November 2025 Data Clinic EDFacts Data Analysis and Use-20251112_185354UTC-Meeting Recording Transcript

November 12, 2025, 6:00PM

53m 21s

Alex Gabriel 0:12

Hello, welcome everyone.

It's good to see all of you here. Um, and we'll give folks a couple more minutes to trickle in.

All right. Well, I think in the interest of time, we can get started. And I know Lynn wanted to say a few words to start us off.

Lynn Bauer 1:33

Thanks, Alex. Good afternoon, everyone. Thanks so much for joining us. It's great to see we've got a good group in the room here. For those of you that I haven't met, I'm the project director for NDTAC and I wanted to join just very briefly here to just discuss.

To matters that are going on with the contract and any concerns that may be out there related to the shutdown and your accessibility to us really at the at the root of what I wanted to say here is that we're here, we're our team is fully here. We are in a great position in that respect.

We know that some of our colleagues in in other technical assistance centers are not in the same position. You may be experiencing that where it's difficult to reach even the vendor side. So that would be me, Alex, Obioma, Darrin, you know in other centers where you, you know sort of the lights are out.

Not only at Ed, but on the vendor side of things. That is not the case with us. All of our staff is fully available. Our ticketing system to receive your requests for support is fully functional and.

We will take any in our requests. We are available to meet with you one-on-one. We are responding to your requests as they come in. We are hosting monthly events to make sure that we get together and maybe anticipate some of your questions or answer them live together.

And we know we we know that you know the the government may come back soon

looking promising. We'll hold any requests for technical assistance that we are not able to answer. Those are typically legal in nature for the department and upon the return they would.
Will be in touch with you and we will be sure to get with them as soon as possible when they do come back to brief them and anything that feels outstanding and they will reach to you one-on-one to make sure that you have the support that you need.
So that was all I wanted to say just to make sure that you know we're.
Business as usual as far as the center is concerned and will continue to be. So hopefully, like I said, it'll be back soon. Our program officers, Heather and Jeff will come back and join us and you'll see them on our next live engagement.
That's all. Thanks, Alex.

AG **Alex Gabriel** 3:59
All right. Thank you, Lynn.

CK **Claire Kelley** 4:00
Thanks for the introduction. So thank you all for joining us today. Today's webinar is on ED Facts Data Analysis and Use. We'll walk you through the whole pipeline from data collection through data analysis trends and all sorts of exciting ways that you can apply your data to make your programs better.
And our speakers today are me, Claire Kelly. I'm a Senior Data Scientist and the Data Lead for NDTAC. And I'm joined by Alex Gabriel, who's a TA and data specialist on this project and sort of our resident expert on this data.
So we're going to go through basically 4 main steps. First, we're going to talk about an introduction to Ed Facts as well as the data collection, submission and review process. This will be sort of the 101 level introduction. We'll then follow that up with the Title One Part D data analysis, talking about data sources.
Introduction.
Analytic concepts, a demonstration of how you can use the data to answer some key questions as well as some sort of basics of data calculation. And then we'll move into a bigger picture view of how title one Part D data can be used, as well as some data, some use cases of data-driven decision making, and how this data can support your own goals.
Setting We'll finish up with question and answers in any remaining time, as well as an

evaluation survey where you can tell us how we did today. So I'm going to pass it off to Alex, who's going to start off with the Ed Facts introduction.

Alex Gabriel 5:31

Yes, thank you, Claire. And if you have any questions during the presentation, feel free to use the chat and we'll be monitoring that.

And so yes, I'm going to be going over an introduction to the data that's collected and how it's collected. So we have everyone on the same page and so to start the process in view involves collecting and reviewing the data.

Submitting it to Ed and reviewing it and updating and correcting it, and then ultimately releasing that data and using it to make decisions.

The collection and submission timeline is from July 1st to June 30th. That's when the data is collected. So for 2425 school year that would be from July 1st, 2024 to June 30th, 2025.

And that data is submitted to ADD in February of 2026.

And the purpose of ED Facts data collection is to have robust performance data at the core decision making and policy making education that is both at local, state and federal levels. The goal of this is to reduce state and district data burden, streamline data practices.

