**WorkforceGPS**

**Transcript of Webinar**

**Building Upon Performance Data and Reporting to Conduct Evaluations**

**Thursday, January 21, 2021**

*Transcript by*

*Noble Transcription Services*

*Menifee, CA*

JONATHAN VEHLOW: Welcome to "Building Upon Performance Data and Reporting to Conduct Evaluations." Without further ado, I'd like to kick things off to our moderator today, Ingrid Schonfield, work system strategies/evaluation and research hub dissemination lead, Employment and Training Administration with the U.S. Department of Labor/Safal Partners, Inc. Ingrid?

INGRID SCHONFIELD: Thank you so much, Jon. And welcome, everyone. We are so excited to have you with us today. We have a really wonderful group of presenters to discuss how you can build on your performance and administrative data for research and evaluation. And we're going to have multiple opportunities for you to pose questions and participate. So we're so happy to have you here. Let's just get started. I'd love to introduce the presenters. You'll hear their voices in a little bit.

First off, we're going to have Wayne Gordon, who's the director of the Division of Research and Evaluation for ETA, Office of Policy Development and Research Division. Then we'll move to a very well-known person in our field, Dr. Randall Eberts from the Upjohn Institute, who will give us sort of a higher level view of what kind of idea can be used and how to apply different approaches.

And then we are going to finish up with Ms. Carrie Mayne, who's currently the chief economist with Utah System of Higher Education, but in very recent past was the grand leader of all Utah workforce research and evaluation. And she's going to give us some very concrete tips and advise on how to use data and how to get championships as well in your state to do the research and evaluation required under WIOA. So that's a little bit about who we'll be hearing from. Let me just run through the objectives today.

I think the main thing is we wanted to take a look at – let me give you some backstory for just a second. In case you haven't joined us in the past, this is the third of four public webinars that the Department of Labor Employment and Training Administration Division of Research and Evaluation is presenting on research and evaluation. It's all happening in the current program year. And the content that will be presented today is intended to enhance workforce development professionals at all levels of our system to better understand the right data to use that's available to support research and evaluation.

In the next 55 minutes or so, we're going to really take a look at how we define administrative data and how to find and use those data. And we'll look at the distinction between the goals of performance reporting versus the goals of research and evaluation. We'll also take a look at how these data may be used for rapid cycle, quasi-experimental, and other evaluation methods. And as they say on late night TV, and much, much more. So let us move to you, because this is all about you. We want to know what matters.

We'd love it if you could use that chat feature to please give us a sense of one thing you're hoping to gain from the session today or learn by the end of our time together. And by the way, we saw that some folks are noting already they're interested in using WIOA PIRL data for evaluation, so noted. We'll take a couple more minutes. Oh, great comment there from Christie about wanting to get a better sense of what is happening in the field.

And I'm assuming you mean research and evaluation that's happening, perhaps? Welcome, Ashley. We're so glad you're here. It's great to have fresh eyes with us. And I think that Ms. Mayne will be speaking today about very much some best practices, as well Dr. Eberts and possibly Mr. Gordon as well. Oh, great, great point. I think that was from Benjamin, saying he wants to make sure that we can help grantees by improving quality of programs through R&E. Great comment – what are some unique ways of using data to determine efficacy of programs? Fantastic. Oh, wonderful.

Oh, great. I love that – someone mentioned that they're new in their role, and we need to know what to measure and how to do so effectively. Fantastic. You guys are in the right space. Well done. I'm so glad you're here. Let me move us along, because we have the great good fortune to have Wayne Gordon with us today, who is going to set the framework for us. Wayne, do you want to take it away?

WAYNE GORDON: Yes. Thank you, Ingrid. I will, thank you. And thanks, all, for joining us today. Well, let's start with what the Workforce Innovation and Opportunity Act expects of us, or WIOA, as we call it. Federal and state expectations in WIOA around research and evaluation are generally the same. But for our purposes today, we'll focus on the state requirements under Section 116(e), that are highlighted on this slide. And I'll read those out loud for those just joining us on the phone.

I'll paraphrase this from the text of the regulations in law: "The State and local boards, and State agencies shall conduct evaluations in order to promote, establish, implement, and utilize methods for continuously improving core program activities in order to achieve high-level performance within, and high-level outcomes from the workforce development system." Later on, to paraphrase, "Results of evaluations…are to be used to promote the efficiency and effectiveness of the workforce development system."

This is very welcome news from WIOA. However, our interest here at the federal level – specifically within my division – has been in supporting the states in their endeavors to conduct research and evaluation. And that predates WIOA. We have worked with many volunteer states in federal research and evaluation activities. Over the years, we've conducted demonstrations and pilot studies with a parallel evaluation associated with those things. And during this time that I've been with ETA, we've learned a great deal, both with regard to the studies that we've done and also in the capacity of the states that we have worked with.

Some states, they had an established research and evaluation team that we worked with directly. And some states there wasn't. In those cases, we worked with key staff identified by state labor administrators. And all efforts really were successful. Some of those states that did not have a dedicated R&E staff, later established one. And we like to think our interactions with them at the time was some motivating factor in that. So we'll take as much credit as we can. But we're glad to see that states are moving toward using research and evaluation in a regular way, building that capacity.

