>> DERRICK VARGASON: Hello and welcome to today's webinar Student Engagement: Are Your Test Scores Valid?

My name is Derrick Vargason. I will be your host today. On behalf of NWEA and Ed Week I want to thank you for joining us for this webinar.

Before we begin, now's a good time to review some of the technical aspects of today's presentation. Please check the audio setting on your computer as well as your speaker volume settings. If you are having audio trouble. If you are still having issues, please see the detailed audio trouble shooting file, available in the handouts folder at the bottom of the console.

There also are some other icons located at the bottom of the webinar console that open some additional feature panels. You can read about today's speakers in the bio panel and click the handouts panel to download a copy of today's slides. To submit a question for our speakers to answer during the Q&A session, type them in the Ask a Question Box, located above the Chat Window. Finally, an on-demand archive of today's presentation will be available online in the next 24 hours. Both the archive and a free-to-download version of the Power Point slides will be accessible through edweek.org.

With that out of the way, let's go ahead and get started.

What if? What if a well designed test isn't enough to guarantee a valid score? Our research indicates that the level of student engagement with a test impacts the score. How would educators recognize or measure that engagement especially at a school or district wide level?
Today we'll cover why student engagement test results, which students tend to be rapid guessing and how your assumptions about student growth can be. You'll learn about how map growth assess when students are guessing on a test. I would like to get feedback from the audience. Using the poll function respond yes or no to the following. Do you see disengaged test taking during assessments a the your school? In other words, are you seeing kind of students guessing on tests?

We'll pause for a minute for you to submit your response.

(Pause.)

>> DERRICK VARGASON: Let's go ahead and look at that response and not surprisingly it looks like just about everyone here is experiencing that phenomenon in their schools. That probably explains why you are joining us today.

At this point I would like to turn it over to my colleague, Steve Wise, Senior Research Fellow at NWEA. He has published extensively during the last three decades in applied measurement with particular emphasis on the psychology of test taking. He is the foremost researcher in the nation on student assessment engagement.

Steve, take it away.

>> STEVE WISE: Thanks, Derrick. Welcome to everybody. Today I'm going to talk about this problem of disengaged test taking. I'm going to approach it in three sort of phases. First I'm going to talk about our assessment, what we use at NWEA. It's a little bit different than other assessments.

Then I'm going to talk about just start addressing probably the topic that is most interesting to people listening. What can we do about it? What do we do to rye to minimize? It's something we don't want to observe.

Now, MAP, we are going to talk about adaptive tests and focus on something called MAP Growth. It's a computerized adaptive test. The computerized adaptive test is a special sort of test that adapts or adjusts to the achievement level of the person taking the test. It turns out we can measure most efficiently when we give people moderately challenging items. Unlike a fixed form test where everybody gets the same test, an adaptive test will adjust. If you are doing well, it gives you harder items. If you are not doing well, it gives you easier items. It takes the difficulty of items into account when scoring the test. It is designed to provide the most sort of bang for your buck. The most information per item. That's the advantage of it. And so MAP Growth is an adaptive test.

And it produces a score. Not greatly important that you know this. Something called a RIT score. That's the thing we put it on, like SAT, like an vertical equal interval cross grade scales. We can measure different students and see how they are doing at any one point in terms of achievement level but allows us to track growth over time. As they move through a particular grade or across grades you can see the RIT scores go up reflecting the amount of achievement growth they have demonstrated.

And this is a graph. I hope you can read this. It is a graph that shows typically how an adaptive test works. What an adaptive test does is you may recall I said that we measure most efficiently when we give somebody moderately challenging items. The problem with that logic, though, is if we knew what the person's achievement level was we wouldn't need to test them. What we need to do as part of the adaptive test is figure out where they are. So you start out getting items. If you are doing well, you get harder items, not doing as well, you
get easier items. Essentially the algorithm that is selecting items reacts to how you are doing on the test.

In this particular case you can see a person started off on this RIT scale around 200 at the beginning of the test. They started doing pretty well on the test. And they got up to around the 220 level. Then you can see they kind of, we kind of target where they are and start getting a lot of items, items correct and incorrect. So what we call our provisional level achievement level or provisional RIT score stays around 220. And the items we give them are matched to that and they stay in the same spot. It locates the person and then it throws a lot of items at them at where they seem to be. That a lust us to get the most precise measure possible of what they know and can do.

