Assessments Friday Focus 2026-2027 2026-2027 Letter Grades, Appeals, and Accountability



Smith, Sean

Thanks again to everyone who's joining us today. My name is Sean Smith. I'm the Chief Accountability Officer at the Arizona Department of Education. What I've got for you today are just a few updates on indicator scoring. So, that's the latest changes that we're making to the A-F accountability system as required in the statute. This is going to take place for the 26-27 school year. So, we just released letter grades for the 25-26 school year. Those look similar to that would have in the past and these changes will be for planning purposes going forward, our hope is always to give you as much lead time and heads up as we can as you're planning for this, recognizing that this is taking place as the school year is happening. So, I'll go ahead and start sharing my screen here.

So, up front, just a real quick timeline of what we're expecting. Next month, we've been working with our IT team to roll out historical what-if data. Initially, that'll be for 24-25 fiscal years. And then for the latest 2026 data, we will process all that data that's recently been worked on and get that in there as well. But what we're looking to do is push that into ADEConnect and show everybody for their schools, you know, what would it what it would have looked like if in 2024 and 2025 we've done indicator scoring, right? So, give everybody kind of a feel for it be able to plan communications and understand a little bit of how it works.

January, February of 2027, we plan on bringing new business rules to the Board of Education for their approval. That will, again, have all the calculations, the formulas, and the specifics on how everything works. But our, like I said, our hope is that the most useful portion will be as we publish that what-if data as early as we can.

Then June-July 2027, that should be our typical timeline for release of the static file and the preliminary data and calculations. And Then November 1, 2027 is that required deadline for publishing the finalized letter grades.

And if you have questions as we go along, please feel free to drop those into the

chat. Candis will be monitoring those for us as we go. And Candis, if you're seeing anything that looks urgent or needs to be addressed immediately, just let me know.

To talk about and understand the change of from the current system to the indicator scoring system, I put these slides together to just kind of give an idea. The current system is what's called in literature a compensatory system, right? So the idea is that you have an indicator or a set of indicators that balance out other indicators, right? So, it's going to, for instance, you may have really low readings on one indicator and that's balanced out with really high readings on another indicator. Currently in our current system, that is, you know, we see typically low ratings in proficiency, and those are balanced out by most schools getting most of the points in the other indicators. So, most people are probably familiar with that, and that makes sense in a compensatory model where we're just giving a summative grade but it starts to break down when all the indicators are standing by themselves and they're each individually accessible rather than just combining to form the whole, right? So again, if they're not all working together like they are currently in the compensatory system and they break apart, then this compensatory idea that one balances out the other doesn't quite work. So, the balance of that has to shift to a conjunctive model. And this is where everything is designed to work together. Each of the indicators stands alone, but instead of to get the highest grade, you have some elements balancing out others and can cancel out some signals. The highest grades and the positive signals come from essentially all the indicators pointing in the same direction. It's doing good across multiple indicators.

This is just a data visualization that shows that dynamic. On the left-hand side here, you see the balance of points in the current system. So, you can kind of see at the top, proficiency, about 61% of K-8 schools currently get an F in proficiency. They get 50% or fewer of the points. And that's balanced out by growth, EL and acceleration readiness, where schools are getting between 45 and 61% of schools are getting A's. So, the A's balance out the F's in proficiency.

And I think you can kind of see why this would be a problematic approach to indicator scoring. We just said everybody's an F in proficiency, but everybody's an A in everything else. It creates a lot of problems for any schools that aren't in these buckets to explain and to understand and to be able to react to. And it also, it misses

what's really happening. It doesn't include the context of any of these indicators, right? The idea currently is just that you add these points up and then you add these on top and you add these on top and it all comes out together in the end. But like I said, from an information perspective, that results in sometimes you have some indicators that kind of cancel out and send a signal opposite to what the other indicator is saying.

The problem with that is the research is really clear that the best, most effective accountability systems function as information systems, right? They provide information back to school leaders as far as what needs attention, what's working well. So, these on the right-hand side, so we see concordance, concordance plus half, and concordance plus 12. I'll go into what that means exactly. But these were the three options that we presented to the board. So, look at each of these as a column. And the board approved the concordance approach. And when I say concordance, what we're meaning by that is a high degree of similarity to the outcomes of the current system, right? So, if we're seeing a certain percentage of A's, B's, C's, D's, and F's, the board has directed us to come up with an indicator, a set of indicator scoring cuts that would produce similar outcomes to what we see currently, right? So again, this shows over here that the meaning of some of these is going to change. But the idea of concordance is that there's a high degree of similarity between the outcomes of the new system and the old. In other words, there's going to be, we didn't want to shift to something that was, you know, all proficiency where 61% of schools were getting F's. We needed it to map more in line with the communication that people are used to seeing. And that's important because one, these letter grades get used all over the place by all kinds of people outside of education, within education communities. We didn't want the communication to shift so radically that they became completely useless and that there was absolutely no point of comparison between the two.

