IES Learning Acceleration Challenges: RCT Webinar Transcript

Eden Baker (00:00):

Welcome everyone. We are recording today’s webinar and we will share a recording, as well as a copy of the slides, following today’s session. We will also share a copy of written FAQs. So if we don't get to your question today, we will hopefully publish that on Challenge.gov, following the webinar. So my name is Eden Baker and I use she/her pronouns. I am a white woman with shoulder length, blonde hair and I am sitting in front of a Zoom background that has the logo for the Learning Acceleration Challenges. For anyone who was not at our first webinar last week, I'm an engagement manager at Luminary Labs, and we are an innovation and consultancy based in New York. We've been engaged by the Institute of Education Sciences to run the Learning Acceleration Challenges.

Eden Baker (00:44):

So we are really excited to have everybody here today. The purpose of today's session is to provide an overview of the randomized control trial efficacy study, that is part of these challenges. This includes explaining what entrants need to do and consider as part of their Phase 1 submission, as well as what finalists will be required to do to support this study in Phase 2. As I mentioned earlier, we are going to have a Q and A at the end of the session, and we'll be using the Q and A function in Zoom. So if we do have questions, as we're talking, feel free to share them in the Q and A box down the bottom of your Zoom window, and we'll come back to these during that part of the session.

Eden Baker (01:21):

If you missed me saying, this is at the beginning, we are recording today’s webinar and we'll be sharing a recording following today's session as well. We're really pleased to be joined by two excellent subject matter experts today. So Andrew and Sivan, I'll pass it over to you to introduce yourselves and Andrew, we'll start with you.

Andrew McEachin (01:41):

Yeah. Thanks Eden. So I'm Andrew McEachin. I use he/his pronouns. I'm a white man with short, brown and graying hair. I blame the pandemic. I ,too, am sitting in front of a similar background that has the Learning Sciences logo behind me, and I'm the director of the Collaborative for Student Growth at NWEA which is a nonprofit testing and research organization.

Sivan Tuchman (02:05):

Hi, my name is Sivan Tuchman, and I work at Datability Education Consulting and I am white woman with dark, brown hair and glasses. I am in front of a light gray background with the project logo behind me and we'll pass it back to Eden or Andrew to get started.
I'll jump in just quickly. So before I hand it over to Andrew, to jump into details of the RCT. I wanted to just quickly provide a bit of a recap of the timeline for Phase 2. So if you are selected to progress to Phase 2, which is the implementation phase, you and your partner schools will participate in an efficacy study using a randomized control trial or RCT design. In Phase 2, finalists will implement their interventions at partner schools in both prizes, so in math or science or out of school time programs are for the Science Prize only, under routine conditions. Participating students will take the NWEA Map Growth assessment before the implementation period begins and really importantly, this has to be done by November 1st, 2022.

Eden Baker (03:18):

So this does mean that your partner schools will need to have completed this assessment before you know whether your intervention has been selected to progress to Phase 2. Finalists will be notified in early November and then, will be able to start implementing their interventions. The duration of the interventions may vary, but they must be implemented between November, 2022 and April, 2023. Then, participating students will take the NWEA Map Growth assessment again at the end of the implementation period. This needs to be done by May 1st, 2023. Schools will then be asked to provide student and school level data to NWEA by May 8th, 2023. And Andrew will be showing a little bit more about what this data involves later in today's session.

Eden Baker (04:03):

NWEA will then use this data and scores from the two assessments to prepare evaluation reports. Finalists will also be asked to submit additional data as part of their Phase 2 submissions. So this will include things like information about implementation, as well as the scalability of the intervention. Judges will then review these Phase 2 submissions, as well as the NWEA evaluation reports against the Phase 2 selection criteria. These criteria are available on Challenge.gov, and one of my colleagues can also pop a link in the chat so that you can take a look at these in your own time. So that's just some extra context for today's session. Andrew, I'll pass it back over to you to get into the details of the RCT.

Andrew McEachin (04:47):

Thank you, Eden. So for today, we're going to start with some high level overview of kind of what the RCT or randomized controlled trial looks like as part of the efficacy trial for the both challenges, and we're going to split it between the Phase 1 and Phase 2 components. So during Phase 1, you'll have an iterative process where you both need to kind of figure out how the intervention is going to be implemented for eligible students. Meaning, are you going to randomly assign at the classroom or school or other various levels? And we'll get into those details in a couple slides. Along with that, figuring out how many students, classrooms and/or schools are you going to need to have a randomized control trial of a size that will make it so you can be reasonably competitive in the challenge.

