

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Panelists:

Barrett Johnson
Cynthia F. Parry
Leslie W. Zeitler

Please note: The following is a direct transcription and has not been edited.

Barry Johnson: Good afternoon everybody. I am Barry Johnson, and I am the Director of In-Service Training at California Social Work Education Center at UC Berkeley, School of Social Welfare. And I wanted to have our panelists introduce each other and then I will do my...

Cyndy Parry: Okay.

Barry Johnson: My duties.

Cindy Parry: And I am Cindy Parry. I am currently an Evaluation Consultant for CalSWEC. I have been now for quite some time, but I work in a number of different states and have worked with California since 2004 on this project.

Leslie Zeitler: My name is Leslie Zeitler. I am Training and Evaluation Specialist at CalSWEC and I have been there for several years. Oh, thank you. And prior to that I was doing direct services for children involved with child welfare, but from a nonprofit prospective.

Barry Johnson: Okay. So, I have this pink sheet that I am supposed to read to you with your rights and responsibilities. So I am a real follower at heart, so I will read it.

As a reminder the audio for this session will be digitally recorded and once formatted for accessibility standards will be made available through the Summit website in lieu of written consent. Participants who ask questions or provide comments during the session will be giving their “permission or consent” to this recording. Permission and consent are in quotes, I just want to point that out. If you have any questions about this recording, please feel free to talk to one of the summit support staff.

So the other thing that means is if we invite questions and I think we are fine with people unless we’d start running short on time people asking them as we go along, but, and we will hurry it along if that doesn’t work. But do try to speak into the microphone or we can try to repeat the question, because they are recording the sessions.

Alright, and to give you a little bit of context about CalSWEC and what we do. California is a county administered state oversight child welfare system, and it has quite strong counties meaning they have a lot of impact on program decisions. And we also have a very complex training system. There are five regions each with their own regional

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

training academy that delivers the training for the child welfare staff including the training we are talking about here, the Common Core Training and then CalSWEC is the coordinator of the training and the holder of the curriculum, so we updated and we coordinate and provide leadership for the evaluation of the Common Core Training. So that's a little bit of context about who we are before we start.

I am going to go up here so I can. So we come from the training world and we know that you are supposed to always tell people what your learning objectives are when you come into the training, at training sessions, so we've done that. A little idea what we're going to talk about, we're going to talk about our system that's very complex system that I was talking about and how we managed to get to having an evaluation that provides us with some meaningful results about our training. And talk about the different decisions that we, we came up with a set of decision points that really helped us to clarify what we were doing as we did this process and so that's about, that's the second learning objective. And then we will have probably just a question and answer session at the end where you can get a chance to think about where you are in terms of any project that you are working on in terms of training evaluation and where you might be in terms of these training evaluation decisions.

Alright, so why evaluate child welfare training? This is something we've asked ourselves having spent a lot of time doing this. I always, I think it comes back to why we are here, and that there is, you know, the increased focus on outcomes in our system. If you see me present before, you might have heard this story before but I will never forget our first child and family service review, the big meeting where the feds came and the big, the county leadership came and the state leadership came and I was there representing the training community and we got to the training part of the discussion and one of the current top administrator at a very large county in California raised his hand before I could say any, before anyone could say anything said, "Well we don't really know whether the training impact outcomes at all, so I don't know why we are spending so much time and resources on it".

So we unfortunately began this process before that meeting. So we were able to give a sort of an outline of what we were doing, but we've come, we've fully designed the evaluation process and I think that's kind of its key value I think, is to be able to not directly answer that question, but to be able to provide information about that in a sort of a meaningful and rigorous way. I think that's one of the big values that it has.

And you can see the bullets there. We do spend a lot of money on training. We don't know often what people know when they leave the training room. Yet we, most of our program improvement efforts and large initiatives, first thing they do is train our staff. And then we look at outcomes and we wonder why they weren't achieved. So we've sort of tried to tie that together with our training evaluation.

And the last bullet instead of what we've learned through this process, and I call it "The tail wags the dog", because when you start into designing and evaluation mechanism for your curriculum one thing you, usually you realize quickly is that your curriculum

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

probably isn't specific enough, so you end up going the evaluation people, end up talking to the tail, the evaluation people end up talking to the training people and saying hey, you know, so what is it that you really are training them. What do you, what are the specific knowledge and skills and how are you doing that. And then usually that cascades upward from the training people to the program people, because they realize, you know, we don't really know exactly what our administration wants in terms of training. And it's not the case of every training program obviously. But I found that focusing and evaluation really helps you focus, not only on what your training is but also on your program and what's really essentially and what you really want to know that your staff gets from it.