So this is how it typically works. And when Derrick started framing this, he talked about the idea of the what if question. And there's no, we spend a lot of time and resources making MAP Growth. It's a pretty good assessment. It can give a pretty good idea of what a student can do. But that's not enough. Unless the student engages in the task of taking it, we are not going to get a valid score.

And that probably makes sense from a common sense standpoint. When we don't try our hardest on a task we have we probably don't do as well as we could and it probably doesn't reflect what we are capable of doing. That's what we see in MAP when somebody disengages. And so what happens as I mentioned a moment ago, if a person disengages, the score you are going to receive is probably an underestimate of what they know and can do. And it is going to create problems. What we are trying to do is give information back to educators about where a particular student is and it is misleading if it's too low. They may feel that the student doesn't know as much as they really do or have cared to show you.

This can have implications such if the student scores too low they may be put in an instructional program they don't even need. This might be compounded by about the fact that somebody who needs to get into that program because somebody else is there. Time and resources are wasted in this type of problem. Disengaged test taking when it occurs can really distort the scores and really mislead people. That's why we need to deal with it.

Now, this is an example of a student who again like in the last graph, adaptive test. You see a person who started out around 240. They are pretty well targeted. Their items, their provisional RIT score stays around 240. They keep getting items about that difficulty. Along the bottom, there are some red boxes. The red boxes indicate things we identified as disengaged or what we call rapid guesses. I'll explain more about that in a moment.

You can see for the first half of the test they did pretty consistently. There were a couple rapid guesses, not very many. Suddenly they start showing a the look of rapid guesses, a lot of disengaged responses.

As you expect, their score started going down. Now, why does this occur? One of the reasons it occurs is, one of the features of an adaptive test, if you give somebody an item that is well targeted to you, you have about a fifty-fifty chance of getting it correct. But if you are rapidly guessing, just throwing down answers to a multiple choice test, your chances of getting that correct are far lower, maybe 20, 25 percent depending on how many choices there were. So you are probably going to get those items wrong which will tend to see your achievement estimate drift downward as shown in this graph. This gives you an idea of how far that thing
can go. That's a sizable amount in this case. The estimate they got at the end is far off from
the 240 we would think we would say they are at at the end of the first half of the test.

So to summarize, the reality is disengaged test taking occurs with MAP Growth. I'm not
going to say it happens terribly frequently. Most students try hard on our assessments and
get good scores. When students are disengaged we like to know about that because the
scores we are going to provide to educators are going to be off, misleading. We would like to
know something about that and do something about it. About you this problem isn't just
constrained to MAP Growth. It occurs with any assessment you take. If the student is not
engaged you are going to see this problem. It doesn't have to be an adaptive test. It can be
any test. It doesn't have to be computer based. It could be a paper an pencil test as well. It's
just our computerized adaptive test, we have a chance to get a better idea of what is going on.

And I say NWEA is committed to addressing the problem. We realize that this problem is
there. And we have sought ways, I have been working on research in ways to do something
proactive about it. That's what we are going to talk about next.

So how does this all work? How do we know this? I mentioned the term rapid guessing.
What does it mean? It turns out when people are engaging in multiple choice tests and they
are disengaged, they tend to speed up and give an answer before they have even had time to
even read the problem. They give you a sense of perspective on this, you get items for which
somebody, people on average maybe take 40 seconds to answer a particular item. If you get
an answer within two or three seconds, you have a pretty good idea that they didn't spend
much time on that and didn't try to really understand the challenge posed by the problem and
answer it in an effort-forth fashion. So what this means is, you can look at every item
response and make a judgment about whether it seals to be very rapid or not. And rapid
guesses are the first part or the first type. The second part are something we call solution
behaviors. As far as we know, they try hard to answer the question. Rapid guesses, solution
behavior. If you aggregate that information across all of the items a test taker looks at, we can
calculate an index called response time effort or RTE. That gives a general overall way of
assessing or measuring how hard the students seem to be engaged. How engaged they seem
to be on this particular assessment.