So what you can see here is that in contrast to the way things currently work. These are not designed to compensate for each other. Rather, each indicator functions within its own context so that we see a range of outcomes that's similarly balanced. The schools who are doing the best in any of the indicators are going to receive A's, B's, C's, D's with roughly a similar level of rigor, right? Here you can tell it's much harder in the current system to receive an A in proficiency than it is to

receive an A in acceleration readiness. The biggest change that's going to happen is that now that will be similar levels of rigor across each of the indicators. It's going to be just about as hard to earn an A in growth as it is in proficiency, as it is in acceleration readiness, as it is in English learners. And forgive me, I'll show these graphs for high schools and alternative schools as well. The trend is the pattern is going to remain the same. I'm just kind of talking to K-8 because that's the majority of schools.

So, again, big change being that there's going to be a similar level of rigor and balance across each of the indicators, and that each of the goals is going to be at a threshold that responds to the context of that indicator. So, we don't have an indicator where it says everybody's an F in this one, everybody's an A in that one. And then just to reiterate, the board adopted this approach, which would have a high degree of concordance and consistency with the outcomes that we currently see.

So, here's what that looks like with high schools. Again, you can see over here. These indicators, growth, English learners, graduation rate, and college career readiness, balancing out where we're seeing 55% of schools earning an F in proficiency. So, all these ones have a high percentage of A's, this one has a high percentage of F's. Again, in any of these models that we proposed, a function on that conjunctive approach that has a similar degree of balance between all the methods.

Alternative schools. Same idea, 100% of schools earning an F, balanced out by some of these other indicators like college career readiness where 82% of schools are in A.

Things that we have considered and designed as we're making changes to the system as well to the underlying indicators. So that idea of compensatory conjunctive, that's kind of an overall system dynamic approach. When we looked at each of the indicators, what we, one of the things we wanted to measure was the association with non-school factors, right? So, a school's location was rural or urban, size, and poverty. So, we know that these are factors that are not necessarily within a school's control. And so, we didn't want to just pick up, we know that with the current proficiency measure, there's a very high association with some of these non-school factors. And that's true for most of the indicators as well. They have a lot of many of them have a high degree of association with non-school factors.

Acceleration readiness, for example, had a high degree of association with, it was strongly associated with the size of a school and where the school is located, whether it was rural or whether it was in a more populous county.

What we would want to see is that along this line, and we did this for each indicator, but you can measure the association using the statistical test. It kind of shows you if everything's kind of along that line, that means the odds of any school, or excuse me, the odds, for example, of a rural school are the same as the odds of any other school getting an A in that particular indicator. So, we looked at this for each of the indicators. And what I can tell you is that it's not perfect. There are, there is still going to be an association with non-school factors, in particular measures like proficiency. But we've seen drastic reductions as we've designed new approaches to it and how we set those cut scores that greatly reduce the association with non-school factors. So, one of the things that you're going to feel noticeably, almost as kind of biofeedback, is that the new system should respond in a greater degree to the actions that you're taking as school leaders, as educators, and it's going to feel less like just destiny based on where a school is located or how big a school is.

The other factor that we wanted to look at, and as we were setting cuts and proposed those three different approaches to the State Board of Education, we wanted to make sure that it was associated as well with a change in student outcomes, right? So, improvement in student outcomes. We didn't want to just create a math problem that produced a bunch of numbers. We wanted a sanity check to make sure that each of the indicators was corresponding to improvements in student performance. We looked at this for each of the indicators, and I don't want to misrepresent it and say, like, hey, we solved it. It's all perfect. Obviously, there's no perfect accountability system. But what I can say is we've had a high degree of success in making these indicators both more tied to student achievement, as well as less tied to non-school factors. So again, the big through line there is that as you're approaching things in the new system, a lot of the concepts are the same, but you should notice a real difference in the results responding to the actions that you're taking rather than factors that are beyond your control or that seem to go up and down. And I mean, a good example of this one was the growth indicator. A lot of you probably experienced this where you'd see that depending on the kids that came in in a given year, even if they were growing for those kids, the point dynamics could

result in, you could have really high growth for those kids and get fewer points depending on where they started off. You know, the current dynamics, have the way they award additional points for students who are minimally proficient in prior years, if you made big gains and then didn't have as many minimally proficient students, that could end up hurting you, right? So those weights really came into effect again. So, the student achievement may not have been closely tied to how the points were earned. We've tightened that up again while still maintaining or reducing that association with non-school factors.