So I started to say this. This is just a list of our partners. You know, it's a complicated system when there is a lot of logos involved and there are -- and so we had our State Department of Social Services and our Accounting Welfare Directors Association are in the middle of that because they are the key folks that are administering the system, and then these are all of our different training entities around the edge including CalSWEC. And I want to emphasize that we could not do anything without all these people. They really administer, not only administer the evaluation instruments, but play integral part in designing them, and that's one of the reasons it's worked is that they have really been part of the process throughout it.

So we are in the evaluation Summits. So I don't need to preach this to you too much. But this is just a slide that gives it for practitioners and for administrators what sort of links evaluation of training to them. You know, we have training regulations, prior to doing this we have a level system and I call it Level-0 is tracking and training because prior to doing this we didn't track our training in California. You know, we did a lot of training and people got the training and we knew it, but we didn't systematically track who had complete the training and who didn't, and it made us, this process made us look at that much more carefully and make that more systematic. It also allows, as I guess it allows to participate more in the training process in the curriculum review process. Your administrators and program people if you do it right. And then they get more investment in what you are actually training and it becomes more of a reflection of what they want. And it gives them a structure to support the transfer of the training and the transfer of the learning into the practice environment, because they, you've clarified what it, you know, exactly the skills are and how you are measuring them that you want to get across.

Alright, so our process you can see it started a long time ago. But it did start before, it started just as a, before our first program improvement plan and our first child and family services review because we just wanted to do it. We knew it was important and then it really had a fire lit under it by the first pip process, where the pip included developing a framework for training evaluation and implementing it. The pip also included standardizing our training statewide. We have not done that for our Common Core Training. I would say about the CSFR process, that you can complaint about the methodology, but they sort of get it. They understood that behind all of our language about all of the great training we are providing them, we didn't have standardized training and they made us do it.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

So we started our framework and it's a very collaborative process with that macro evaluation team. The macro evaluation team includes county participants, people from that training academies and our consultants and people from CalSWEC. And we completed the framework in 2004. Notice there is a five year sort of ramp up to get the more of it, the most, not the most basic, to get sort of an intermediate level because we haven't really gotten all the way to top of our framework yet, it's sort of an ongoing process. But it was a lot, because we are also standardizing the curriculum during that time. So it did take a while. And then we decided at the end of that period, near the end of that period that we would do a second strategic and Leslie is going to talk a little bit about what that did to refine our plan. So, we now have a strategic planning process in place. We are about to go into our third process of strategic planning. And this is just a graphic display of our process.

Alright, and Cindy is going to come up and talk a little bit about those decision points that I mentioned.

Cindy Perry: Okay. As Barry mentioned, we started out with one of the first macro evaluation team meetings, laying out a series of decision points for the group to consider, to try to start a structure, the training evaluation program. And these are some of the things that we started our considering and again, preaching to the quire as evaluator as you all know you need to start out with what's the purpose for this evaluation, and that was a key point for people to think about, to discuss. You can see up on the slide, there were a number of things that we did consider there and a little bit later I'll talk about where we landed on that decision. But this would drive a lot of future decisions. So we really wanted to get this right. We wanted to get people clear on what they wanted to do with this process and now as we are approaching the third strategic planning we are kind of coming back and revisiting our initial thoughts about this. But what this did was this determined thing like test construction, test length, how we would go about covering the content from the classes and how we would collect information about who is taking these tests, what the ID code structure would look like. Whether we would be able to identify people as a result of the ID code or whether it would be more or less anonymous as far as we were concerned.

And then accountability, we've had a lot of discussions about accountability. As Barry mentioned it is a complex system. There are 58 counties. We still didn't come out completely uniform on the accountability question. Los Angeles has a little bit different approach to that. They have a little bit more accountability and they already built into their system and wanted to maintain that. So they are a little bit different yet in terms of what's done with the information. And again talk about more of that a little bit later.

