive tests, "Oh, I just didn't want to do that part. It was too much reading, so I just guessed."

Whereas with the universal screeners, they're right in front of you and they have to read this for one minute and they have to do it right? It's not a long time. I think that's what people need to realize. This is not a long time, and yet we know from research that they give us so much information and that even the oral reading fluency has a high correlation with comprehension, which is fascinating.

I think if a teacher is using a CAT and they're being told to form instructional groups based on the responses - what would you say to that?

Alisa Dorman: Well, I think it is, my experience having reviewed assessments as part of policy initiatives in two state departments of education. I'll bring that lens, to my voice right now.

I find that they are harder to make those refined decisions for placements into the structure. They might be good at separating risk from no risk, but when it comes to differentiation and really separating students into likenesses or groups that are similar, I find that curriculum-based measurement or general outcome measurement has stronger characteristics or features that support educators to do that.

I think sometimes computer adaptive assessments are used to report out. Somebody says we must screen, and I can assess all students at one time. I can get it done in one class period and I have my answer for reporting.

What I often find when I talk to educators is they don't use that assessment after that point. There isn't anything instructionally relevant there. There is a sort of a one and done type assessment experience, and I think that maybe what CBMs or general outcome measures offer is a lot more of that refinement. The ability to, as you state, hear students produce tasks to be able to get to item level.

When I mean item level, I don't mean an item, like pulled out of a bag of items. I mean like the actual item, a letter sound, a phoneme, a word. And to be able to look at that and in some way, very nearly diagnostically evaluate where the student is, particularly in these early measures so that we can get more refined. We're not just going to say they need help in a big foundational skill, let's say, like letter sound correspondence. We're going to know exactly which letters because we're going to have tested the threshold of as many letters as possible and as many experiences of items that produce sort of that sampling.

And we're going to be able to know whether it's confusion in particular consonants or vowels. We're going to know if it's a one-time error or if it's a repeated error. We're going to know if they are accurate,

but not fluent, fluent but not accurate, or both. We're going to know a lot more in production-based assessment tasks, maybe than we would if we just have students who are taking items, where we can't really know if that item is mastered in a production way or it was lucky selection as you mentioned, like through guessing.

Anna Geiger: Yeah. And also I think Acadience and the other tools really lend themselves to helping you when you see where the strength or weaknesses are, knowing what diagnostic tool to use to really dial in. Right?

So maybe they're not doing well on nonsense word fluency. We might notice very quickly they don't know many sounds, so I'm going to give them a letter sound diagnostic assessment to find out exactly what the problem is. And I think, like you said, sometimes with those CATs, they might not even have received items on a particular skill.

Maybe it just pushed them right past, or they maybe had one. And so they might be good for reporting purposes, but in terms of data-based decision making, that's where they're lacking and that's what we really want to be focusing on.

I'll be diving into that hopefully with some other guests coming up, but thank you for helping us really lay out the difference between CBMs and CATs and why it's so important to do universal screening.

Alisa Dorman: Yes, thank you. I think you are exactly right. One of the features I would say of any good assessment is its ability to provide data that educators can use. Something that empowers the educator to really make decisions and to take actions.

We all know educators are here to support students to better outcomes, and if we can provide the right tools, particularly measurement tools, to sort of help them have the feedback at the individual student level, which we talked about, but also at a systems level, are the tools I'm using to teach with, is the training I'm providing and support to teachers, are those things coming together in such a way that students are meeting expected outcomes?

Assessment and screeners allow you to be able to see that, to be able to see that very clearly. I will be anxious to hear, as you continue to have conversations with others, I know they will all agree that data-based decision making is a critically important contribution assessments can and should be making. If we are doing tests for compliance, we're doing it for the wrong reason. If we're doing testing or screening to inform instruction and to make data-based decisions, we're definitely on the right track.

I look forward to hearing more.

Anna Geiger: Thanks so much, Alisa.

Alisa Dorman: Thank you.