More tangible effort over the years that we've undertaken was work with some states and various efforts to assist them in linking their administrative data across their state agency that served the same customer – be it workforce, education, or health and human services – and use that information to perform analysis and research either on their own or with a third-party evaluator. This effort of linking administrative data culminated into what many of you know as the WDQI, or Workforce Data Quality Initiatives, which we still award grants on today.

Some states, like Florida, have really taken off in using their administrative data in pursuit of answering research and evaluation questions. Other efforts, if I could briefly mention, are within the SWIS agreements. Even they make a nod to the extent that we are able to within the context of those agreements to support states in the use of administrative data for research and evaluation. We here at our Division of Research and Evaluation have undertaken a significant effort to provide technical assistance through the evaluation peer learning community, which we offer to assist states in building their evaluation capacity.

And we're now in our second cohort of six state teams in that endeavor. Further, ETA and NASWA have collectively been emphasizing state agency involvement in the Coleridge Initiatives, which provides a secure platform and training for state staff to analyze and generally play around with their own administrative data. Our work with NASWA goes back a long way if there's some discussion about state capacity. Several years back, we enlisted NASWA to do a scan of state research and evaluation capacity.

And we released a report – it's available on our website – which, at the time, was a snapshot of where the states were in that evolution from simply submitting required performance reports all the way through conducting research and evaluation within their states independently, either on their own or with a third-party evaluator; which takes us to my next slide, which is a quote of mine that my boss liked. So I'll continue to use it: "Research and evaluation begins where performance reporting ends."

The point here is to recognize that using administrative data for purposes other than managing your programs and simply performance reporting to the Feds is an important point to make. You already collect this data. And you own it. And then supplemented with other data that's both internal and nonpublic data and external data, increases its power and reach for states to answer questions and fulfill other evidence needs that are important to the states. Performance data allows you to detect the smoke and perhaps where to look for the fire, if you will.

Research and data analysis are needed prior to conducting evaluations to find the right questions to pose. So there's work there that can be done in order to develop well-formed and answerable questions. Each use of it beyond performance reporting and administrative data really reduces the per-unit cost of collecting it in the first place, and is an added value. Walk through here sort of a side-by-side comparison of performance measurement versus research and evaluation. And this information is taken from the Foundations for Evidence-based Policymaking Act of 2018, which is often referred to as the Evidence Act.

Just reading off from side to side, performance measurement can help you identify priority and questions and indicators. Think about this for yourself as I go through this list as to what else you think that performance measurement can provide and what, and in contrast, or in addition to what can evaluations provide?

Identify underperformance measurement, identify variations in performance over time across sites within your state. Perhaps it can suggest that many participants in a program experience a particular outcome. You might notice that. You might ask a question as to why. Any other ideas with performance measurement, please use the chat as the ideas come to you. In addition, for evaluations under the Evidence Act description, evaluations can clarify which indicators are predictive of an activity's success, provide context of potential explanations for variation over time, or across sites that were initially revealed by performance measurement.

Evaluations can confirm whether findings are directly attributable to the program or some aspect of that program. Again, please use chat for other ideas or other uses you see for evaluations. I'm seeing some really good ideas here. I'm glad everyone is throwing in their opinion into the mix. This is helpful. To get us started now, we'd like to take a quick pulse of your knowledge of and/or familiarity with administrative data. If you see the poll there for you, if you all could select an answer, multiple answers, we'll get a feel for where folks are coming from. And each person has a different perspective. We appreciate that.

It's important to know what all the perspectives are. Building capacity within states successfully around research and evaluation is to understand that everyone in your agency is coming from a different perspective, and often using different languages. So everyone's ability to understand that and become well versed in each other's language, then there's a translation opportunity there that, once exploited, everyone can get on the same page. All of the above is the clear winner. Is that a loaded question? Perhaps. Next slide, please. Here we are.

Here's a list of various approaches or techniques that one can undertake when conducting research and evaluation activities. Everyone has a different purpose. Everyone has a different time line. Everyone has a different cost and different uses of data. But all use administrative data. It is most often used in rigorous evaluations. This slide identifies the types of studies undertaken in an evaluation, be it experimental or quasi-experimental studies, that use sampling methods and sampling frames with statistical methods to test which impacts are determined to be impactful and statistically significant.

Rapid cycle and outcome studies may also use longitudinal and cross-sectional approaches that include administrative data. And lastly, cost studies may include cost analysis, cost effectiveness, and benefit cost and use administrative data for statistical or cost research methods. Administrative costs such as staff time, budget aspects are expected to be changed. Total costs and cost breakdowns of interest may also be used for evaluations. In this information, the cost information is an example of the internal administrative data you might have and you might want to bring to bear.

Cost and cost breakdowns of interest may also be used. Implementation studies are also used, administrative data such as participant enrollment, as well as demographic and other characteristics type of data.

Now, with this frame in mind, I hope I've painted the picture for you of where we're going. I would now like to turn the presentation over to Dr. Randall Eberts, who I've worked with in the past and can take it from here. Thank you, Dr. Eberts.

DR. RANDALL EBERTS: Well, thanks, Wayne. Thanks very much for that great introduction. I'd like to start my brief discussion today by going through what I consider to be the five key elements of an intelligent workforce system.