Improve data capabilities by providing resources and provide data for planning, policy and management.

Now the datas that are data files are collected through Edfacts are organized into different file specifications. These are both at the state agency level and the local education agency level. So there are files for academic achievement.

Participation or demographics outcomes while students are in facilities, outcomes after students exit those facilities and an optional data file on.

Proficiency on a state standardized tests.

Now there are several different pieces and data points that are collected through Edfacts. Those participation files include data on age, sex, race, English learner status, student disability status, long number of long term students and the program type.

For academic achievement, this includes growth on reading and math, initial and follow-up assessments, looking at if students improved more than one full grade up to one full grade.

If there is no grade level change or a negative grade level change.

The outcomes in facility and outcomes 90 days after exit. A look at the data points

for students who earned high school course credits, enrolled in GED programs, earned a GED, earned high school diploma, enrolled in job training.
Obtained appointment was accepted or enrolled into post secondary education.
And then finally is the optional state standardized assessment data files, which includes students who took statewide assessments, students who took statewide standardized assessments and achieved proficiency.
And the number who took those assessments and did not achieve proficiency. And once again, those are optional for the state standardized assessment.
Now the data collection and reporting process flows from facilities all the way to the Department of Education. So subpart one facilities would report their data to the state agency.
And the Support 2 facilities would report their data to that LEA and then the LEA and state agency will report to the State Education Agency who then send it to the US Department of Education.
There are also several resources on data quality. We won't be covering data quality in depth in this presentation, but if you want to learn more about it, you can review the Business Rule Single Inventory on the Department of Education website. We also have a Data Quality tip sheet on our website.
And then the our 2023 Data Clinic focused heavily on data quality. So that would be one to rewatch and review if you're interested in that topic.
So next I am going to cover analyzing data.
There are several different ways you can access Taiwan Part D Ed Facts data. There is the Fast Facts website, the Data Explorer, Ed Data Express, and then of course the internal data that you'll have in your.
Or state data systems. So the Fast Facts website or Fast Facts is on the NDTAC website and includes very high level information on state level demographics and programs.
And it's good for looking at three-year trends.
The Data Explorer is more interactive. It lets you get more granular information and you have up to five years of data trends in there, and it's usual for generating visual and customizable summaries and views.
Both of those data are pulled from ED Data Express, and so ED Data Express is the public release of all Title I, Part D and all ED Facts data, and so it's a useful way to download that data.
It's comprehensive and includes data starting from school year 2010, 2011, and you

also have a geographic map if you want to explore visuals in that way.
Now finally there is the internal data that you all collect and we are suggesting that this is the best option for data-driven decision making because it's timely complete. All these other data sources go through a process of.
Data cleaning and then uploading, which can often cause delays and then internal data can be more granular and more complete because the public release data is. Often suppressed and so you'll have less information. So these are the different ways to look at and access your data and depending on what you want to find or use it for.
Will guide you in which of these sources will be the best to use.
I see a question from Sam about in the chat about when the Data Explorer will be updated and so I'll just share that the yes the.
Data has been delayed, obviously because of lots of things are going on in the federal government and so even the Title 1 Part D data for unedited expressed has not yet been released. We are working on updating the Fast Facts page and so that's something to look.
Out for, but yes, those other sources of data are not up to date due to no public files being released.
And this is another reason why internal data will be useful if you are looking for to make timely decisions.
So I'm going to cover some data vocabulary that'll be important when we're looking at and thinking about how to analyze data. So measures to keep in mind, we'll talk about proportions and counts. Trends is looking across, looking for patterns across multiple years of data.
Outliers are data points that are outside an expected range, and comparative analysis involves examining differences and similarities between groups, time periods, or sites in order to gain insights.
So first I'm gonna talk about trends.
Uh, first thing we think about when we're looking at trends. So this is of course looking at for patterns overtime we look for how big the change was.
What direction is the change? If it's going U down, if it's retty flat, or if it's kind of going zigzagging U and down?
And so in this example, even though we do see a decrease from 2017 to 2018, we see a general trend of the percent of students.
Improving on academic tests in reading and language arts, increasing over the