Andrew McEachin (05:38):
So Phase 1 is really figuring out how you're going to implement it, both in style and in sample size. Then, Phase 2 really is doing the actual intervention and randomized control trials. So the very first part of Phase 2 is going to be to randomize students and/or teachers, classroom, schools, whatever the randomization unit is, and we'll talk about that again, in a couple slides, to the treatment and control condition. We at NWEA will help support you with that. Once you have the randomized control or randomized conditions in hand. So you have kids assigned to either treatment or control conditions, you'll need to maintain and track the assignment throughout the school year.

Andrew McEachin (06:23):

Then, also track how kids are participating in the intervention. So kind of think about it as a dosage, how they're engaging with the intervention itself. This is all done in Phase 2 and feeds into the Phase 2 judging criteria. Throughout Phase 2, Abt Associates will be providing technical assistance to kind of help with kind of conducting and implementing and maintaining a randomized controlled trial throughout the school year. So right now, we're just going to talk at a high level of a Phase 1 requirements. Then, after we do this, we'll talk about the Phase 2 requirements. So as we talk about the Phase 1 requirements, there's two kind of overarching things that kind of should be in the background as decisions are being made.

Andrew McEachin (07:15):

The first is what we kind of call sample size. Do you have the right number of students, classes, school, whatever sample is being randomized, whatever entity is being randomized, do you have enough of them for us to be reasonably confident that we can find a benefit for students? If you don't have enough students or don't have enough schools, if your sample size is not large enough, it's difficult to find a statistical difference between the two groups and that then would hurt the Phase 2 judging. So the first part is making sure we have a large enough sample.

Andrew McEachin (07:50):

The second and related to that is this idea of spillover that might not be as straightforward, but this generally happens when you implement an intervention in a way that also benefits the control condition. So an example of this could be, you take a given classroom, call it Classroom A, half of the students get access to the intervention, half of the students do not, but there's something about the intervention within a classroom that not only benefits the kids who get exposed, but it spills over to the control kids. Maybe classroom management increases. Maybe the teacher can go at the math instruction at a faster pace, because not only are the treated kid is asking fewer questions, but because of their classroom management, the overall experience for kids are better.

Andrew McEachin (08:37):
What's going to happen is you can see that not only is the intervention working for treated kids, but it's working for the control kids. Then, when you go to look at the difference between these two groups, at the end of the year, you might not see that large of an effect because everybody benefited, and if the goal of this exercise is to figure out how the intervention benefits students who otherwise, wouldn't have been exposed to it, we want to make sure that this chance of contamination doesn't happen. So as you think about implementation and randomized ... the design of the RCT, you need to think about spillover. And at a high level, the best possible way to minimize this is to implement the intervention in a way where the benefit of the program does not kind of bleed into or spill over into the control condition.

Andrew McEachin (09:27):

Yeah, we'll go to the next slide. So conceptually, what could this look like? There are a number of ways in which you can think about randomly assigning students to either a treatment and control condition. So the first is really to think about what these two conditions are. So the treatment condition are the group of kids who are eligible in this instance for the challenge, who are going to be receiving the intervention. Then, the control group are going to be the kids who continue to receive the business as usual instruction. This is important for a number of reasons. One, the main way in which we're going to judge whether or not this intervention is working and/or impacting student achievement is essentially comparing the outcomes at the end of the school year, between these two groups, between the treatment and control condition.

Andrew McEachin (10:19):

So we want the control condition to really represent what would otherwise happen for students during the school day in absence of the intervention. So it's important to think about what that looks like for your school site. If the control condition is a really effective intervention, then you're not comparing an intervention versus the absence of one, you're comparing interventions. So it's important to kind of think about what the control condition is and to document it. There's also a variety of ways in which we can think about randomly assigning, and we'll talk about the trade-offs for this in a second. At the highest level, commonly is to think about randomly assigning entire schools to an intervention.

Andrew McEachin (11:01):

So school A gets it, school B does not. School C gets it, school D does not, and every kid in one of these two schools is either in the treatment or the control condition. In this instance, spillover is quite tricky because it would require the behaviors that are happening in one school to spillover into the behaviors that are happening into another school. On the other hand, as we'll talk about in a few slides, if you randomly assign whole schools, you're going to need a large sample size, in this instance, in terms of a number of schools, because really the randomization unit in this instance, it would be schools, dictates your sample size.

Andrew McEachin (11:38):
So it’s not really the number of kids that matter, it's the number of schools that you have, and that would require you to have tens of schools, even if that meant thousands of kids, it's really the number of schools that matter here. A second way to think about this is to randomly assign whole classes. So you could take third grade math in a given elementary school, and some of those classes get the intervention, and some of them do not. You’re less likely to have spillover in that instance because what's happening in class A, minute by minute is less likely to affect class B. The chances of spillover are potentially higher than randomly assigning at the school level. One benefit is if the random assignment unit dictates your sample size, you’re going to be able to get to a larger number of classes quicker than you are a larger number of schools.