I can't go into much depth in this presentation, so I refer you to a chapter I wrote for a book put out by the Federal Reserve Bank of Atlanta in the Heldrich Center at Rutgers University called, "Transforming U.S. Workforce Development Policies of the 21st Century" – kind of a long title. The copies of this chapter are available on the Upjohn Institute website. I referenced it on page 23 along with some other references there.

So my first point here – and a very important reminder I think to all of us – is the job search is a complex decision process for both the jobseeker and the employer. Jobseekers in particular may not be familiar with the various steps they have to go through in order to find a job for the primary reason that many of them just don't look for a job. That's why an intelligent system must customize information and have it available when it is needed. But the data must be organized in a way that meets the needs of customers demanding.

When I go through this list, see how these points apply to what you as administrators or frontline staff of a workforce program need. So my first of these five bullets on this slide is decisions must be driven by data. And this means data put together in a longitudinal manner that reflects the history and demographic of each participant in the workforce program. Second, data must be customized to ensure customers see the relevance in the decision process and can easily access certain information when needed.

Third, decisions must be evidence-based, which means the data must be analyzed and placed in the proper form make good decisions; that is, good decisions made by both customers and management. The fourth bullet is one that says the management must take the information and use it to make useful decisions. And so for example, if the data is customized to individual customers, then services must be targeted to reflect the needs of those customers, whether it's the jobseeker or an employer.

And finally, performance management – that is, performance-related decisions – must be based on value-added measure; not simply based on performance measures. As you've probably noticed, the first three elements – and I think went too fast here – the first elements are, in my judgment, essential for understanding and constructing value-added measures that can be used by the management and are the focus of today's webinar. The last two help us understand how we should and can base our management decisions on workforce programs and base it on that information that's available.

So slide 14 – at this point, I'd like to start discussing the administrative data that are available at the state level, and then move to briefly describe how these data can be used in research and evaluation. First, I should note that data are generated by local workforce programs. These are transactional data that talk about how an individual works through a program. Are they enrolled? What date? When did they exit? And what services they provide.

But at the same time, that information also has elements about the demographics of individuals who participate in the program. And these are very rich. They include educational attainment, barriers to employment, veteran status, persons with disabilities, low-income status. And the list goes on and on. And when we merge these data to UI wage records, we have employment and earnings both before and after enrollment. In addition to the transactional data or the administrative data that's compiled by the workforce program, they also have information that can tell us something about the local labor market condition, the tightness of it today.

These include things like unemployment rate, job openings, maybe even mass layoff, employment characteristics – by this, I mean the percentage of people in certain industries or occupations – and job postings. With this information, we can construct the participant's employment history, as I've shown here on slide 15. This history shows us the things that have happened and the characteristics before they enter the program and after they exit from the program. As you can see, all of the elements I mentioned on the previous slide are incorporated here.

And if your state administrative records do not offer this sequence of information, then the PIRL – which stands for participant individual record layout, which is part of the WIOA way of collecting information, and it's constructed by the ETA – the PIRL does provide some of the sequences that are needed to put together longitudinal data. The information included on the last two slides comes from a variety of sources besides your own state-level management information system.

Except for private companies that scrub the internet for job postings like Burning Glass, all other data can be – and I should say – should be compiled from government sources. One reason that UI wage records and other data coming from official government sites are seen as objective is that they are, because they leave little room for manipulation by local administrators or others that could have access – (inaudible). Accountability is critical for both accountability in reporting and for research and evaluation.

So if you go back to Wayne's sliding up and down that list there, we have to have data that can be objective throughout this entire spectrum of information gathering and reporting, as well as evaluation. Now that we understand the data that we have available, this brings me to the distinction between performance and research and evaluation. For an economist like myself, the difference between performance that's typically used in WIOA workforce programs and for research and evaluation, it's the difference between what's called gross outcomes and net outcomes.

Now, the primary goal with workforce programs is to find someone a job. That's why performance is measured by labor market outcomes employment and earnings, let's say. Now, there are, of course, other measures – (inaudible) – and credential attainment, but keep it simple and focus only on employment. I should also mention that we define these two performance measures by quarter – employment second quarter after exit or employment fourth quarter after exit – something like that – since we use UI records which are collected quarterly.

For research and evaluation, we take the performance measure and adjust it usually statistically for key factors that figure into performance. So more specifically, research goals relate performance goals to key factors, and typically allow us to better understand what factors contribute to employment and by how much. In this way, questions can be asked like, does gender positively or negatively affect an individual's ability to find a job? Or is this relationship statistically significant?

An evaluation goal relate performance to the value-added of a workforce program. It is the full contribution of the services provided by workforce programs to an individual's ability to find a job. Simply put, our workforce service is helpful in that person finding a job. Now, for those of you who like equations – I'm sure that's most of you out there in the audience – I have written down an equation that explains the difference between performance goals, or gross outcome, and value-added, or net outcome.

It's performance goals minus administrative factors –- now, administrative factors include all those demographics and other things that I included in one of those slides – local labor market conditions, unemployment rates, and then unknown motivation factors. If you subtract all that from your performance goal – which is employment quarter after exit or fourth quarter after exit – you have what's called value-added. Therefore, from that equation, the difference between performance measures and value-added measures are the adjustment factors.