course of five years.
Now we're going to talk a little bit about different measures and how to make sense of those. So counts show the total number of students. So on the graph on the right, we see the racial and ethnic composition for all students in.
The United States in subpart one programs.
Percentages. We show the proportions of students.
And so in this example, we're looking across time again and then we can see that the percentages are relatively stable, that these different categories aren't changing much.
But when we can go to the top and we look at the counts, we do see a pattern of the overall number of students in Title 1, Part D programs decreasing. So while the proportions of students.
Are remaining the same across race ethnicity. The number of students are decreasing over time.
Next is outliers. So an outlier is a value that is notably larger, smaller, or has sits outside the cluster of other values.
So in the graph on the right here, we're looking at career and academic outcomes in facility. We are looking at students enrolled in GED programs and those who earned a GED.
And.
So all of these are different LEAS, each of these points. And so we see this point here at the top in the red circle and that is showing a an LEA in which there are a lot of students enrolling in GEDs.
But not many obtaining the GED. So because this is point is so different from the rest of the data points, it we're considering an outlier and from there you can decide to.
Have investigate that further, understanding what that programming looks like in that LEA and understanding.
Why they're so good at enrolling students in GED programs and why fewer students are graduating or obtaining a GED. And there could be a number of explanations.
This could be a data quality issue, something could be misreported.
This could be the kind of focus of the program and just given the duration of time that students are in the facility, they may be able to enroll, but they're not able to obtain a GED.
Next I want to talk about comparative analysis and so this is the different a few different examples of how you can make comparisons. So there's state to national

which compares state outcomes with national title one Part D trends to identify alignment.
Areas for improvement.
You can look at the difference between Title 1, Part D programs and public school populations.
And you can look at the title and part D population compared to other high needs students like students who are in foster care or special education.
This will look similar to the what we shared in the data collection, but yes, there's several measures that you can use and ultimately from the data that you collect and you'll be able to use these measures to calculate.
Different outcomes and metrics.
So for example, for demographics, you can use any of that information to calculate demographic proportions, like the percentage of EL students, the percentage of long-term students.
Using the outcome data, you can calculate outcome proportions and so this is using information both from the demographics file and the outcomes file. So you can calculate the percent of students who earned a earned high school course credits.
In assessment taking proportions, you can look for percentage of long-term students who took an assessment and calculate that using the students who took an assessment out of the total number of long-term students.
And there's also academic achievement proportions, which looks at the percentage of students improving on academic assessments, and you can calculate that by taking the count of students to improve more than one grade.
Or up to one grade divided by the count of students who took an initial and follow-up assessment.
It.
So some example questions when analyzing the data are what proportion of students are identified as English learners in the most current year? How does the proportion of identified English learner students compare to the proportion of Hispanic students?
How does the proportion of identified English learner students in Title 1, Part D come prepared to the proportions of students identified as English learners in traditional public schools? And how does the proportion of English learners and Hispanic students differ across LEAS?
So we're going to take these questions and apply them using an example.

So here we have a table that looks at students in Title I, Part D and students in public schools for different LEAS.
We have the number of students in Title I, Part D, the number of Hispanic students, percent of Hispanic students, number of English learners, and the percent of English learners in Title I, Part D.
And we are comparing this to students in public schools, the number of English learners and the percentage of English learners.
And so this data is also from Edfacts. It's submitted by Title 3, and it's also publicly available.
So what we are going to be looking at is the percent of Els in Title One Part D compared to the percent of Els in public schools.
Uh, for each LEA.
And so when we're looking across each LEA and comparing these two columns, we can see that there are a lot more English learners in public schools than are reported in Title 1, Part D programs.
In fact, we can see just looking at the English learners reporting Title I, Part D, there are very small amounts. Most are zero and beyond that they're all the LEAS report single digits.
Whereas in public schools we see a larger proportion of students being English learners.
Next, we're going to look at the percent of English learners and percent of Hispanic students in Title I Part D programs. This is because in some areas of the country, the population of English learners are Spanish speakers, so you can use Hispanics.
Someone as a proxy and to see if there are similarities. Now of course you have to consider your local context and then in some in some states and of course some LEA's this will be widely different. Of course in like Hawaii, Alaska and Montana, those are all examples where.
Um.
Spanish is the not the most common language spoken among English learners, so always consider your local context when making this type of comparison. But looking at these two data points we can see that.
Even in facilities where there are.
A higher proportion of Hispanic students. The percent of English learners remains relatively low.
So all this could indicate that the percent of E LS reported in title one by D might be