The other thing we had to consider was how we were going to define success, whether we were going to be looking at for example change from pre to post or whether we were going to be looking at progress toward a standard. And again we came out with slightly different decisions depending on the course module for that one. So this was some of the initial around some of these purposes for evaluation, what we were going to be using the evaluation information for.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

We also wanted to talk about -- in connection with how we would be using the information, whether there would be any stakes for the participants. You know, everybody was interested at first in high stakes testing, was interested in certifying people possibly as confident in specific areas and that as we discussed what that meant, we began to layout what you would need to do in terms of designing the system if you wanted to have a high stakes decision, at the end of it, the, I know, are all of you familiar with the APA guidelines for test development when you are doing a standardized test. Well the American Psychological Association has a, and jointly with National Council on Measurement and Education and someone else I am blanking on at the moment has a book standards for test development if you are going to be using a test for high stakes decision making. The book is honestly that thick. So, it's not a light undertaking. There are also union issues; there were human subject's issues, because CalSWEC is located in the university. There was a lot of complexity to consider there and that's not ultimately where we ended up going. But we did consider issues of rigor, we did consider what we wanted to put into our test development process, and I will talk about that also in a few minutes.

We wanted to narrow down the scope of the evaluation. There is several possibilities that we considered. One being sort of the structure and functioning of the training system itself, another being course specific evaluation or series of courses or a content area. We also considered whether we would get down to the level of a specific learning objective focused on a knowledge, skill or ability. And what we ended up with, Barry spoke a little bit about some of the curriculum development that was going on at the same time that we first started this process. California did not have one standardized new worker core curriculum, and so we had, as we were discussing this whole scope issue we also had to consider it simultaneously with curriculum development pieces. So we were, we go into that the habit of talking about the core of the core, because we realized that we couldn't get to standardize curriculum in a way that would support evaluation in every single course. So we did make some focus decisions about where we were going to go with this.

We also needed to consider what level of evaluation we were most interested in or levels of evaluation. Again, some of you very familiar with Kirkpatrick, I mentioned it as the training evaluator, I am sure most of you are. We started with a Kirkpatrick kind of conception of level, but we did get a lot more refined and I will show you that in a few minutes as well. But we did need to consider what levels of information we both wanted to achieve to begin with and what the curriculum could support to begin with and that's an important point too. We may have wanted to get the outcomes, but when we didn't have standardized curriculum, and we didn't know how it was being delivered and we didn't know if it was being delivered with fidelity as we hear a lot about in these couple of days. We really couldn't begin to answer those questions yet. So we had to be thinking of those questions of level in terms of what's most appropriate for the curriculum, what it would support at that time, where do we want to move toward, what's realistic and what can be phase in over the next few years of this strategic plan process.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

We also spent some time talking about how we will be sharing results, who would get what and for what purposes and how often, and who would have access to what information since this was a statewide system of training evaluation. But it's a county based child welfare system and a regionally based kind of training academy system. People were very concerned about who would see their information and how that would be shared. So we had a lot of decisions to make about what would be appropriate and inappropriate uses for the data, who would see what, and this has evolved and we'll talk about that too when we get to the decision making piece of this. I think when we first started considering this people hadn't really gotten to the place where they fully understood what they would be getting into at this point.

And then the last set of questions, we talked about were resources. What did we have available in terms of previous evaluation for example, what was sort of already in place? What skills did people have? What could be pulled from the universities? What do we need to bring in from people like myself, from outside? What kinds of sort of infrastructure needs where there as Barry mentioned there was a beginning of this no way statewide to answer the question of who is getting what training and who has completed what requirements. So that requires sort of a learning management system discussion. So we had a number of sort of infrastructure pieces that we had to talk about as a group as well.

And here is where we sort of wound up. After considering the whole possible range of purposes for our evaluation data, what we came to was, where we felt was most appropriate to begin with was to use the data for program improvement, for course improvement, for refinement of the curriculum and we have done a lot of that over the past few years with this information we collect and the evaluation feeds back. We have meetings from time-to-time to go over areas where people are not performing particularly well in specific questions on the exams and people look at that and in relation to is it covered in the curriculum, is there a problem with the question. So we have a group of trainers, a group of evaluators a group of, from all the different training academies sitting go over these kinds of things and feed that information back. So it has, that has been its primary purpose.