Let me say that again – the difference between performance measures and value-added measures are the adjustment factors. And that's administrative factors and local labor market conditions, those motivation factors. I should also mention that one of the big stumbling blocks in estimating value-added using this equation is the third factor – the unknown motivation factors. The problem is it's just not observable, but it can be proxied by other variables. But still, there's some part of that that isn't captured by those other variables.

Eliminated, however, by constructing appropriate comparison groups, such as like what pharmaceutical trials do with treatment and control looking at the difference between the two. We've heard a lot about that recently with the COVID vaccination. Now, this equation reminds us that performance goals can't – and I should say – cannot really be used to tell us about value-added; that is, the contribution of the workforce programs to the outcome. And it's because of these intervening factors that I've listed here.

Now, this I really think is one of the most important takeaways of my presentation. So you might want to just underscore that, if you can, in your mind or on a piece of paper. Without subtracting out key factors from performance measures – which value-added does – we don't know the true contribution of workforce programs to the likelihood of employment. Let me say that once again. Without subtracting out those key factors or performance measures – which value-added does, remember – we don't know the true contribution of workforce programs to the likelihood of employment.

Now, turning to the next slide, suppose that one local labor market investment area on a local WIB or something, they have more highly educated participants than another. And I'm looking at what key factors and how they relate to employment here. So let's say that we are comparing a diamond at the far right of the graph with one at the far left side of the graph, so one with more education versus a local workforce investment area that has less education.

And we found from our research that those with higher educational attainment are more likely to find employment. We have this upward sloping line there that's drawn there. OK? So there we see right away that there are intervening factors that affect what happened to the outcomes of this workforce investment area. Now let's also include something on the other set of factors. That would be the unemployment rate, the tightness of the labor market. So let's say that that same local workforce area – the one of the right with the higher educational attainment – may have a higher unemployment rate than, let's say, that local workforce area on the left, particularly in industries that hire more educated workers.

So this mitigates the likelihood of finding employment for the first local workforce area compared to the second one. You see we need to take into account as many factors as possible; not just education, but other factors have this type of relationship as well. All the variables that are listed on slide 14 we need to put in here and more. Another way to think about this is we need to counterfactual the same individual facing the same local labor market conditions – so without receiving services – in order to satisfy the research goal. Let me move to the next slide 20. Now, this slide shows several things.

First, it shows relationship between performance measures and value-added. The dot represents the predicted value with state workforce program. That is what we would expect when all the factors are accounted for. And the star represents the actual or performance value of employment. The length and distance between the star and the dot represents the extent the value added in the state. The longer that stick the greater the value added.

Now, this slide takes all those sticks that we saw in the previous slide – the difference between the actual and the predictive – and lines them up from highest to lowest. Or stated differently, it lines them up from greatest contribution of workforce programs to employment to the least contribution.

And these are actual numbers that were gleaned from the research we've done. Now that you're all experts on research and evaluation – just joking, don't worry – how could we use this information? We will focus on how to improve workforce programs and leave for another time the question about customized information for our customers to help with their decision. So here on this slide, I summarized some of the points already made in this presentation. And the first two bullets on this slide 22. So let me move to the third bullet.

Using value-added measures or evaluation goals, we can compare the effectiveness of states or local workforce investment boards in state by making these comparisons. We can learn from high performing programs to help inform low performing areas. And remember, this is one of the mandates that Wayne mentioned of the WIOA legislation. I should just put a plug in for what we're doing at the Upjohn Institute; and that is, we're engaged in something like this with the help of an outside foundation.

I'd like to invite any of you to learn more about this program if you like. The fourth bullet on this slide calls for an extension of culture – not a replacement of a culture accountability reporting, but an extension of that culture to one that also focuses on data and data analysis as the primary tool to identify, refine, test, and implement improvements in the workforce programs.

For those of you who have seen the movie, "Moneyball" starring Brad Pitt – does that ring a bell – or read the book by Michael Lewis, this is really the Moneyball phenomenon: using data to make better decisions. And to do this and to finish out this slide, we must do analysis that is research and evaluation often. And finally, we must find ways that local administrators can understand how their decisions now – that is, they make today – can affect employment outcomes in the future.

And finally, on this last slide of mine, I offer some light reading for your enjoyment – particularly before you go to bed – and illumination. So thanks for your kind attention. And now I have the pleasure of turning the program over to Ingrid.

MS. SCHONFIELD: Thank you so much, Dr. Eberts. And I'm glad to hear that we don't have to be experts just quite yet. And thank you for these studies. We have lots of questions coming in through the Q&A. And I think that folks may not be able to see other people's questions. I want to turn to one – there are so many good comments. I know we're running a little bit short on time. I feel like in the interest of time, I'm going to just turn over the mic to Ms. Carrie Mayne, who is going really speak to how the state of Utah has put these practices that Dr. Eberts is pointing to, to work.

And she'll give us some very specific examples of how that looked in the state of Utah. And then we'll turn to Q&A session after Ms. Mayne's presentation. So with that, I'm going to skip over this and welcome, Carrie Mayne. Thank you so much for being here. We really appreciate it. And we can't wait to hear about how you have really put these theories and concepts into action in Utah.