under reporting.
And here are the key takeaways we discussed. Most LEAS reported 0 English learners and at most 4%. The percentage of English learners reported in LEAS is lower than those in public schools in the same LEA.
Facilities with high proportion of Hispanic students still have a low proportion of English learners, which ultimately may indicate that there is an issue of under identifying English learners in facilities.
Next I want to walk through a another example. This is long-term students and math assessment progress and participation in juvenile detention facilities by LEA. So this uses data from.
The participation files and the assessment files.
So from the participation file we are using the number of long term students.
Looking at that data point, we have the number of students who took assessments and then we have the assessment score changes from negative grade change, no grade change up to one full grade change and more than one grade change.
In the outcome measures we calculated here are the percent of students who took assessments and the percent of students who improved.
So we are going to be looking at.
Two data points, the number of long term students and the number of or the percent of students who took assessments.
And so in this first LEA, we see we have 168 students and only 11 took the assessment.
In the second LEA we see 210 long term students and 79 took the assessment.
In 3rd facility we have 78 students and 100% took the assessment and in the final one we have 31 students and 0% took the assessment. And so looking back at these three, we do see quite a bit of variation.
And so to remind everyone that the long-term students refers to those who are in facilities for at least 90 days, and in this time they're expected to take both an initial assessment and a follow-up assessment.
And so when we're seeing.
I think a couple of things to keep in mind is the facility size. So we see a large facility of 168 students and only 11% are taking tests and we can compare that to LEA where 210 students.
Are present and 79% took tests. So this is an interesting comparison because we can see that these facilities could be peers in the way that they have a similar number of

long term students.
But they have very different rates in the percent of students who have taken tests. So one thing to do here is be able to learn from LEA 2 about and what are the practices that they do to ensure that students are able to get those tests.
And pair them up or use those findings to help inform practices that are going on in LEA one.
Now similarly we have LEA 3, which 100% of students are taking tests.
And this is also an opportunity to learn from them. But one thing to consider also is the facility size. So given that they're a smaller facility that might be more feasible or easier to test all students, but I think it's still an opportunity to learn about those different practices that they are doing.
Now in this final one, we have 0% of students are taking assessments, but we have around 31 students. So while I would like to see them improve, it is understandable that that 0% also comes from the fact that it is overall a small facility.
And one last thing I want to say is before we even look at the percent of students improved, it's really important to see the percent of students who took assessments, because in this first LEA, 0% improved, but you're only testing 11% of students. So when you have such a low rate of students.
Taking tests, the percentage who improved is not going to tell you a lot of information, so that is won't be a useful measure until the percent of those who took assessments increases.
With LEA two and three, we can see that with higher percentage of students taking tests, we can really start to look at the percent of students who improved and be able to make decisions on that.
And this slide summarizes kind of those key takeaways. Again, the LEA one assessed only 11% of the 168 long-term students, showing a need for improvement in their testing practices. LEA two assessed 79% of the 210 long-term students and can serve as a.
Model for LEA one.
LA 3 assessed 100% of the 78 long term students, also offering as a an example of effective practices.
And then comparing LE's of similar size can help identify strategies that are transferable and practices that are transferable.
Transferable.
And keeping in mind that when you look at test improvement rates, you really need

to first look at the participation because low participation reduces reliability of those results.
Alright. And so next I will pass it over Claire to talk about what to do with the data after you've analyzed it.