And so using RTE, we are able to amount of amount of test taking effort the student exhibits
during a MAP Growth event. This gives us a sense of whether they are engaged or not. It is
important to note, remember, I said you can look item-by-item to responses. It turns out that
people are rarely all disengaged. They tend to be disengaged for part of their assessment or
part of their test event. This gives us a sense of being able to understand not overall how
engaged they were but instances in which items they seem to be disengaged from. We can
use that information as I'll talk about in a little while.

Well, I mentioned before this doesn't happen terribly often but let's get a sense of its
dynamics. Now, if we see one rapid guess during a test event it probably doesn't distort a
score too much. But we tend to use a rule of thumb. Once we see at least 10 percent of the
responses being rapid guesses it probably distorts the score to a degree that we ought to be
cconcerned about it. And so we can look at the number of students in a sample that have at
least 10 percent or more of their responses being rapid guesses. Here, this gives you a sense
for just one of our samples of one of our districts in terms of how they are doing in math. You
can see that the rate at which this occurs varies by grade. The lowest bar is third grade. Goes
all the way up to eighth grade. You can see as you go up through grades, the percent of students who show this behavior increases from about oh, third or fourth grade, about 2 percent in this sample. Up to seventh and eighth grade, seven or 8 percent. It tends to go up with age. If you think about it and think about what age is this representing, kids going from elementary school up to being middle school and being teenagers, the notion that kids may become a little bit less compliant doesn’t surprise many parents or teachers. So this mirrors what we might expect to occur.

Now, if you look at reading, we see the same trend of an increase from third grade up to eighth grade, but it's at a higher rate. This is one of the things that we know, so what do we know? We know as you go up through grade disengaged test taking seems to he increase. We also know that it tend to be higher for reading than for math.

Why would that be? That is probably, my studies of rapid guessing behavior suggest that when a student sees an item, they decide whether they want to engage with it or not. One of the biggest factors of whether they engage is how much work does this look like? And reading tests, reading items tend to have more words in them, more reading associated with them. That seems to be a large determinant of when people judge the amount of effort needed for this item. Items that tend to have more reading tend to invite more rapid guessing.

Again it goes up by grade. It goes up some subject areas, sort of elict this behavior more than others. There are other factors we have been learning. I won't go into many of them but for instance it occurs more often for boys than for girls. I like to joke and say for some reason it doesn't surprise people when they hear that. Boys are more likely to give rapid guessing or be disengaged during a test, and that's what we observed.

Not only can it affect the score from a particular test event, you may recall earlier I said that one of the things we use MAP Growth for is to study student growth through time. What this requires is more than one administration of MAP Growth. And so disengaged test taking, though, can affect MAP Growth in two distinctly different ways. I'm going to illustrate that right now. On the graph you are observing ugh you can see results for two different administrations, time one and time two for a particular student. They start out at time one, they had a RIT score of about 200. At time two they had a RIT for of 210. Simply we would compute it as a growth of ten points.

Let's imagine this is what is really going on. The achievement level of this student at time one is 200. At time two, it's 210. What if the student said at time one, I don't feel like engaging very much in this test today. And they decide to show a lot of rapid guessing behavior. What is going to happen? Their time one score is going to tend to be lower. Now, if they are fine at time two, what is that going to do to growth? What it is going to do is exaggerate it. That is, instead of 200, maybe they get a 194, a lower score than they would have otherwise. And you can see if you compute growth now between time one and time two, it is no longer ten points of growth. It's 16 points of growth.

You can see in this case that growth is overestimated by the occurrence of rapid guessing at time one.

Well, what if it occurred instead at time two? What if it occurred, the person was engaged during the first test event where they got the 200. But the 210 test event they decided to be disengaged. The opposite is going to occur. Instead of 210, they get a 204. Now instead of ten points of growth, they appear to have only four points of growth. These are examples. By
the way, the magnitude of these can be far greater. In fact, if somebody was really disengaged sometimes we see a phenomenon called score -- the test score at time 2 seems to be lower than test score time one, which appears that the student knows less at time two than at time one. It is confusing to educators and doesn’t inspire confidence in what we are trying to do. It is often due to disengagement at the second of the two times.

So when we are talking about growth, disengaged test taking can either exaggerate in a positive direction the amount of growth you observe or it can make it lower than you expect. So in that way, growth scores get distorted as well when we have a disengaged test taker.