CARRIE MAYNE: Thank you, Ingrid. And thank you, everyone who has joined this webinar. I appreciate the opportunity. And my introduction stated I had several years at the Utah State Workforce Agency running the research and LMI programs. I'm now at the System of Higher Education. We are a WIOA partner. And we continue to do research and evaluation for education programs. I think you'll find similarities in both these arenas. I'm going to present to you all some background information that relates my presentation to everyone else's, and also to give you some examples that will show Dr. Eberts' framework in action with specific WIOA programs and also with education programs.

So discerning reports versus research – I think this is very important whether you are conducting the report or assembling the research. Or if you are someone who is going to be requesting this, it helps you to understand the magnitude of that which you request. And later, I'm going to talk about resources so you'll better understand what you'll need to prepare for should you be posing one of these evaluation questions to any of your partners who would conduct your research. So reports are valuable. They measure progress. They measure participation. And they often give us data that can help to track finance and budgeting.

It's an excellent way for us to measure activities and outcomes. So reports I often think of as displaying volume of activity, and outcomes relating to that activity. The reports are assembling data that tell you who, what, when, and where. But if you're noticing correlations or you're detecting patterns when you report this data across different demographic groups or across time, you perhaps will want to understand why – especially if you're in the position of making, changing, or informing policy, of which policies will be most effective – depend on the why behind the correlations in the patterns that you find.

Reports will indicate the correlation to you, but further research needs to be done to further understand causation. And Dr. Eberts explained very thoroughly why that would be the case. You may see perhaps a job training program resulting in an increase in employment over time, but you may also be experiencing a local economic boom. So the question is, was it the intervention of the job training program? Or is it simply a manifestation of the robust economy in which the seeker is looking for employment?

So with the research, we have to isolate those covariate factors, those other factors – as Dr. Eberts listed out – that could also be driving what you're seeing in reports. Isolate those in some manner, be it through experiment, quasi-experiment, or other approaches. And try to specifically measure that value-added that's coming from your specific training program. And that is where you gather the best information to inform policy. So think of it as building evidence in this pyramid type of data.

This is commonly used in science, but I think it applies here as well – it's understanding what your evidence is and what it can tell you. So expert opinion, if you will, is probably the most readily available information; information that people who have studied the field or studied the program for many years have understood to be common manifestations of policies or common effective policies, all the information that's gathered from prior research. But moving on up the pyramid, you get to better and more scientifically informed information resulting from whatever experiment or whatever evidence you're collecting.

So reports would reside in the next two levels of the pyramid – observational studies, perhaps you're displaying cross tables with comparison groups, or you have data from case studies or case reports at your fingertips. That's an increased level on our evidence pyramid. What we have been discussing today moves us more into the nonrandomized control trial. And in some instances, we have the resources available to actually run randomized control trials. The difference between the report and research is moving up to more scientific evidence that is better suited to inform whatever policy decisions you may be making as a result of what your evidence shows.

So the first example I want to explain to you all is research that I and a colleague conducted back in August 2011. We were asked a question similar to the example I gave earlier. We were seeing at our workforce agency improved outcomes with customers who were receiving job training programs. And I do use the term "job training" and not WIOA specifically for a couple of reasons. The Utah workforce agency encompasses more than just the WIOA programs. For those of you that have worked with Utah partners, you know it's a broader defined agency that includes things such as food stamps and TANFF.

So we were looking at all job training programs relating to those funding streams. Additionally, at the time we were conducting this research, WIOA was just beginning. So it was actually the WIA Program that were the workforce programs in our data in the research we conducted. So the report that our agency leadership was seeing was showing the volume of customers that were coming through and some success measures. We were seeing trends of improvement and wanted to understand what the interventions from our state were that were the reason for the success.

At the time, our local economy was also gaining significant momentum in the post-recession recovery. So that was clearly something that needed to be controlled for, if you will, as we understood and isolated specifically the effects of the training program. So we conducted an experiment that is more specifically a quasi-experiment – quasi-experiment meaning we did not assign a control and a treatment group prior to individuals entering a job training program. Rather, we were retroactively looking at populations of individuals who had and had not received the intervention.

They had not been chosen specifically for us, but we had data on the two groups. What is then done to control for all those other variables – as Dr. Eberts mentioned – was through a statistical model, find matches of individuals whose other variables that could have been affecting their labor market outcomes; things such as education, industry of employment, experience in the industry, length of employment, educational level, things of this nature. Those all matched statistically in a one-to-one ratio so that, essentially, if we have defined our variables correctly, the only difference between the individuals would be one received a job training program and one did not.

Then we would observe the individuals going forward past the training event and look at their wage and employment outcomes, assuming that through statistical controls and accurate measurements that what we would be measuring in that difference is the marginal effect that could be attributed to the intervention, i.e., the job training program. We brought in, as I said, not just WIOA and WIA data, but also Wagner-Peyser and other social assistance programs. We also data-source usage including wage records to show where the individual was employed, if they were employed, and in what industry.