CK **Claire Kelley** 31:46
So at this point, you should feel like you know a little bit about how we collect the data and also about how you can analyze it. And in this section, sorry, excuse the coughing. In this section, I'm going to talk more about how you can use it to inform the way you practice and the way you work within your organization.
So before you start any data related questions, it makes a lot of sense to be very clear about your end goals. So if you start a project knowing what questions are we trying to answer.
What does the data tell us? Hopefully using some of the skills Alex just talked about. What is missing from the data? What contextual pieces or additional factors are not represented in the data? What are things that are going really well? What are things that are going really poorly?
Or things that we could do better next time? And how do these directly inform changes we can make within our organization? So if you start any data project with these sort of 30,000 foot view level questions, you'll end up with a product that is more useful and questions that are answered more accurately.
The ways you can use your own internal data are essentially infinite, but a few sort of most common use cases include improving program operations, so looking at how many people are taking assessments.
And what you could do to change that. I know that's something that a lot of programs struggle with or ensuring accurate identification of students with specific needs. So these are all cases where your analysis of the data would lead into changes in the way your program operates.
You can also use data as a metric to measure how your program is improving outcomes. So after you've identified the relevant outcomes, you could look at how changes impact the number of students receiving a high school degree or diploma. You could look at how a new educational program.
This program impacted the percent of students who are proficient on different assessments and this this allows you to have a sort of more evidence based take on programmatic changes you're making. This is also another common use case is to

examine the relationship between variables, so this allows.
To answer questions like which outcomes are related to long-term outcomes, those it can be very difficult to see change in long-term outcomes or how do the demographics of your own program align with the demographics that we would expect based on comparison to other populations and a lot of these.
The relationship between variables can get as complicated as you want it to, and is really only limited by your imagination and the your statistical abilities.
So thinking first about improving program operations, we can imagine we're asking the question what proportion of long-term students are taking initial and follow-up assessments. Following the analytic strategies that Alex just shared with you, you'd start by taking the count of students who are take who have completed the assessment.
And divide that by the count of long-term students. You're using your understanding of how the program works to know that the correct denominator is only long-term students, because assessments are only offered to those who are within facilities for a sort of long-term status.
Then you can dive into the data and start pulling it apart. Can you identify any sub grantees that are excelling or on the flip side, any that look like they need sufficient support when you're looking at the distribution of the percent of students taking a? In an assessment, are there any outliers that sort of jump out to you either in a positive or negative direction? You can also use this to look at data quality pieces. Are there any sub grantees in which you see more students reported as taking assessments than there are students?
Now, you might think that is a very low-hanging fruit, but we do see multiple people every year supporting data or submitting data that suggests there are more students taking assessments than they had students, which is sort of a clear clue of a reporting error. And this can be a place where we then open the dialogue with sub-grantees to learn more.
About either their data submission process, including verification or the sort of underlying processes and practices that go into supporting more students to take these assessments.
Another example you might ask, how does the proportion of students with disabilities in this facility compare to the proportion of students with disabilities in public schools in my area? So this question is basically getting at does the profile of disabled students in my school look different than the public school?

To support this analysis, you start by calculating the proportion of students with disabilities and compare to data on students with disabilities in the state or the LEA. Now, which level of comparison you use really depends on your own local knowledge if you know that your.
LEA is very different than surrounding LEAs within your state. You might be better off comparing to the public school information from your LEA, but if you're interested in a bigger picture state level analysis, that might be when you zoom out and compare yourself to the state in general. And this can help identify sub grantees with either over identification or under identification of students with a.
Disabilities and be a way to open the conversation with sub grantees about the screening processes that they use.
Um.
When you think about improving program outcomes, you can also think about things on a time scale. So if you start your analysis thinking about early outcomes, so these are milestones like earning course credits or enrolling in a GED program that are then necessary for later outcomes.
A lot of the time, as I think Sam already mentioned in the chat, students may not be in the program long enough to achieve the final milestones, like a high school diploma or a GID. So starting by measuring the foundational milestones gives you some insight into the data, even if you don't have the full picture.
Short-term outcomes are also most directly linked to service delivery, and these are the most sort of immediate and most tractable in terms of making programmatic changes. And these are things like performance on assessments or earning high school course credits.
When we're thinking about these kind of questions about how to improve program outcomes, it's also important to look at what sub-grantees themselves think are important. So which data points do they outline in their application, or which ones are most specifically related to the purposes of the program?
There are a lot of LEA level or sub-grantee level programs that have a particular focus, especially on things like employment training or workforce preparedness. So if you're evaluating programs like that, you'd want to hone in on the outcomes most directly related to the program's focus or the program's application.
As we prefaced in the beginning, there's sort of an infinite number of variable relationships you can explore within the data. For example, you might want to know what could be driving the increase in the number of students identified as English