This is a document that is still in progress, but it goes over the changes that we've made to the, you know, what we call the small M methodology, type of the big M methodology, how the system works overall, changing from compensatory to conjunctive. These are the small M methodologies. All right, how are we going to actually measure each of the things. What you'll see if you reference this document is that conceptually, a lot of it is really similar. It's not going to be a complete overhaul that you'll have to study for weeks and weeks to try and make sense of. A lot of this is really going to feel familiar and again, we've tried to tailor it to student outcomes and school efforts. So ideally, the things that you would already be doing and the things that you naturally inclined to as educators, we're trying to tune it so that it picks up on those things that you do for student success. So, our hope is that
This creates a reduced burden to where you're having to do a wholesale curriculum or structural change to fit this system. We're trying to fit the system to good practice and make it more useful. So, a good example here, the proficiency indicator currently, that operates on a scale of 0 to 1.3, and then we multiply that times the number of points that you get. It's capped. There's a bunch of things going on. If you get minimally proficient, it's zero points, partially proficient, it's.6, 1.0 for proficient, and 1.3 for highly proficient. It makes it a little bit hard to understand and convert that in your head as far as, you know, if you're scoring a point 8 on that, you'd get 24 points in K-8 or high school. And so, what does that mean? So, we've tried to change it, or excuse me, we've changed this under the hood. The goal, the intent of it is to make it better communication. Where same idea, we're looking at kids across the range of performance who get points for kids no matter where they are and how they're moving. But just on the scale of minimally proficient one, partially proficient 2, proficient 3, highly proficient 4, just so it maps to what you're already getting back and familiar with from assessment. And that way, you know, if you're scoring a 3.3

under the hood there, that makes a little more sense in terms of what does that mean in terms of average performance. And almost everybody's getting proficient and moving towards highly proficient if you had a 3.3. If you had a 1.3, you would know the vast majority of your students are close to minimally proficient.

Again, that document is accessible, still a work in progress, but it'll outline what's staying the same conceptually, what differences we anticipate. We'll have full entire details in the business rules, but our hope is that this gives you a quick guide as far as what's different, anything that you would need to plan and adjust accordingly.

Bonus points, if you've been to the several of the last state board meetings, you may already know this. The bonus points are being removed from the 26-27 letter grades. This would have been, if we remove them from the 25-26 letter grades, this would have made a radical sea change difference. It would have seen a lot of disruption. The good news is that although we did not anticipate bonus points being removed, we had already anticipated as we've been designing indicator score over the last couple of years, that bonus points had some problems with them to begin with. They were a little unpredictable as far as what schools we're going to get them and what schools we're going to get them year to year. There were problems dealing with sometimes a school would be eligible for some points and not others. And they were there were a lot of different points. You know, there was nine points on the table for high school, 6 for K-8. And then again, just the range of letter grades, nine points in high school means something very different than six points in K-8. So, there was a lot of inconsistencies. There was a lot of tie to, again, some of the non-school factors, in particular school size, that made it difficult for schools to always rely on these or predict where they were coming from. So, we've kind of seen those problems.
And we'd already done some work to address some of those issues in the design, so that the positive thing is the removal of those bonus points is much less, it's going to have a much less significant impact on 26-27 letter grades in an indicator scoring paradigm than it would have if, if we had simply done that in 2025-2026.

But a lot of the impetus for removing them was from the Auditor General report. There was concern from the Auditor General, from the legislature that it inflated grades, and that Arizona was the only state doing bonus points. Again, we'd known already that there were some issues with bonus points that they caused some

problems already. We kind of designed around that, and it's going to mitigate some of the impacts that you feel from removal of the bonus points. But we also want to let you know that that is there. That's something to consider.