And we also had for most of our data a significant amount of demographic data to use for controls as well. So the table here is an example of what we were able to see in terms of the four-year sum of differences in earnings and the four-year average employment differences, signifying to us that there were reasons statistically to believe that our job training programs were making a difference beyond that which just was being experienced from the economic recovery. And my example, I'm turning away from the workforce arena and into education.

I understand that most of you are in workforce, but I like this example because it shows you the same concepts that takes us outside of our usual mental framework of workforce programs. In this particular example at my current place of place of employment, the Utah System of Higher Education, we were producing reports to understand the performance of students in their first year of math; math being one of the difficulties that students experience in completing their degree.

We tracked on an annual basis the performance of students in that first math course taken in their freshman year of college. What we wanted to understand further is not just what courses they were choosing to take, what courses they were being placed in, and what their grades were in these courses, but why was it the case that some students performed well while others struggled. Did placement specifically play a role in this? Were students being placed in the wrong level of math, thus resulting in lower performance in those courses in their first year?

Gathering data on these students, which included not just their math course grades and their math course enrollment, but also data that we would consider measures of college readiness. So from the high school records of these students, we were looking at their ACT scores, their grade point averages in high school, additionally what they were experiencing in their first year of college – what course load they had, what other courses they were taking while they were taking their math course.

We then ran logistic regression with our regressions predict the probability of a dichotomous variable – so yes or no – the probability of a positive value. In this case, it was passing the math course. We were able then to control for demographic factors and understand whether GPA or ACT or neither or both were good indicators of future performance in the first math course in their freshman year. At the time, the composite of their GPA and their ACT had been used by the majority of our universities to place students in their first year math course. The results are shown here.

And where we see the coefficients – and this relates back to the equation that Dr. Eberts showed – we are measuring the effects of these individual variables. So you can see that age had a positive effect on their performance in this particular math 1010 course. Ethnicity where we had white and nonwhite as two groups had a negative impact on their performance in passing their math 1010 course. Cumulative GPA in the previous term from when they took the course had a positive indication, meaning that it does positively correlate to passing their math 1010 course. And ACT had a positive, but smaller probability for predicting that logistic regression for predicting the performance on that math 1010 course as their first course.

So these are two concrete examples, if you will, of what Dr. Eberts debut in terms of framework and methodology. How do you get from the idea of conducting research to on-the-ground work for completing research? In my prior experience at the state workforce agency, I understood this to be a very difficult task to surmount. So I reflected upon my experiences and tried to understand where the key elements were that led to successfully completing research agendas such as the ones I've presented today.

Here I will give you some of what I thought were the successful components. First of all, you need a well-supported prioritization. You will need leadership to want research. Research will not happen if it's not given a high priority and adequate resources. And part of that buy-in comes from helping your leadership understand the difference between research and reporting, and what research gets you that reporting can't.

So explaining to leadership the difference between research and reporting, explaining to them the ability to measure that value-added, and explaining that that value-added is really the information that will move and change, guide, and direct policy is where you will get the support that you need. You also need a project management plan. Research can't just happen because you point a finger at one or two researchers and say, go do the research. You have to have a deliberate project management plan.

You need to set clear milestones to progress towards that research goal and put resources toward the task of managing the project itself. So to simply send researchers off to conduct is not enough. Leadership must be put in place to manage the project, check on progress, set milestones, have clear goals, have clear due dates so that research does get completed and the result is actionable advice for your leadership. You also need a reasonable timetable. I can't express how many times I was asked a very profound research question and expected to answer it in a week or less.

You have to have a clear distinction in leadership's mind between reports and research. Reports can often be put together in a week, a day, a month. Research takes – and this is a simple estimate – 6-12 months or more. Short turnaround force researchers to create cross-tab reports. If you're asked to report something in a week, you're going to go back to correlation and you won't be able to explore and get closer to causation. You will not be able to do things such as logistic regressions or propensity score matching in a week timetable. It must be months dedicated to the project. Easier said than done, I'm sure you're all thinking.

Other components – probably the most important, but I put them later because I wanted you to focus on the first few – data. Most research questions will require data beyond program counts and basic demographics. You really have to have a data source beyond this one MIS system. You'll likely need to aggregate and link data together. It's not often the case that good, thorough research can happen from one program's data collection system. You'll want to have joined data in most cases. So you'll need data from cross systems.

Sometimes you need to put data sharing agreements in place depending on how your agencies are organized. Often it's the case that your FLDS or your WDQI resources in your state are a great place where this data is already being joined together. And for some states, there is dedicated FLDS or WDQI resources towards assembling data to be shared for research. And I know we already also have been talking about the PIRL, which is something I have since left the workforce agency so I'm not as familiar with. But I know people were asking, and Dr. Eberts touched on that system as a good source as well.

You're also going to need researchers. And how you're going to get those researchers depends on how your workforce agency is funded. In the case of Utah, the LMI Shop was funded and staffed in such a way that we were able to use the workforce information grant deliverables and the money attached to that grant to conduct some of the research for the WIOA. And that is very dependent on how the state is structured. So there's no guarantee that an LMI Shop could support at the level to which my example required. In other instances, you can use your program money, WIOA money, to hire consultants. I'm sure you have many partners. We have them in think tanks, research groups, and universities.