language learners.
If you were to go back to your data and find that there's a notable increase in the number of students identified as Hispanic, this is giving you a clue into both why the identified the number of identified ELL students is increasing and also what supports they might need in terms of what language they're speaking.
And this could be most accurately identified by looking for the count of both English language students, English language learner students and Hispanic students across time.
You can also use some of the short-term outcomes as I think we already prefaced as a sort of early indicator. So if you're looking at, if you want to drive earning GEDs, looking at trends in GED enrollment is a good sort of leading indicator and may allow you to get.
A sense of where the trend is going before you're able to get the final results, and you can also use that to look at where in the sort of intervention pipeline subcountries can best meet the students.
This is also useful for supporting Title One Part D administrative processes. You have a wealth of data that can be used to support application review and selection, and during this process you'd want to use the data to design applications.
Assess the needs of your subgrantees and ensure that the activities proposed by applicants fill well documented gaps. Similarly, this data can be used to support subgrantee monitoring and this can be both a performance and a data quality perspective, so.
Are the programs submitting high quality data that's free from errors and is complete? And is that data showing progress on the most relevant performance indicators? Is your investment in the specific applications translating into improved outcomes on the relevant assessments?
This can also support facility eligibility, help identify common problems with in demographic identification and other steps to make sure that programs are aligned with their intended purpose. And then of course, as a technical assistance center, we are also very keen on the ways that data can be used.
To support technical assistance. So you can look at trends in your own data to figure out what topics might people need additional instruction on, which sub-grantees should be targeted for support or be provided with specific peer learning opportunities. And I think this is one of the most powerful and sort of often under or overlooked.

Aspects of data.
If you've been around the technical assistance space for a while, you've probably heard ad nauseam about SMART goals, and we're not exempt from them here. When we're thinking about data-driven decision making, developing SMART goals is one of the most effective ways to translate your data into action. What do you need to be a SMART goal? You need to be.
Specific, measurable, achievable, relevant and time bound. So specific. That means we need to be named in a clear way so that the goal is so that everyone who's working towards the goal knows what they're working towards.
An example might be we want to increase the accuracy and consistency of identifying English language learner students in juvenile detention facilities. We now now need to make this ask measurable.
So that means we need to operationalize this and we're saying the metric we're measuring is the proportion of facilities that adopt our standardized English language screening protocols and complete.
And the comparison between their English language learner identification data with the public school's English language identification. So the benefit here, sorry, excuse the coughing. The benefit here is that now we are measuring our goal, which encourages people to drive in the same direction.
Direction essentially in their attempting to make these changes. It also is reasonable to assume that these goals should be achievable. So if you make goals that are not feasible within the time period, it demotivates people to work towards them and it devalues the goals in terms of being a realistic.
Representation of the kind of process progress that is possible. So for example, in this use case where we're trying to improve the identification of English language learner students, achievable steps might be that within this year we'll identify standardized identification.
For calls and provide targeted training. An example of something that would be less achievable would be making your initial goals something like 100% of students will receive a full English language learner screening in all facilities at all times. And the the less realistic your goals are, the less sort of.
Easy it will be to keep people on track working towards those goals. Of course, these goals should also be relevant. They're an infinite number of things we might want to fix about our programs, but we want to focus on the most relevant goals, especially those tied to the sub grantees applications.

And time bound. So that is there should be a specific named time period. In this example we might want to make multiple time bound goals starting at in a shorter time frame saying within 12 months we want to implement these standardised protocols and going out through.
A longer term goal and saying you know after 24 months we want to be in a place where we are really able to evaluate the improvements in English language learner identification within our sub grantee.
Alex, how are we doing on time?

45:35
We're good.