Andrew McEachin (12:38):

This is more relevant potentially to the science challenge, but you could also think about a randomization design that pulls kids out of their classrooms into small group settings, for science that might look like an out of school context, for either of the challenge that might look like a kind of study hall or something like that, a setting where they're pulled out of their classrooms into smaller groups. Then, for science in particular, you could also think about a similar pullout structure, but it happens out of the school day. So it can be potentially off the campus after school hours, but it's something that happens in out of school programs.

Andrew McEachin (13:24):

Cool, so in guiding our conversation both today and some of the documentation that we’ve provided, that we’ll have links to, on Challenge.gov, we provide some overview for sample size considerations for a few common research designs, and in guiding that, we’ve anchored our calculations to an effect size of roughly 0.3 standard deviations. Conceptually, you can think about that as taking a student who’s scoring at the 25th percentile on a given test and moving them to the 35th percentile. So the effect size of 0.3 is fairly large and fairly meaningful, and that effect size guides our sample size consideration. So if you see the number of students or schools or teachers in these documents, it’s based on a design that would adequately find an effect size of 0.3 in an RCT.

Andrew McEachin (14:25):

Again, in the following slides, we'll kind of flush this out in a little bit more detail. It's really the number of students and/or classes, teachers or schools that you need for a given RCT is really dictated both by this number above 0.3 standard deviations, but also how your intervention itself is being implemented. So if it's best implemented uniquely across classrooms, then the power calculations and the research design, et cetera, would need to randomly assign across classes. If there's a really strong threat of spillover, then maybe it needs to be done across schools or if there's something about the partnership between the intervention provider and the schools that really suggests it needs to be done in a certain way, maybe that's what guides it.

Andrew McEachin (15:13):
This is also going to dictate then, based on how you're going to implement it, whether you're doing it at the school level, whether you're doing it with classes, individual students, whether it's a pullout, that's going to dictate the sample size that you need, and we provide for the math challenge and the science challenge kind of three high level ways in which it might be common to implement it, but it's definitely not going to be all encompassing. So, it's important to keep that in mind. For those that have kind of the background or experience doing power calculations, kind of the numbers that we've used are using the power up tool or come from the power up tool, and we provide the link here in the slides and we can provide the link both in the chat and these slides will be available.

Andrew McEachin (16:01):

There's a variety of ways in which you can use the tools on that website. There's an R Shiny tool, there's an R program, there's a web-based platform, there's formulas available, et cetera. So all of the numbers and stuff that we provide are generated using this tool. I should say the last thing I'll note too, with this is that for the help with the matching process, there is an email, it's challenges@nwea.org, that I think we can drop in the chat as well, and you want to email that email to get help with the matching process and in doing so, not only do you want to use that email, but it'd be best to provide up to 10 potential districts in which you think you'd want to match with for this challenge.

Andrew McEachin (16:53):

We can try to work with you to see if those would be good fits. So the first example that we're going to talk about today, and that you would also find in the accompanying PDF for the challenge is a design, where we're randomly assigning full classes within a school, and what this table does is it gives you three different sample or account, three different examples for eligible students per class for the Math Prize. So you could think about settings where you might have three, you might have six, or you might have nine eligible students in a given class. We realized that in any given class, these numbers are going to kind of shift, so it's best to think about these as averages.

Andrew McEachin (17:42):

So on average, you would have three eligible students per class. We also think about settings where you would have three, five or eight schools that would be willing to implement the intervention in this RCT challenge. Then, given those two parameters, you see how many classes you would need across the third, fourth, and fifth grade. So, as an example, if we see kind of the middle row, if we have six eligible students per class, and you're in a setting where you have five schools, you would need each of those schools on average to have five classrooms across grades three, four and five, it could be two and third grade, two and fourth, one and fifth.
If you have more classes than that, that's great, you just want to try and avoid a setting where you have fewer than that, and we provide as an example, the total number of students that you would need, but really the most important thing in all of these tables is not the total number of students, but the number of classes and schools, because we'll see soon that for some settings, you'll end up with a lot of students, but it's because you need a lot of either classes or schools or teachers to have an adequately powered RCT. So really the focus is on the schools and classes or teachers and some of the other slides, not so much the total number of students.

Andrew McEachin (19:15):

Yeah, so a second way that we might think about doing this for the Math Prize is something called an AB design and one of the benefits for this is it allows a little bit more flexibility in the day to day practices within schools. As an example, if we're randomly assigning individual classes, as we did in the previous slide, that design makes it so teachers can't co-teach or directly collaborate because if one teacher is assigned to the treatment condition and one is assigned to the control condition and the treatment itself is changing the teacher's behaviors, her behaviors might now spill over into the control condition and it might muddy the treatment effect or it might ... it'll water down our ability to find a statistical difference between the treatment and control, which will ultimately hurt Phase 2 judging.