And depending on state of privacy issues that you have in your state, you may be able to hire graduate students. Sometimes FLDS system is set up so that they can de-identify data, and so it would be allowed for sharing for graduate students. I know this is very difficult at times, though. But sometimes that can be a good partnership with you and a local university. I hope this information was helpful for everyone. And we'll be fielding questions. And I'm going to turn it back to Ingrid now.

MS. SCHONFIELD: Thank you so much, Carrie. Boy, this is a master class today for sure. In the interest of time, I'm going to group some of the questions together. Thank you, everyone, for your great questions. Hopefully, the folks who asked how do we distinguish administrative from performance data, that has been answered. So I'd like to offer Wayne up for the question about, are WIOA partners working with other federal agencies to try to develop easier access to high value data sets for performance reporting, especially research and evaluation purposes? Having a mechanism like SWIS could be useful. Wayne, could you briefly speak to that?

MR. GORDON: Yes. Thanks, Ingrid. And I think some of this question has been touched upon. But both at the federal level, we're working with our partners under WIOA. At HHS, in education, and others, we've done it over many, many, years. It's been episodic. WIOA gives an opportunity to work more collaboratively and more purposefully. We're sharing discussions with our other federal agency partners now. And we would expect the states to do the same. So we're encouraging that reaching across the street, reaching across the building, whatever it may be at the state level – as we have done – to work with our partners that share the same customer.

Further, with the use of the data, the Evidence Act is pushing us toward transparency, toward availability, and making public more data and ease of access in that data, certainly at the federal level. And we hope this trickles down in the sharing of data. I mentioned the Coleridge Initiative. And I'm not well-versed. I haven't sat through any of their training, but I think some of the activity that participating states are doing there is combining data across agencies.

MS. SCHONFIELD: Great. Thank you so much, Wayne. And if folks are able to stay just a couple more minutes, we're going to try to get to a couple more groups of the questions that came in. I want to turn now to Ms. Carrie Mayne in Utah. There was a great question that came in early on about, how can local labor market conditions be examined – county level, ZIP code, city? If you could start with that one. And Dr. Eberts, if you want to chime in just briefly?

MS. MAYNE: In the research we have done, we got down to county level for some measures and WIB, a local workforce region definition for other data. Given that policies are used at the region level, it's often not the case that the ZIP code level data is going to – that level of granularity may over-fit your model, if you will. So the county level – which is what is commonly used for a lot of the labor market information – was sufficient for our research.

MS. SCHONFIELD: Perfect. Thank you.

DR. EBERTS: I agree. If I can just add this to what Carrie had to say. And that is that what you want is to match the area that people are looking for jobs with the data that you're collecting. So most people look for jobs within the workforce investment area if you're looking to evaluate workforce investment boards. And those are primarily designated by county, except for a few of them which are within a city or within a metropolitan area.

If you go to our website at the Upjohn Institute, you'll find a map of the United States which has all the workforce investment areas in it. Each one has an unemployment rate. And I think it's just the unemployment rate designated by the borders of that WIB. And so that might be helpful in answering the question.

MS. SCHONFIELD: Wonderful. Well, Dr. Eberts, since we're on you – and actually, before I go to that, let me just say that John's going to paste an email address into the chat for folks to use – I know we're running short on time, a little bit over time even – in case you want to share some of your questions directly with my technical assistance team and we can get answers to you. I want to turn Dr. Eberts while you have the mic to the question about, isn't program evaluation more about determining the merit value or worth of a program relative to a counterfactual versus simply adjusting a measure using regression?

DR. EBERTS: Right. No. I looked at that. And I think that was a very important question. I think we're saying the same thing actually. When we talk about merit value or worth of a program relative to a counterfactual, that's what we're doing even with the value-added that I was talking about. Because the value-added, if you're looking at WIBS within a state, the value-added would be the average characteristics of those factors within that state. So that would be your value-added.

You can also create counterfactuals by using propensity score matching or other ways of trying to find a similar group of people. The best way, of course, is to use random assignment. And that creates both a treatment and a control group, but there's ethical problems with that as well. So it's very difficult. It's never been done for a workforce program to my knowledge. But the last part is a little bit of your question kind of got me thinking. And that is you said something about except making value judgments. OK?

But what we're trying to do is not make value judgments, not subjective judgments, but objective judgments. And that's what I meant about objectivity with the UI wage records and other things. So it's not just coming up with somebody's subjective evaluation of a workforce program – it's really using hard core data to make that evaluation.

MS. SCHONFIELD: Great. That's a really important distinction. Thank you. There were quite a few questions that came in about how to pay for this work. I know that Carrie mentioned that. Carrie or Wayne, do you want to take another pass at that question? Any other suggestions about how to – (inaudible) –?

MR. GORDON: Yeah. This is Wayne. I can start with that, and if Carrie wants to jump in. Yes. We know paying for this is obviously a question that comes to mind first and foremost. WIOA and the Office of WIOA, of course, have pointed to the governor's set-aside funds. I've described it as that bottomless fountain of money that everything else at Capitol Hill – everything else has to be paid for. Capitol Hill has said it's there.