Claire Kelley 45:36
OK, this is another example that I'm I also think is is a good way of illustrating this. For example, if we want to increase the proportion of students who participate in initial and follow-up assessments, this is probably one of the more common problems that Title One Part D programs face.
We could make that measurable by saying let us increase from 46% to 61% or 51%. This should be achievable. So if you know that your existing level of.
Assessment is 46%, you want to make an achievable goal that isn't let's say 100% completion because that's not going to be feasible and we're going to build on existing practices and scale strategies that are already in use within the most high performing facilities.
So that we can improve our overall average. This is relevant to the program goals because better data collection helps us have more comprehensive information about program effectiveness as well as allows us to better assess the effectiveness of reading and math interventions.
And in order to make this a true SMART goal, we should also make it time-bound. So we say let's make this 15% total increase over the next three years a 5% annual increase. So now at this point, our goal is both specific, measurable, achievable, relevant and we hope time-bound.
This is just a series of resources that we have. Any of these are available for you to explore. I'd encourage everyone to start with the ED Facts tip sheet. That's a great place to get started and you'll see also the resources that were discussed in different segments of the presentation. For example, we have the recordings from the 2023.

Free Data Clinic, the business rule inventory that Alex mentioned as one of the data quality checks, as well as the Ed Data Express and Fast Facts places where you can access the publicly available data.
And at this point, Sam asks whether the PowerPoint will be emailed out. That is true. Yes, the PowerPoint will be emailed out.

AG **Alex Gabriel** 47:55
Yeah, and we have reached the end of the presentation, so we can open it up for questions. Feel free to put it in the chat or use the raise hand function.

CK **Claire Kelley** 48:10
Oh, Francis is starting us off with a very interesting question and one I'm particularly interested in. So Francis says Connecticut is a small state and the number of students participating could range from 10 to 1000. Analyzing small data sets can be very challenging. Any ideas, Alex?

AG **Alex Gabriel** 48:28
So yeah, this is a good question. And yeah, it's a good point that you want to consider the size of your programs because it can affect the way you look at things. And so for small studies, we say first starting at looking at counts or then percentages, because if you're looking at proportions, you might see.
Lots of variations simply due to a small number of students, changes in a small number of students being in the program, but in some ways you have an advantage in being able to understand the whole all.
See a big picture of all the students that are in there. So I think with a small data set, it's nice to be able to get granular and be able to see looking at very specific, looking at very specific programs and looking at specific demographics fields that have.
Have a fuller understanding of what is who all the students are and what is happening in those programs.
And.
I think, yeah, there are different ways you can look at the data. So for example, if we are looking just at the assessment piece, you can just kind of look at the numbers here of the number of long term students and the number who took assessments and really use that more as a guide to.
What is um of who might need support rather than looking at those percentages,

which could be very volatile.
Is there anything you would add, Claire?

Claire Kelley 50:04
I would also add that it depends on what question you're trying to answer. So for a question that is looking at multiple LEAS together on multiple time periods, if you have several 100 or especially in the case of 1000 students, that's not small data and you're.
Relatively, you can relatively confidently use any of the techniques we discussed today. If you have truly smaller sample sizes like your LEA's of 10, those are going to always be trickier. They're going to involve more manual review because the influence of each outlier will be larger. It's also important to.
If you have deeper questions to consider more advanced statistical techniques, there's a whole branch of statistics focused on nonparametric analysis of small data sets. So that's something you you or your statistician should consider if you do want to go down the route of doing more analysis with truly small data sets.

Alex Gabriel 51:00
I'd say I'll add that one of the benefits of also having a small data set is that you can get better quality of information. So one of the big things that quantitative analysis is good for and a lot of things that we covered is because.
You have a lot of data, but when you have a smaller data set, it's easier to get information and kind of get those individual stories from each facility and on those students. So it's a different approach to thinking about.
You know, getting information and kind of getting a fuller picture of what's going on.

Claire Kelley 51:44
I also wanted to elevate that somebody in the chat, they've put the link for the evaluation of the session, so feel free to tell us about what you did or didn't like.
Does anyone else have any other questions or things that were raised in the presentation that you'd like to hear more details about?

Alex Gabriel 52:30
I think if folks do have any other questions, feel free to reach out to us. As Lynn said,

we are still operating and available. Otherwise, I just want to say thank you all for coming and it's good to see all of you again.

Claire Kelley 53:03

Oh yes, someone asked if we're sharing the slide deck. Yes, this will be posted on the website along with a recording of today's session.
But.

stopped transcription