Andrew McEachin (20:11):

So one way to get around that is to think about an entire grade level, being assigned to treatment or control. That way, if the teachers are working together, it is not creating any sort of spillover. So in an AB design, what that would look like is that you would create a group called the A schools potentially, and you would assign two of the grade levels to treatment. So, in this example, we have grades three and grades four, are assigned to treatment and grade five is assigned to control. Then, the B schools are the opposite of that, so in grades three and four, would be assigned to control and then, grade five is assigned to treatment. The other thing that's nice about this is that you could work with your schools to figure out what grades would be best to be for A, and what this pattern could look like.

Andrew McEachin (21:00):

You could also do grade three is treatment, grade four is control, grade five is treatment. The main thing to think about in this design is whatever you do for A designation, you do something that's the opposite for the B designation. Again, why you might think about this design relative to just randomly assigning classes is if you think that there's something about how teachers could elaborate across classes within a grade, that might create a spillover issue. This AB design can help get around that. So in this table we provide A, again, number of kids. Here, when we say grades, it just means the number of grade levels. So this isn't referencing third grade. It means like you would have grades three, four and five.

Andrew McEachin (21:52):
So these numbers might look different if you're working with a school that ends in fourth grade and you only have two eligible grades, but assuming you're in a system where schools have grades three, four, and five, you can see how, if you go from 10 to 30 eligible students, how many schools you would need and roughly in this design, you're going to need nine or 10 schools to do a well-powered AV design, and you can also see that you could increase the number of students per grade by an order of three, and it really has a marginal effect on the number of schools. So increasing the number of students by a factor of three only decreases the number of schools that you would need by one.

Andrew McEachin (22:33):

Again, it's another example of when you're thinking about sample size, once you go beyond randomly assigning individual kids to some higher level grouping, whether it's classes, teachers or schools, it's really that higher level grouping or what sometimes referred to as clustering or nesting that dictates the number of groups that you would need. It's not necessarily the number of individual kids. As an example here at the bottom, if you have 30 eligible students per grade, you have three grade levels per school and you have nine schools. You're going to have 810 total students, which is a lot more students than the previous design, and it's driven by the nuances of how the grades and schools are assigned in this setting.

Andrew McEachin (23:22):

And then the final one, which might be the easiest to implement, but has a trade-off of needing quite a large sample is you could think about just randomly assigning whole schools to either treatment or control, and if you do this, one benefit, as we mentioned is that the chance of spillover is quite small because the practices that are happening in school A are less likely to contaminate what's happening in school B. The trade off is in this setting, if we think about having anywhere from 15 to 50 eligible students per school, you're going to need somewhere between 15 and 22 schools to be able to randomize into treatment and control conditions, to find an effect size of 0.3.

Andrew McEachin (24:04):

Also, for this slide, we show that for the Math Prize, if you're worried about the grand prize, which for math is a threshold of 0.77 standard deviations, you would need at least nine schools to have an adequately powered RCT for this design. So again, for this design of randomly assigning schools. There is far less of a potential of spillover and it's also potentially maybe easier to implement because an entire school is doing one practice or the other. The trade-off is though, you're going to need far more schools and then, also far more students than you would for the first two designs that we showed. So for the next three, we're going to walk through something similar for the Science Prize.

Andrew McEachin (24:52):
The Science Prize has kind of two differences that well, we'll talk about between the math challenge and the science challenge. The first one is the science challenge is focused on what are traditionally middle school grades. And often in middle school, in the modal middle school, teachers have many classrooms that they are responsible for, anywhere from like three to six periods during the day. Whereas in the modal elementary school, a teacher is responsible for a single class. So if we think about the issue of spillover, we can't really randomly assign individual classrooms or class periods to treatment or control because if a teacher has half of her periods exposed to the treatment and the other half of her period is exposed to the control condition, then if something about the treatment is improving her practice that might also then spill over to the control condition itself.

Andrew McEachin (25:47):

So you need to elevate the unit of randomization from the individual classrooms up to the teacher level in middle school. I'm thinking about we're going to randomly assign a teacher and all of her responsibilities to either treatment or control, and in that instance, you're less likely to have within teachers spillover, because all of her practices are under the same treatment condition, either treatment or control. So again, in this table, we provide the number of eligible students per school, as an example, the number of schools and then, how many teachers you would need in a given school for that. So if you have on average 30 eligible students per teacher, and you have five schools, you would need four teachers in that school setting for about 600 students.

Andrew McEachin (26:44):

Related to the Math Prize, you could also think about, in the science setting again, randomly assigning whole schools. So the main numbers don't change for the science setting, if you're randomly assigning an entire school. Again, it doesn't really matter, even though middle schools are on average, bigger than elementary schools, when you're randomly assigning schools, that's your sample size. So it's 15, 20, 25, 30 schools. It's not the thousands of kids that you might have. So the number of schools that you need for the Science and Math Prize are going to be roughly the same, if you're thinking about randomly assigning the whole school. One difference is if you're worried about the grand prize for the science challenge, the grand prize for the science challenge is set of a threshold of 0.4 standard deviations.