This notion of disengaged test taking is a problem that can occur in an amount that is predictable, but it can reach a level that is concerning at times. It invites the question: What can we do about this? We have been studying this. I have been studying it for more than a decade now. We have some pretty good answers for things that we might be able to do.

The first thing we are going to do and in fact we are going to be doing this weekend, that’s how soon it is coming up. We are going to provide feedback to the people who read score reports. If a score has been affected to a meaningful amount by disengaged test taking, we make sure our score reports interpret that to the person. It is a way of giving them feedback. Be careful interpreting this score. It very well may be off by a bit.

And we can also get a sense, also have a way of statistically estimating -- Derrick is going to illustrate this in a little while. We can statistically estimate about how much distortion occurred. And we will report that as well. Not only distorted, but here is about how much it was distorted.

Both those things are useful to give to educators, but what they do is, there are things we can apply after a test event has completed. In a lot of ways a more exciting possibility is, what if we can do something during a test event to reduce or preempt its occurrence? That’s the second thing we can do. Since this is all based on response time, how long students spend on test items, our MAP Growth test now can monitor this. And watch for the beginning of rapid guessing to occur. If we see a student start to engage in rapid guessing behavior it could identify, or notify, rather, the Proctor in the room. The Proctor also has his or her computer screen that shows things about the progress of the student through the test. It could alert the Proctor that you know, this particular individual seems to be disengaged and what they hopefully can do is intervene somehow. Go over and talk to the student or encourage them or somehow figure out if there's something wrong. If the student is not feeling well or something that we might be able to address in order to get that student reengaged.

Now, we have just started doing that this fall. I can't report any results for this yet but I've done this sort of messaging before direct I to students in some college testing I've done. It seems to have a really large effect on the behavior of the person taking the test. They kind of get the idea that okay, I need to reengage. It really sort of gets them back on track.

We are hopeful that that is going to occur with MAP Growth as well.

Now, the other side of it, though, and we have a lot of questions, a lot of the pre-conference questions about: What can we do about this? There are a number of things we can be thinking about. I guess to answer this you need to think sort of there are things you can do before a test event, during a test event and after a test event to effectively deal with this. You know, before the test event, just try to make sure the teachers, Proctors, parents and students are mindful of this importance of engagement. It may sound obvious that if a person is not
engaged, that they probably won't get a good test score, but it could go into things like the planning for the room where the test occurs, are there distractions? Time of day, maybe you don’t test right before lunch, maybe at a different time. We find that the earlier you test in the day the less likely you are to see disengaged test taking in the form of rapid guessing.

There are things you can be thinking about, what you tell students about the importance of the test and why they are taking it. All these things factor in to the sort of engagement you will get out of students.

Now, during the test you can make sure it is promoted. Make sure that the teacher or Proctor -- usually it's a teacher, will clearly convey the importance of this. This is a serious thing that we are going to be keeping an eye on with engagement. Activity on the part of the Proctor, moving about the room, keeping an eye on things is a good idea.

When we use MAP Growth, one of the things we find particularly useful is to help enhance the student’s sense of ownership over their score. Remember, I mentioned about these growth calculations. One thing we can do is work with students beforehand. Have them set their own growth goals. You can say okay, a student like yours may grow ten points over the next eight months in school. And the student can decide, maybe they say can you do better than ten? Can you make ten? Give them some goals. If they have a goal in mind when they are put in a test event they are interested in seeing how they can do. Not a game, per se. But they have ownership over this and so it's more meaningful to them. We found just enhancing the ownership of the students over their own test taking seems to make a real difference.

After the test event, basically we talked about the score reports and what we do. Paying attention to any information you can get about how engaged the student was during the test. And taking that into account when you interpret scores. If you see a particularly low score, much lower than you thought, it makes a difference of whether that student was engaged or not. If the student was disengaged, you might say that probably explains what I'm seeing. If the student seems engaged you may worry more about whether they are really learning as well as you thought they were. It helps you. It's additional type of information that you can provide educators’ I ways of interpreting what they are seeing in the score reports.