Again, you can kind of see here, this is across all models, high school and K-8. If we'd taken the 2025 results and compared them to the 2026 results, oh, sorry. I want to skip this slide. I'm forgetting what I'm even talking about here. Oh, yeah. Sorry. Here's what would have kind of been the results. Without bonus points, we would have seen this sharp increase in B's, large decrease in A's. What you can see is that with the concordance levels that the board adopted, there is still a decrease in A's, but it's mitigated fairly significantly. Same with the increase that you see in C's. This is just what would have happened if we'd applied them to 25-26 letter grades. This is what 25-26 letter grades, or excuse me, 2025 letter grades would have looked like no bonus points. This is what they would have looked like if we were using the indicator scoring approach that the board adopted.

And I think we're about to wrap up, but I want to be upfront about benefits and challenges that we anticipate. There's a lot of good that we hope to see from the indicator scoring approach. This was a change to the law in 2018 that was introduced by a school superintendent. The intent of it being that there was a desire to make the grades clearer and be able to highlight where schools are doing well and not just lump everything together into one overall grade that may not represent all the things that our schools are doing well. And the areas that That they still want to work on, so we've tried to design with that intent in mind. Again, the idea is to show accomplishments and nuance that there's a lot of things schools are working on, even, you know, even of all the things that are missed in an accountability system. There's still a lot going on, condensing it all into one single grade can miss some of what's happening there. So, it highlights this. On the challenge side, that does bring additional scrutiny. I'm not going to read through this entire slide, but these top two I think are worth looking at. The other side of the coin is this should be much clearer communication to your boards, to your communities, instead of trying to explain why, you know, 43 out of 50 points in growth is maybe a problem area or why, you know, 13 points out of 30 is not as terrible as it sounds in proficiency. This is going to come back as an A, B, C, D, F within the context of each of those. It's going to be a little bit easier to have a conversation starting on common ground rather than trying

rework the business rules and talk about how the points are earned. You can just kind of say, you know, here's where we're doing well, here's areas we still want to work on. The challenge of that is that because it's potentially easier communication, that's going to bring a lot of people into the conversation that maybe haven't been able to access it before, right?

So, it will require some additional support as far as communicating, helping people understand what they're seeing, telling the stories that they need to hear, and helping them make sense of what's now a little bit more accessible. So, again, don't want to go through all of that in detail.

But I will open the time up for questions.



Middlebrook, Candis
Thank you, Sean. We do not currently have any questions in the chat, but everyone, if you have questions, please put them in and we will get to them.



Smith, Sean
I'll try and leave good wait time like a like a good teacher, but then if you guys are all clear, I think, ready to go. I'll hang out for a few minutes in case anything comes in. But thanks again for joining us. Please reach out to either of our teams if you have questions. We want to be here to support you as we're making this change. Thanks, everybody.

Okay, hi, Tara. With the indicator scoring, we won't have cut scores until after the data is all compiled, and then you will use the concordance percentages to determine, correct? Is that correct?

Here's what we know so far. We'll still probably need, I don't think they'll be official until the board votes on them after data is already out and everything. But the plan is now with what the boards voted on; we're going to set those initial goals that we understand to be the cut scores as they are.

But I'll use English learners as an example. You know, we saw this big increase in reclassification this year for English learners. And I taught English learners for a long

time. I know sometimes I have a big year one year, and then I have a lot of kids who are still working on it the next, and so I'd see some declines. That's the kind of thing where we may need to see, just like we do currently, we bring to the board, say, here's last year's cuts. That was kind of our framework and our starting point. Did an input change that requires readjustment? And so what we'd say, what we understand now is the board has voted on concordance. So, we kind of bring back and say, all right, here's, you know, if the scores went way down, here's then how we'd have to kind of readjust that to achieve that concordance level. So yeah, we'd anticipate that the process still looks similar to that where we get final cut scores officially approved in October, but that we have some initial cut scores that barring any change, initial goals that let people plan accordingly and say, all right, here's where I need to be focused, recognizing that that could, you know, be adjusted downward, possibly, if if the inputs change radically.

Oh, great question, Catherine. Is the concordance data based on one year or multiple years? So, we've done primarily as we looked at it in the analysis, 2024 and 2025, and we saw real pretty consistent results in those two years. On a few of the indicators are already pulled in 2026 data. Those final goals will look at, you know, what does concordance look like in that lens of 2025-2026 for the latest data. But yeah, everything has stayed pretty consistent even across years. We've seen just enough scale that the ranges stay very similar. Thanks, Catherine. Nice to virtually see you. Love back at you, Brian.
Okay. Candis, do you want to call it good?



Middlebrook, Candis
Yes, I think so. If you have additional questions, like Sean said, you can reach out to either of our teams and we'll be happy to assist you. Hope you all have a great weekend.



Smith, Sean
Thanks, everybody.