Andrew McEachin (27:31):

For that, you would need 11 schools. So again, the whole school design, the benefit is, it really reduces the chance of spillover, but at the trade-off of needing a potentially larger sample size, in terms of the number of schools and then, also the number of students that would come with them. The final design that's unique to the science challenge is thinking through an out of school time program. So in this figure here, we have triangles and circles and each of those represents individual kids and then, the shape is the school that they're in. So school one here is represented by the circles and the blue triangles are represented for school two. So you can see you have an eligible pool of sixth graders in this simple example.
There, you're coming from two different schools and they're going to be randomly assigned to one of two conditions in this out of school program. We'll start with the control students. You'll see that the control students are kind of just randomly assigned into their regular afternoon. So whatever these kids would've done, if they don't go to this program, that's what they're going to do. So then it might be going home, soccer practice, whatever it is, you're not providing anything to the control students. They're just doing whatever their normal afternoon or out of school experiences are. For the treated students, they're going to be randomly assigned ... sorry, for the students who are randomly assigned into the treatment group, they're going to be exposed into the intervention.

Andrew McEachin (29:06):

Not always, but often, and out of school programs, there might be further grouping. It could be small group tutoring. It could be because of the physical capacity of school rooms. So, kids have to go into room one, two, three or four. Whatever it is, you might have to further randomize and group students into treated subgroups, and this adds complication to this design because the experiences that the treatment kids get are shared within this different treatment design. So kids who are assigned to tutor one, share similar experiences. Kids who are assigned to tutor two, receive similar experiences, but none of those experiences spillover are experienced by the control students.

Andrew McEachin (29:57):

So we have to do kind of a special calculation to figure out the right sample size in this kind of a setting and with a big caveat being there's a lot of ways in which you could implement this. You could have no groups. You could have really small groups, really large groups and all of that will influence the number of ultimate students in schools that you would potentially need. So we provide one example of what this might look like, but it's not the only example. And if your intervention is implemented differently, you all have to think about the sample size slightly differently than what we've provided on the next slide. So in this simple example, we think about ... we're providing an example where students are grouped.

Andrew McEachin (30:45):

The students in the treatment group are grouped in two smaller groups. So in this instance, if we look at the first row, we're going to have two groups of students for the intervention. So it'd be like tutor A and tutor B, and you're going to have 10 students per group. In that sort of a structure, you would need 11 schools for a total of 220 students, and you'll see that as the students per group increases, the total number of students increases rather quickly, but the schools needed does not change very much, again because when we think about sample size for these research ... randomized controlled trial designs, it's really the number of groupings above the students that matter more than the actual number of students.

Andrew McEachin (31:41):
So as an example, if you have 25 students per group and you have four groups per school. You’re going to need eight schools in that setting, to have a well-powered design, and you'll have 800 students, which is quite a few more than the first row. Then again, it's driven by this kind of messiness of grouping students into various smaller groups. We'll show at the end of this slide, the slide deck, where you can find these resources. Again, they're meant to be used as guidance towards your decision making, but they're not going to cover all the different ways in which you might implement your intervention. So now we're going to move from thinking about sample size to actually implementing the efficacy trial during Phase 2 and what the key components of that are.

Andrew McEachin (32:37):

So the very first thing to do when Phase 2 starts is to randomize students into the treatment and control conditions. So by this point, you would've identified the sample size that you need and how the intervention itself is going to be implemented. So you'll need to document that thought process and will be in the Phase 1 material and NWEA, we will work with you all to do the randomization. So in order to do that, we're going to need what we're calling a randomization roster. It really is a roster of students who are eligible for the intervention and has a few key components to it so that we can randomly assign students based on how the intervention is being implemented and then, provide that file back to you, so you can make sure that you’re providing an invitation to the intervention to the right students.

Andrew McEachin (33:26):

Then, you're also maintaining the control condition for the right students. So the file must include the following things and the first and the most important part is having the right student ID so that we can merge it with our assessment data and then, so it's linkable to the other parts of the data that are needed for the Phase 2 part of the efficacy trial. This is what we're calling the NWEA Map Growth ID. So for the school systems that are using Map Growth, they will provide us as part of the normal practices, a local ID that matches whatever they're using within their systems. Sometimes this is a statewide ID. Sometimes this is district specific, but whatever the idea is, that they're using for the Map Growth assessment is the ID that we need for the roster, because that's how we will ultimately link the test data with intervention data.