That's some basic ideas. This can go on and on. You can be talking about this, as people become more aware and mindful of test taking engagement, they can in fact, it begins to affect the way teachers behave, the way they plan to assess, the way they behave themselves as they are introducing the test to the students. Making sure that students get the message that this is something they should take seriously. The students get a sense of if I disengage on the test people will notice. It's going to be somehow noted in the score report that I wasn't.

So hopefully just that general awareness of the importance and the fact that engagement is being assessed is going to help improve in ways that may decrease the amount of disengagement that occurs.

At this point I'm going to turn things over to Derrick. Derrick is going to give you a little bit more information about these new features we have in MAP Growth that I have described in terms of Proctor notification and the score reports. Derrick?

>> DERRICK VARGASON: Thanks, Steve. I want to start by taking a look at the Proctor console which Steve mentioned earlier. The view that you are seeing on your screen now is the view that a Proctor has while a student or group of students are taking a MAP Growth assessment. It's the point of access, point of control during the testing process. During the
live test session the will Proctor will receive notification when a student submits three consecutive answers determined by the system to be rapid guesses. The top right of the Proctor console, a box labeled students disengaged illuminates to an orange color notifying the Proctor of the student’s who is doing the guessing behavior’s name. The Proctor can choose to click on that notification to gather more information.

This is what they see as they expand that information. This dialogue box is telling the Proctor the student name, how long ago the rapid guessing behavior happened. Maybe if they were away from their console monitoring the room they have some information about when the behavior was starting to happen and in green text a recommendation to encourage the student to make their best effort.

The recommendation in green is optional activity. As Steve mentioned it's part of a corrective best practice to keep students engaged. During our research in user testing of the students in these features, educators consistently tell us that Proctors should receive a prompt to encourage the students to do their best. It's a suggestion not a requirement. The MAP Growth assessment is not relying on the Proctor to encourage the student. The test doesn't pause nor does it feel any different to the student doing the assessment.

Take a look at reports. What you are seeing here is the student profile report. The student profile report, how this information is captured for teachers starting November 11, which is coming right up and beginning with tests delivered this fall, the student profile will indicate if the student exhibited rapid guess behavior on a test. There will be pop-up boxes within the student timeline view of student assessment results built right into the student report. This will show two new metrics, first, the percentage of disengaged responses. The percentage of items on the test answered too quickly for a student to have applied effortful behavior to answer that question. If 30 percent or more of test questions are marked as disengaged the entire test score should be considered invalid and a retest might be appropriate for your student at this point.

The second metric that is being introduced here is the estimated impact of disengagement on reiterate. That shows how much RIT points higher the student might have scored if they had been fully engaged in the test process. With the RIT score of 200 estimated impact of negative 2, the student might have the potential to score 233 if they were fully engaged during testing.

This information helps educators determine how wealth test event captured what the students know and what they are ready to learn next. Like Steve said this impacts all tests not just MAP. Isn't it better to know for sure how engaged your students are?

Before we get to the discussion portion of today's presentation, we did receive a few questions before the webinar. Asking if we were able to correlate test taking engagement to other social emotional learning factors that have an impact on students' success. So building on Steve's work, our researchers made significant process on this work. Jim and Nate partnered with the Santa Ana unified School District checked self management and check regulation, the result won first place in the social emotional assessment design challenge sponsored by the collaborative for the social and emotional learning. I encourage you to check this out at NWEA and read more about this from Ed Week, regarding this research from our team.
Earlier we asked how many of you were seeing this rapid guessing in the schools. We encourage you to place into the group chat window some of your approaches that have worked for you in your schools and your district to reduce this rapid guessing behavior. So we have the chance to learn from each other in this environment. So the group chat window us available on the bottom of the console. Pop it open with a click and put in your suggestions of how you are addressing this behavior in your schools or district. In a way that you found to be effective.

That brings us to the end of our content and the presentation today. We will turn it over to questions. We do have quite a few coming in. So again thanks for joining us. I'm going to transition right into questions.

So Steve, there are a number of questions that have come in about over testing and test fatigue. And how much of that might be playing a role in disengaged test taking behavior.