As I mentioned it's a little different in Los Angeles. They had in effect already a training academy that had a test, and they had a system where they would report the results of that test back to the supervisors, when the workers went back and were assigned to their units, and they wanted to maintain that. They don't have consequences attached. You don't lose your job if you don't do well, but they wanted to maintain that ability to do that. So they continue to do that with this evaluation data, and we have a special waiver from the IRB to permit that.

Barry Johnson: I just wanted to add, that's what we meant by stakes is having any, we...

Cindy Parry: Personal consequences.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Barry Johnson: Yeah, personal consequences for us where high, at least medium stakes to high stakes. So, you know, in fact they, the IRB approval, they have to sign and say they are not going to use, Los Angeles kind of will not use it for those purposes and then they can link their data up, so they know who the people are.

Cindy Parry: Right now as we are moving toward this third cycle of strategic planning, people are beginning to want to get more diagnostic, which is another possible purpose with this information. So we are beginning to see certain patterns of knowledge, skills, abilities where people are not performing as well and they would like to be able to provide that information back to the counties and it, as it were currently structured we can't do that. The tests, right now are setup, because they are only for measuring program improvement. To be relatively brief the most I could bring out of the trainers was half an hour. So, there are 30 items for the multiple choices. We also have what we call embedded skills assessments. They are sort of a hybrid between a paper-pencil knowledge test and a true skill demonstration kind of in the classroom. But what it does is it's, the task is setup to mirror something that they would have to do on the job and to require them to demonstrate a little deeper understanding and ability to apply what they are learning to a situation. Two of them are in child maltreatment identification models where there are, they have been given a set of, in the fiscal abuse I think it's 13 possible flags to be looking for, in the sexual abuse it's a little bit more, it's 17.

Leslie Zeitler: About 20.

Cindy Parry: Oh, 20! I mean I have looked at that lately. And what we do is we give them a scenario, actually four scenarios and we have them go through the scenarios, analyze the scenarios for some, the presence or absence of some of these indicators, then pull that information together and say well what will your decision be in this case, would you say that this was substantiated or not and would you pursue it. So that's one, there is also one in the risk and safety area. California uses the structure decision making in safety and risk and in that one again I get a scenario that get the actual structure decision making risk assessment tool and they have to analyze the scenario for risk and then that get scored as well.

So we have that kind of assessment, we also have the, as I said 30 item multiple choice questions per module. And the modules that we focused on, that we agreed with the core of the core, the ones you see up there sort of risk and safety, engaging families in case planning and case management, human development, placement improvement and then child maltreatment identification.

Okay. We already talked about a little bit about the stakes. Even though we decided that what we were primarily interested it was program improvement. We wanted to do as careful a job of developing these measures whether they were the knowledge tests or the embedded evaluations as we possibly could. We wanted to have high quality, we wanted to have some validity to the decisions that we are making as a result of this.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

So we have a multistep process that we go through on. Starting usually with the expert group from the macro evaluation team, we get a group of volunteers who will sit down and try to develop testing materials, right questions, and then we go through multiple levels of item review starting with sort of an editorial review just to make sure that we are clear in the questions that we kept, we've done some item writing seminars for people to kind of bring them up to speed on some of the dos and don'ts and some of the things to avoid when they are writing questions. We go through the multiple levels of sort of editorial, of training, item writing, editorial review. We pilot each item. We can do some statistical item analysis and then we also look at differential item functioning because one of the major themes in California for their practice is fairness and equity and we want to model fairness and equity and this is, this process is well for the workers. We don't want any particular group whether based on race gender, we also look at whether you have English as second language and we try to make sure that we don't have items that perform differently for some subgroup of trainees, and if we do find those, they cycle back through that process. I spoke about where we take the group and we get the content experts, we get needs the representation of the testing experts and we get together. We discuss and hash it out and try to identify what's going wrong with it and modify it.

What we have done, this is the classic case and probably all of you have been in this situation. We've been working on evaluation of training and evaluation in your states. It's a classic case of you are building the plane, well we are flying the plane, and we needed to be developing this system in concert with standardizing and refining curriculum, getting that in place across the state. And we needed to have usable data while we are doing this. We didn't want to have to wait 10 years before we get start to look at data. So what we've been doing is using Rasch, an item response theory model for doing the test developments that, item scaling. Are you already familiar with, familiar with the IRT models? I talk about that for a minute.