Andrew McEachin (34:21):

So the very first component, and this is crucially important, is it's going to have the appropriate ID that the district or school or school system is using for the Map Growth assessment so that we can match kids, and this is true for both the math and science challenge. The second is we need to have the right ... we need to be able to link students to their class. So is it for math? It might be third grade math for science. It could be ninth grade biology or something, but we need to have an anonymized ID at the class level. We need to have an anonymized ID that links the teacher and students, and an anonymized ID that links the school and the student. These IDs need to follow the student throughout the trial.

Andrew McEachin (35:11):
Importantly, what we mean here by anonymization or anonymized ID is that by looking at the IDs themselves, we should not be able to identify who that teacher is, who that class is or who that school is. So you'll want to work with your schools or implementation sites on creating these IDs so that we can track kids across these different groupings, but that they're also not revealing the identity of these individuals. Then finally, the course name. So what is the name of the math class that students are in or what is the name of the science class that students are in? The final thing ... Sorry this is not mentioned on the previous slide, it'll be important to work quickly to sign the data sharing agreement with NWEA and the intervention provider and/or the school system to make sure that these data can be transferred.

Andrew McEachin (36:10):

And there's an example of the data sharing agreement as well on Challenge.gov With the other resources. So the great thing is once you randomly assign students into the conditions, you can start intervening and you can provide the students the great program that you all have to offer and change the course of their school year. The important thing to keep in mind as a school year is unfolding is to track the original assignment of students, classes, teachers or schools. So whatever unit that we were randomizing and how the intervention itself is being implemented, you need to track whether or not there are changes to that structure. So our students changing classes, our teachers changing schools, do the school drop out.

Andrew McEachin (37:00):

Ideally, these changes wouldn't happen during the process, but we acknowledge that the school year is just a ... it's a very dynamic setting. Things come up. The most important thing to do is to document any of this, document these changes, because they're important for the efficacy trial. Again, there will be Phase 2 TA available or technical assistance, from apps to help with any of these things that come up, but it is really crucial to track these changes, and we provide an example of what that can look like on the Challenge.gov site. At the end of the trial period, so in May, part of what you'll provide NWEA is a data set that includes these sorts of changes that might have happened to the original random assignment during the school year.

Andrew McEachin (37:50):

Then, the last set of data and really crucially the most important part is tracking student participation in the intervention during the efficacy trial period. So without the school year or throughout the school year, you'll want to track the student's engagement with the intervention. Again, depending on the design of the intervention, you'll also want to measure what we might call dosage at a high level. So the number of minutes students are receiving, the number of lessons that they were assigned and completed, completion towards learning goals. Some other measure of the actual intensity of the intervention that students were both assigned to and completed. These data are important to help think about whether and how the intervention was or was not working.

Andrew McEachin (38:44):
These are also data that you will send NWEA at the end of the efficacy trial period and these data feed into the Phase 2 scoring, and we also provide examples of what this might look like in a simplified way, on the resource or the Challenge.gov site. So the last slide, before we go into a Q and A is on both of the challenge sites, so if you go to the Challenge.gov and you find the math and science challenges, if you go to the resource tab, which here is circled in a dashed blue, underlined in orange, you'll find a variety of resource documents that will be helpful for you. Many of which were highlighted today. Some of them provide guidance on the actual design of the intervention itself, which would be the practice guide.

Andrew McEachin (39:40):

Then, the next five documents really pertain to the efficacy trial and providing guidance to various components, including an example of what the data sharing agreement is, that you all would have to sign. Great. So I think now I'll turn it back over to Eden, to play emcee for the Q and A session.

Eden Baker (40:02):

Great. Thanks so much, Andrew. So as I said at the beginning, we are going to open up for question and answer now. So if you do have questions, please share them by the Q and A window in Zoom. We will get to as many as we can, live. If we don't get to your question today, we will be sharing written answers by the FAQ page on Challenge.gov. So before we jump into live questions, I just wanted to cover a couple that we received via the Eventbrite registration before today's session. So the first one was why is IES running this as a challenge rather than as a research grant? So we actually covered that question during our last webinar, and we're going to be sharing a recording of that session and a response to that particular question by the FAQs on Challenge.gov shortly.

Eden Baker (40:47):

So I would encourage you if you haven't already, to follow the challenges on Challenge.gov. So you're notified when the recording and the FAQs are published. Then another question that we received via Eventbrite, which I believe Andrew has largely covered during today's session, but I'll put it out there again, which was, is there a minimum number of students that providers should include in their interventions? So Andrew I'll pass it over to you just in case there's any extra context you would like to share for that one.

Andrew McEachin (41:16):

Yeah. The simple answer is yes, but the actual number to answer that question is going to depend on the design and implementation of the intervention itself. So, if it falls into one of the examples that we showed, those tables could be quite helpful. If it doesn't deviate too much, the tables will also be helpful. If it's quite different than those, then the power up tool and the links to that would be helpful resource, to answer that question.