>> STEVE WISE: Well, I think it might play a large role. You know, we use tests to get information from students. And everybody thinks that their test is the most important test. And I don't think people often attend to the degree to which testing occurs. Again, I'm not somebody who is saying we necessarily over test and we need to cut back, but we need to be at least mindful that the amount of testing that occurs does matter. And one way I sort of characterize this and we are familiar with the opt out movement where students are being told, parents are essentially saying I'm not allowing you to assess my student.

I view rapid guessing as a way of essentially opting out by the test taker essentially saying I choose not to show you what I know and can do on this item. And they may reengage or not on the next item, but that notion that they are just not choosing to engage, this opting out behavior is something that they are actively doing. I think it's also more important not only the number of tests but the type of tests. Tests that provide information about the individual student and their performance and their growth is probably going to be more meaningful than a test in which they don't get much information back and I'm going to use an example. I'm not trying to pick on statewide accountability tests, but as importance tan as they are, if you picture from the student's point of view they are being asked to take a test and they are not going to get -- they may not get a score and the scores may not come back until the summertime for tests they took in the spring. So the instructional value for a teacher of that student's performance is pretty low. And so while it is understandable why we have these tests we also have to recognize that from a student standpoint they may say well, what's in this for me? And for those few students who disengage, they may see it as something that is not particularly useful to them. This notion of looking at it from a students, the test taker's point of view is a good way to think about it. Is it meaningful to them? For tests that have a greater value from their perspective, the less likely you are to see disengaged test taking. So it is the tests and how meaningful they are to students.

(Voice breaking up.)

>> STEVE WISE: That's a great question. In fact, when we choose, when we choose to define rapid guesses, to do that you have to come up with something called a time threshold. In other words, what is too rapid? We try to intentionally do this in a very conservative way such that if we say that a student responded too quickly, we want to be sure that we can infer that it occurred, reflected something that was not effortful or disengaged. That means that
when we are in doubt we won't assume it is disengaged. So when you have a student who is
... when you have a student who is ... hang on a second.

When we have a student who is engaged but missing quickly, we work very hard not to have
those misidentified as rapid guesses. So we have statistical ways of looking at the behavior for
large numbers of test takers for particular items. We have a pretty good way of identifying
when we are starting to get into times that are very high performing student could give an
effortful answer.

Again, it is part of the art of this, of trying to understand what is too quick and we err on the
side of being conservative even at the expense of identifying, not identifying some non-
effortful responses because we don't want to take a high performing student's response and
say they weren't really trying.

(No audio from Derrick's line.)

>> STEVE WISE: One of the questions we got had to do with how much rapid guessing is
attributable to fatigue. That's part of it. Certainly when we make tests we try to get a sense
of how many items can you productively give somebody within a time period. And at some
point students get tired. I mean, it is commonly seen that the farther back in a test you tend
to go, the more likely rapid guessing is to occur. So in part it could be fatigue. Part of it just
lagging attention to the test. I'm not sure how different those things are in reality. But that
notion of fatigue does factor into it.

I like to view it from the perspective of what does it take for us to get a good score, a score
that really indicates what the student knows and can do. We have to be mindful of any
fatigue factors that can factor in.

It's one of the reasons I mentioned earlier that we tend to do better if we test earlier in the
day. Later in the day you are more vulnerable to things like fatigue.

>> DERRICK VARGASON: Hi, Steve, thanks. A question is: How does the computer
determine that rapid guessing has occurred? Is there a minimum time limit per question?

>> STEVE WISE: Well, it depends on the question. If you are giving somebody a fairly short,
let's say mathematics question, on average it is not going to take people very long to answer
that. And so if you think of generically the response of, how long would it take somebody to
even read the item, comprehend it, work out the answer and submit an answer? You try to
think of it from that sort of perspective. If you have let's say a basic math question it doesn't
take very long for all that to occur, for a higher performing student.

The time threshold of how you define rapid is very short. It might be under a second.
Whereas if you give somebody a reading passage and it would take anybody at least a dozen
seconds to read, even if a quick reader. Then understand the question and then figure out the
answer and submit it, you might say that that time threshold might actually be several
seconds. Could be, oh, as many as five, six, seven seconds. So the threshold we determine
individually for each item is in part a factor of how long it typically takes students to answer
that item. Each item will have its own threshold. That depends on sort of what it needs. We
found that that correlates pretty well with how long it takes to even read and understand the
question.