The idea behind this is usually when you've got a knowledge test for example, you get the score you get someone score by adding a boulder correct responses. And what happens to you is if the curriculum changes and you change a few items, you've got a whole new test that doesn't line up with your previous test. So you have no way of equating performance on last year's test with this year's test. It's also a pretty complex process if you want to use a different pretest and a different posttest because they have to be equivalent in difficulty. What this approach does is it helps you build the bank of items and this whole bank of items is placed on a common scale and the way that you get a score for somebody is you look at not how many of the items on the test they got right, but their probability of success on items of a certain difficulty level. So it doesn't matter which set of items they take necessarily because its role hinges on your probability of success on an item of a specific difficulty. So what we can do is we can change out items on a test and overlap the old form with the new form and we can still make the same kinds of inferences about people's progress, change from pre to post because it's all on a common scale. So as the curriculum changes we don't throw out the whole test and start over. So what we've been doing is we've been gradually building a bank of items that we can use to pull different pretests and posttests from and to look, and we can look at trends across years even though the curriculum may have changed, even though people

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

may have taken slightly different test versions. So that's been enormously helpful. Did that make sense to everybody?

Alright, I spoke about Kirkpatrick a few minutes ago and we did kind of expand on this and not just for the sake of complicating everyone's life, sometimes people accuse me of that. What we wanted to do here when we were talking about our levels of evaluation was to make them correspond more closely to the level of the learning objective, the level of the curriculum because then you get a closer correspondence between how you are measuring success and how you are teaching the material. One of the things that sometimes happens and I see happen a lot as I work with training curricula is that you'll have objective say at a skill level, but the training just gives knowledge, just talks about say interviewing but doesn't actually have people practice interviewing. The test may go back to the knowledge level and just be, so what do you know about what an open ended question looks like. So we wanted to make sure that we weren't doing something where we had a mismatch between how we were measuring the learning objective and the way it was being taught in the sort of desired end result.

So in doing this Jane Berdie and I, when we were at American Humane put together a system of levels that was -- took Kirkpatrick to sort of cross walked it with Bloom's Taxonomy. If anybody is an ex-teacher you might be familiar with Bloom's Taxonomy, but it's sort of levels of questioning tied to levels of learning. So when we brought this to California, Barry's big contribution is he added the tracking level so we now we have a new level and we collapsed a few at the end. But these were the levels that we came up with and what the idea behind this was to make it very painfully clear what the desired end result was of going through the curriculum and going through the learning and how we wanted to measure it. So you can see up there what we have done is we've split out something's typically in Kirkpatrick knowledge and skills lumped together. I think that can cause some issues, so we've split that out. We've also split skill into what we call skill acquisition, which is demonstrated in class versus skill, versus real transfer of learning which is taking the skill and demonstrating it on the job. With the recognition that that's, well once we get past Level-5 a lot of stuff is out of our direct control and this is where we are beginning to have a lot of competing explanations for what we see.

So what we've done is when we do the strategic planning we built our strategic plans around trying to collect information at these various levels, but without the assumption that we do all Level-2 before we get to all, before we go to Level-3 or all Level-3 before we go to Level-4. What we try to do is tie the level of evaluation as I said to the appropriate level of the learning objective and the curriculum itself.

We are also not going to be doing everything in one strategic plan. We have been working on this as, looking at that timeline now for 10 years. We are, we've done quite a bit at our Level-4 which is the knowledge acquisition with our pre and post testing and we've done some sort of, I don't know, we call it four and half with our embedded evaluations with the child maltreatment and notification and the safety and risk where we are getting towards more, something that mirrors more of an actual skill. We are just beginning now to explore some possible options for looking at a transfer of learning

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

studies. So we are kind of got that in process and we are hoping to be able to move into that in the next strategic plan.