Eden Baker (41:47):
Great. Thanks Andrew, and I'll just make another plug for Challenge.gov, and the Resources page. So all of the tables that Andrew just reviewed are available in the guide to RCTs, which is on the Challenge.gov listing. So you can take a closer look at them there as well. So another question that we just received was, can you please describe the trade-off between powering a study to detect an effect size of 0.3 versus 0.4 or 0.77. Andrew, I'll pass that one over to you and then Sivan, you're welcome to jump in as well.

Andrew McEachin (42:21):

Yeah, so it's a really good question. So, this question relates to the grand prizes are at a really high threshold and they are to reward a pretty remarkably impactful intervention. So for math, it's 0.77 standard deviations, and again, for science it's 0.4. In our power calculations, we anchored it, towards something that's both meaningful, but short of that grand prize benchmark. So, the question is about the trade-off between thinking at a sample size that might only hit the grand prize, but doesn't quite meet the requirements for an effective 0.3, and the trade-off here is that it'll be easier to find a sample size to hit one of the grand prize benchmarks because you would need fewer schools and then students, classrooms, teachers, et cetera.

Andrew McEachin (43:12):

The danger in doing that for the challenge is you might have ... for the Math Prize, for example, say your intervention had an effect size of 0.5, which would be pretty remarkable, but is short of the grand prize, and you only found a sample size that was relevant for the grand prize. We might not be able to detect that being statistically different from zero. So you might have a really nice intervention that had a really big impact on students during the school year, but because we can't say it's statistically different from zero, it might hurt your eligibility towards the other prizes in which you might have been eligible for otherwise. So by powering your sample size to 0.3, you're going to qualify in a sample size setting 4.4, 0.77, the grand prizes for the other two, but you also are going to make it, so it's more likely to find an effect that's smaller than those.

Andrew McEachin (44:11):

The reverse is not true, if you only find a sample size that's geared towards the grand prizes, you might make it, so it's very difficult to actually find an effect that's smaller than those.

Eden Baker (44:26):

Great. Thanks Andrew. So another question was why an RCT design for this challenge? What are the specific benefits of this approach?

Andrew McEachin (44:39):

Yeah, that's a good question. So I think there's a couple of them. One of them is given the time of the challenge, it's one of ... it's easier to get information faster back to the relevant stakeholders. The second is it provides a really clean analysis. So it's a lot easier to compare the effect of intervention A versus B, versus C. Then, the other is that it often has a smaller sample size requirement than other designs that you might have. So if we're thinking about the feasibility of actually implementing this in a number of schools, et cetera, you would need fewer schools in an RCT framework than you might for maybe a quasi-experimental design.

Eden Baker (45:31):
Great. Thanks Andrew. Then, another question, which is what can students who are in the control group do while treatment groups are receiving the intervention?

Andrew McEachin (45:47):

Yeah, that's a great question. So it's more or less up to the schools. The way to think about it is you wanted to be as close to the experiment, the experience of what the treated kids would've had in absence of the treatment. So how would they have spent their days if they were not assigned to the intervention? So typically, you'd call that business as usual practices, and I realized that nothing in the past couple years has been business as usual. So it's really how else would the school and/or kids, teachers, et cetera, spent their days if they weren't in the intervention. So they can be in other programs. They can be doing other things.

Andrew McEachin (46:27):

It's important to keep in mind, whatever the control kids are doing, that's what you're comparing the treatment against and that's how you're measuring efficacy. So if they're spending their days in really effective other programs, then you're going to be comparing potentially two effective programs against each other and it's harder to find a sizable impact. Yup.

Eden Baker (46:52):

Great. Thanks Andrew. Then there was another question, can students use any other assessments as a baseline or do they have to do NWEA before the implementation period starts?

Andrew McEachin (47:14):

Yeah, so really, you can ... the students and schools can be doing a variety of activities prior to the start of Phase 2. The most important thing is that after you've randomized ... so after Phase 2 has started, whatever the treat ... after Phase 2 has started, that the treatment kids are in the intervention and that the control kids are in the business as usual setting. So, if the intervention itself was being implemented, kind of across all of the students prior to the start of Phase 2, it does mean that in that example, once you randomize, the kids who were previously exposed to the intervention, but were assigned to the control condition, are no longer going to be exposed to that intervention anymore.

Andrew McEachin (48:01):

So the school system has to be aware of that and really, once randomization starts, they're shut off from the intervention and are back into a business as usual condition. So it can be done just with the understanding that once Phase 2 starts, it can't continue any further for the control condition.

Eden Baker (48:21):

Great. Thank you, and then, I believe this question might have come up last week as well, but we can cover it again now, which was can students who are not eligible also receive the intervention and do we need to track any data for those students?