>> DERRICK VARGASON: Thanks, Steve. The questions are really pouring in now. You
mentioned earlier a percentage of disengaged responses that would trigger a
recommendation to retest. Could you restate what that would be?
STEVE WISE: Well, one of the things that -- and I'm trying to think of this from the standpoint of a what I'll call a responsible assessment organization. You know, at some point the score is going to be so disengaged or distorted by disengagement that you wonder do we even provide it? For instance let's say students engage with half of their items. You are probably going to get a score, but the score probably isn't very valuable.

And so we have a choice at this point. Do we provide the score we know is not valuable and even information that suggests that? Or maybe we ought to at some point say we are not even going to provide a score because we didn't get enough information to provide a score for this student and we've decided to do that latter strategy.

Essentially what we are looking at, if a student disengages more than 30 percent of their responses, their item responses we are going to say that student didn't give us enough information that we can trust that we are going to develop a score. So we are not going to provide you a score. And encourage them. I mean, the school, to retest that student and maybe try to figure out why they weren't engaged. The motion of what do you do in the worst circumstances is a tricky one. Do you give a score you don't trust? At some point do you say let's not even give you a score. You just weren't engaged enough to have that happen.

That's what we do.

DERRICK VARGASON: Steve, there's some questions about if we've done any research on rapid guessing behavior on item types beyond multiple choice.

STEVE WISE: That's a great question. Justly, in fact. We have another assessment we use, the OECD's Piazza based test for schools. One of the -- let me back up for a second. The MAP Growth assessment is a multiple choice assessment, where a student has to give an answer to move on to the next question. In other words, they can't not answer. They have to give some answer. Rapid guessing is a way of determining whether they are moving too quickly.

In the OECD's test, it's a computer based test, not adaptive, but a student can omit answering all together. Or there are some items that ask them to provide a constructed response. You have to type in a reason for something, or why they think, or what is going to occur in some situation.

What we found is we can also use some of the similar dynamics to understand when they are disengaged in that as well. Now, if a student omits an answer, for instance, just omitting doesn't necessarily mean they weren't engaged. They may have looked at it seriously and just said I don't have any answer for you.

But what we found is if this behavior occurs quickly -- so if they omit the answer within a few seconds and move on, we find that is as informative as a rapid guess. We call that a rapid omit.

In a similar fashion, if a student takes a constructive response item, we've also found that sometimes students put nothing or just a little bit they may, instead of giving a full explanation, they type in a few characters. Maybe nonsense. Maybe a few words, but not very much. If we find that what they do is far shorter than what Anne gauged test taker tends to put and it occurs quickly to a rapid, we'll call rapid perfunctory response, that also is indicative of disengaged test taking. They don't say nothing quickly but they say very little quickly and something we decide doesn't really reflect them being engaged as well. We've
done research. We have a publication just coming out, just in a about a month or so, that describes this sort of behavior. So we can extend the idea of disengaged test taking beyond multiple choice items into some new things. That's a brand new development we are working on.

>> DERRICK VARGASON: Hi, Steve. I think we might have moved quickly through this during the presentation. Could you describe quickly how the Proctor notification functionality works again and what the conditions are for that?

>> STEVE WISE: Yes, the Proctor notification is essentially, picture a room or lab and students are at their computers working away. The Proctor also has a computer with a screen that shows where they are -- lists out the roster of the students and sort of where they are in the test, what question they're on and the like. It is a way for them to keep an eye on the progress of the students.

If the student shows a pattern of disengagement. We may change but for now the criteria we are doing, if a student gives three rapid guesses in a row it is a signal that the student seems to be pretty disengaged right now and it might be worthwhile to do something about it. A message gets sent -- a notification is sent to the Proctor's screen saying there is a disengaged student. It notes on the roster which student it is.

In most cases in our testing, the Proctor is going to be a teacher who knows that student and will probably know the best way to intervene with him. And hopefully they will then get up and go over to the student and do something that sort of gets them reengaged. Then they can dismiss that notification. Dismissing the notification is important because then if the student disengages again the Proctor could be renotified. So that is a basics of how it works. The Proctor has a little bit more information than they would have otherwise.