And depending on the state, we know the amount of money that you're applying the governor's set-aside percentage to is vast, and in some cases very, very small. However, the choice here is not to jump into the most expensive, the most time consuming endeavors. The use of administrative data to answer simple questions to begin to look one must crawl before they walk and walk before they run when it comes to undertaking different research and evaluation questions on.

So things can be done cheaply, especially if using existing data that you have or publicly available data. I know Carrie mentioned with regard to staffing, some states have engaged the state university and their Ph.D. candidates; for example, either through details or, in fact, an MOU or some other type of arrangement where, say, workforce agency can leverage the skill sets and knowledge that exist at the university that perhaps is just down the road. Carrie, any more to elaborate on that?

MS. MAYNE: No. I would agree with that. And I do also want to mention and reiterate the ability to, in some instances, leverage your LMI partners. Depending on the structure of the state, they can sometimes hire the research skills that would be needed to conduct these kind of studies.

MR. GORDON: Yes. I would agree with that. That's something I skipped over. And thank you. In my discussions with individuals that work at NASWA, they've made it very, very clear – specifically the LMI work group that they organize – they have conveyed to me that certainly the LMI Shop (seat us ?) is in their lane and in their skill set lane for the individuals that are involved in the LMI Shops. So they see themselves as definitely a resource that's available in house, and in most cases are just waiting to jump on this.

DR. EBERTS: Well, Wayne, if I could just add to that, that I really feel that the data itself can be an asset in the sense that there are some organizations that would like to able to use the data to do evaluations that are helpful to a state. We at the Upjohn Institute have worked closely with the state of Michigan on a variety of projects along with other states as well. And these include WPRS, as well as evaluating workforce program.

MR. GORDON: That's an important point – do a little horse trading with the data that you possess and others would love to use.

DR. EBERTS: Exactly.

MS. SCHONFIELD: Great. Well, thank you so much. This is Ingrid again. I just want to acknowledge that we are quite a bit over time now, although we do have about 350 people still in the room. If folks would be willing, Wayne, I'd love for you to drive. There were a group of questions that came in about how to work with workforce development system partners to get the evaluation done. And then we'll call it a day at that point. And we'll give folks a little bit of information about contacting, and let people go. Wayne?

MR. GORDON: Yes. I alluded to this before earlier in the Q&A. As I said then, we are, in fact, ourselves engaged with our federal partners. We've done so episodically in the past. And we're doing it more formally now. WIOA requires that we develop a five-year plan for research and evaluation activities. And in years past, that was required under WIA as well. And we did have input. We did get input. And we did survey our federal agency partners that share our customer with regard to research and evaluation activities that they would like to see done and that which they have planned themselves.

We're doing it more formally now under WIOA with the specification of partners. And we're encouraging states to do the same with their partners at the state level, their requisite similar partners at the state level. When it comes to looking out as to how states are meeting this requirement under 116, in sessions, webinars, and other forums before, I've described our view of it from a federal standpoint of engaging in this or meeting this requirement. We wanted to be as broad as possible. And that was in either conducting – their own states could conduct their own. Broad as possible, with recognizing efforts toward this goal of research and evaluation.

And that can include conducting in-house. States can conduct it themselves either in-house or with a third-party evaluator. Third-party would include the university resources that were discussed. It could also mean collaborating with your agency partners, with your within-state agency partners. They might have something. HHS or education might be coming down and looking to conduct a study. Or your agency partners may have an idea for a study. They may have resources for that. Your collaboration there is certainly a move toward engaging in research and evaluation that we believe satisfies 116.

And lastly, when the word goes out, our federal partners, we, the Department of Labor and HHS and education and others, when they put out the call for states that are interested in participating in any given pilot study or evaluation, you're willingness to raise your hand and participate and become part of the laboratory, if you will, that is also a way to meet these requirements, we believe. And lastly, you might be contacted by one of our contractors when we hire a third party to share data.

You're willingness to share that data with all of the requisite criteria for preserving privacy and security of that data, sharing that data is another way that a state can undertake, which we believe meets these requirements. So there's a variety of ways that the states can meet Section 116(e). And all of it is toward the goal of using data and using evidence for policy decision making.

MS. SCHONFIELD: Great. Thank you so much, Wayne. And thank you, everyone, who has been very patient to wait a little bit over with us today and for your great questions. I'm so sorry we weren't able to get to all of them. I want to point out – especially for folks who identify themselves as being newer to this world – we've got some great resources for you, particularly the newly revised state workforce agency evaluation tool kit, which is robust. And I think it's very user friendly.

Also, Dr. Eberts referenced earlier his piece that he wrote for the Urban Institute, which is a really helpful publication. And lastly, on WorkforceGPS, a community of practice exactly for this group, the evaluation research hub is constantly being updated with news, articles, profiles, reports on evaluation. So please check those out. And lastly, your fabulous Division of Research and Evaluation contacts are listed here.

And again, at the bottom, if you'd like to send any questions that we didn't address or you want answers to, you're welcome to send them to: evalplc@safalpartners.com. And we will share them with Gloria, Neil, and Wayne.

Thank you so much, everyone, for participating. Thank you so much, Dr. Eberts, Wayne, and Carrie – really fantastic. Jon, I'll turn it back over to you.

(END)