Andrew McEachin (48:36):
Yeah, it's a really good question. So the short answer is yes, they can receive it. The information that is going to go into the Phase 2 judging criteria though, is just the effectiveness of the intervention for those students who are eligible. Really the data tracking, et cetera, at least at the student level is geared towards the eligible students for the challenge. So you don't really have to track what the ineligible students or the students who are not eligible for the intervention, are doing with the exception of, it's somehow like a teacher or school level change affects the treated kids. Then, you do want to track that, but at a student level, the students who are not eligible can receive the intervention and you don't need to track them. It's really about tracking the students who are eligible for the intervention and tracking their data throughout the school year.

Eden Baker (49:35):

Great. Thank you, Andrew. We also just received a question about the eligible students for science. So they have said, "Is that the lowest performing on state tests? Please define." So the eligible students for the Science Prize are those who score on the bottom 25th percentile based on the NWEA Map Growth science assessment national norms. I believe Sivan has just posted a little bit more information about this in the chat. It's also outlined in the rules on Challenge.gov, and we can share a link directly to that as well. One thing to flag here as Andrew mentioned, is that students who are not eligible are still able to receive that intervention.

Eden Baker (50:16):

It's just the data for those eligible students that will be considered as part of the evaluation. IES also does encourage interventions that meet the needs of all students for the Science Prize, including students with or are at risk for disabilities. So I'm looking at the Q and A window now. I believe there was a question about NWEA matching. I'm wondering if that might be specific to that particular entrant, it might be worth us following up with you directly separately. So we can just share the NWEA email address in the chat there so that you can reach out to that team specifically and hopefully, help follow up on a potential match. We've covered all the questions that we have live, I'm going to give people maybe 30 more seconds in case there's anything else that might come up.

Eden Baker (51:22):

Otherwise, you are also welcome, if you think of additional things to reach out to the challenge team. After today's session, if you have follow-up questions and we will get back to you that way too. I'll give people a few more seconds.

Andrew McEachin (51:58):

I see a question about, "Does NWEA have information on students with disabilities?" So those are typically data that we don't necessarily have for all of our partners. It's kind of a case by case basis. It sometimes touches up against things like FERPA, et cetera. Again, that email address that I think has been shared a couple times would be a good starting point. If you have questions specific to a site or context or something around that question, please do use that email to reach out to us.

Eden Baker (52:34):
Great. Thanks Andrew. We also have another question here about ... so it says, "How are teacher interactions included in the analyses? For example, if an intervention includes instructional resources for teachers, for example, supplemental lessons, do we need to track exactly what resources each teacher utilizes?"

Andrew McEachin (52:57):

So for the efficacy evaluation, I would say that's probably not as important for our side, but I don't ... Eden, if you want to speak to the broader implications for the Phase 2 submission that the sites will submit if it is necessary for the intervention, it is something that you'll broadly want to document as part of the application materials in both Phase 1 and Phase 2, but for the efficacy side, in terms of providing data to NWEA, that's not something you would need to provide us.

Eden Baker (53:29):

Yeah. I can chime in with a little bit of extra context about the project more broadly. So this is something that would be considered as part of the Phase 2 criteria. During next week's webinar, we are going to dig into a little bit more around what you should include in your intervention proposal, which is part of your implementation plan, which is where I think you would cover things like resources for teachers and how you are considering some of those things. So Naomi has just popped a link into the chat that has the sign up for that session. So I would definitely encourage you to attend that and then, also take a closer look at the Phase 2 selection criteria, which does have some more information there on considerations for things like scalability.

Eden Baker (54:30):

If there are no more questions live, I think we can jump to the next slide just to remind everybody about the technical assistance that's available over the next few weeks. So, as I mentioned, you're welcome to email the challenge team, if you think of additional questions after today's session that we didn't cover. We will also be hosting a webinar next week, which as I mentioned, we'll be focused specifically on cost analyses and implementation planning. So that will cover some of the things you should be thinking about in terms of your plan for collecting cost data, as well as some of the things you should be considering as part of your intervention proposal in your implementation plan, and the link to sign up for that is in the chat as well.

Eden Baker (55:14):

Both Andrew and Sivan is also going to be available for office hours, and we've popped a link in the chat for how you can sign up for both of those sessions too. Andrew's session is on September 7th and Sivan's session is on September 8th. Both of those will be conducted in a group setting and be available to both potential entrance for the Math and Science Prizes, and there's more information on the sign up pages there. Then, just a general reminder, submissions close at 5:59 Eastern on September 30. Again, the email address for the IES Challenge team is on the screen, so Challenges.IES@ed.gov, so please reach out to that team if you have any additional questions that we didn't get to today.

Eden Baker (55:58):

Thanks again so much for joining. We hope to see you at next week's webinar and yeah. Thanks everyone. Thank you, Andrew, Sivan, and team. Bye.