That doesn't mean that a Proctor can't look around the room and see if a student appears to be disengaged, but this gives them notice of something they may not have noticed, students moving quickly through test items because they are not really trying.

>> DERRICK VARGASON: Is there any correlation between students performing below grade level and the amount of disengagement on the test? Has that been observed?

>> STEVE WISE: It could. I have two answers on this question. First, if it's a fixed form test, everybody gets the same test, if you are a low performer, those items might be very, very difficult for you. At some point the student may give up and say I can't do this. It's too hard. So items that are too hard for them may invite this sort of giving up behavior. Now, also I'll say something like MAP is an adaptive test. The MAP Growth, what it does is if you are struggling it will give you easier items. If a student is well below grade level it tries to move to where the student is and tries to give items that are of appropriate challenge for that student.

In theory at least an adaptive test like MAP shouldn't see that behavior where a student gives up. We are trying to reach them and give students that are matched to their level of achievement and hopefully they will stay engaged longer. One of the benefits of an adaptive test, it allows the test adapt to where the low performers are. And sometimes the high performers are a problem. MAP Growth could give them harder items and hopefully be challenging enough to keep their engagement as well.

>> DERRICK VARGASON: A related question to that, Steve. How do we support those high flying students in the top percentiles when they are getting discouraged because they don't see the same amount of increase of growth like the lower scoring peers?
>> STEVE WISE: Well, okay, this question suggests that there's something we typically see. I think I understand the source of the question because of one of the things we've noticed as we look in terms of growth, we have test norms. We have ways -- we test 10 million students a year. So we get a lot of data we can study and say what happens in terms of growth on average as you go through grades? One of the things we've noticed is that growth in most subjects tends to level off. That is, most of the growth of students on average see from let's say third to are fourth grade is not as high as say seventh an eighth grade and maybe in turn not as high as you see between tenth and 11th grade. There is a notion that as a student gets up to a certain point they don't tend to show as much growth. At this point I would like to say, does that mean you don't have hard enough test items? We are, but the amount of growth take people seem to see given the scale we use isn't really occurring.

I also harken back to something I said a moment ago. Though MAP is trying to give the most challenging items you can to high performing students and trying to give them something that allows them to show the sort of growth that they can, a fixed form test has something we call ceiling effects where if you are really bright relative to that test, you just top out at 100 percent an it won't be able to show you moving much farther ahead.

An adaptive test in principle allows you to get a better sense of that. Hopefully an adaptive test will allow that will sort of assessment to occur, assessment of growth of high performers better.

It is something we need to be aware of. I'll say one more thing. When you think back to things like No Child Left Behind, that type of test was focused on who is proficient. And their focus was are you proficient or not? It suggests that that may have had an effect on higher performing students because they may not have gotten the attention they had before. That in turn could have affected the amount of growth they may have seen. It is partly a function of the test and the behaviors that occur out of educators in reeling to that test.

Now, I'll stop here, though. We can go on and on about this topic but I'll stop here.

>> DERRICK VARGASON: I think we have time for one more question here, Steve. And this one comes in about reports. Is there a report that schools can run to find out the percent of students disengaged and the impact of that disengagement on their RIT score? This is probably talking about a broader level than just the student level.

>> STEVE WISE: I don't know if there is yet, but we are going to get that. I mean, it is -- it's logical that we would want that because teachers see reports about individual students. But these scores get aggregated into the classroom or the school level or grade. So other people see these. Administrators and others. They get a sense of collectively how did people do. And if you had a case where let's say a third of the students were disengaged during a test you would find that out by rolling that up and getting that information into a report.

So we are planning to change our reports at the classroom and school level to reflect that type of information. That's a very good question and we are working on that. We are just now introducing the individual score report, student profile work. But we are going to include it in the other reports as well.

>> DERRICK VARGASON: Great. Well, thanks, everyone, for staying with us an staying through the Q&A and all of your great questions. Again, this presentation will be posted on-demand through Ed Week.org for your review and sharing with your colleagues. On behalf of NWEA, myself and Steve, thank you for joining us today and have a great rest of your day.
(The meeting concluded at 2:00 o'clock p.m. CST.)
(CART provider signing off.)

***

This text is being provided in a rough draft format. Communication Access Realtime Translation (CART) is provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings.

***