One of the things that we've done, I know I heard a lot of speaking this morning, you probably have to, about some of the difficulties with. When you are doing a program evaluation and you are looking at outcomes with attributing outcomes to your intervention same thing in training, I mean we would love to believe that you know if children are safer and we unified more quickly that is because we did a great job of training. We don't have the option of doing any kind of randomized control. We can't randomly assign a brand new worker to get trained or not trained. We can't even do a rolling groups kind of thing and have them hold off for a while and go out and work without being trained. So there are a lot of options. So what we have been doing is we have been working on a kind of a concept that we've been calling the "Chain of Evidence". I just was doing some review for the paper I found that there is someone named Mane who has been doing it as well and he has been calling it "Contribution Analysis". It's a concept that's been around in various forms and the idea is that you establish a linkage between the training and the outcomes, kind of step-by-step so that a reasonable person would agree that you probably had something to do with it. We can't exactly pin down how much we might have had to do with it but we can pretty well demonstrate the training had some contributions here, especially in a situation where we were doing a pretest before a three day module and a posttest at the end of the three day module. We are pretty convinced that, you know, since you went home and went to the hotel and between nothing much happened besides our training.

But the idea here is that if somebody, if the training is really closely linked to what's important on the job, if what we've done is measured someone's sort of engagement with the training through our workshop evaluation forms at the end of the workshop, we see that they see relevance to it that they are interested in it. They feel they have learned from it. They demonstrate that on our measures that they have learned from it. Then we can say that we probably, if we went out and we are able to look at as we are hoping to find out soon, we look at transfer of training that we had an effect on that subsequent behavior. So that's the idea behind this Chain of Evidence concept and we are trying to as we go through our different levels and we tried to structure these strategic plans to start building and filling in some of these pieces, okay.

Barry Johnson: That also helps you answer your questions from administrators. When they question whether you have any, and you can make any linked outcomes we actually have coined a term for it and have a theoretical base behind it which is actually important to be able to talk about.

Cindy Parry: Okay, Leslie?

Leslie Zeitler: I think this one is a super...

Barry Johnson: Yeah it is.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Cindy Parry: Okay, I am sorry.

Barry Johnson: Well you can say that because, well are you coming back?

Cindy Parry: No.

Barry Johnson: Okay.

Cindy Parry: Absolutely.

Barry Johnson: Alright. Levels of Standardization, this was important and it is one of the advantages of doing the curriculum and the evaluation plan at the same time. There are a lot of disadvantages in doing that because things were changing all the time, but the advantage is you are able to change things all the time and you are able to make decisions by the evaluation and then it's not too late. You haven't already created a curriculum that doesn't fit your evaluation methodology. So, this is the level of standardization has a lot to do with that. In fact we determine the level of standardization a lot by how we plan to evaluate the curriculum. So we have multiple levels of standardization in California because it's such a complex system. So there are, it's a 14 I think areas of the common core. We had little names to all and I think we call this the "Gang of 14" right. Yeah, I think there are 14 areas that have common learning objectives.

Cindy Parry: Oh, yeah.

Barry Johnson: So the different regions and counties, because counties also administer their own training. They are responsible to meet those learning objectives. And our learning objective really benefited from our training evaluation work because they are much more sort of rigorous and measurable, so it's more obvious what exactly needs to be in the training from the learning objectives. So that's one level of standardization, that level is at the areas it's not the core of the core, it's, we are not doing a lot of evaluation. There we are doing some, you know, locally they are doing satisfaction evaluation and they are tracking that training. So we are really not getting too far into our levels of the training evaluation at that level and we are doing sort of formative evaluation were people are providing ongoing feedback for updating those systematically overtime.

And there are six core areas that Cindy called the "core and core" that was our little name for that where information is standard. And this is something that we learned is that, you know, often you don't know what exactly the information is in a curriculum to get across in a classroom. So this really established that and made a really centralized structure. I would put in a little plug, I didn't know if any of you have been to our CalSWEC website, but all of our curriculum is online and available and free and we invite you to use it. We also have our knowledge items are behind a password, but we will share them with you if you request it. So if you go to our website and look at our common core, you are more than welcome. In fact we really want people to use it as widely as possible.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

So there are the six core areas where information is standard and on this, this slide needs to be updated actually because there are not two core areas, three topics and two core areas where the delivery and information is standard, okay. So in the six areas that we have standard content we have knowledge testing. So you obviously have to define what the right answers are in your training curriculum if you are going to test people for knowledge. So that's, those are the knowledge test that Cindy was talking about that we did the item development process for the rigorous item development process for.

And then the child maltreatment identification is the embedded evaluations that Cindy talked about for -- child maltreatment identification, part-1 is physical abuse. It has, the curriculum also covers emotional abuse and neglect and Leslie will probably talk about. We know that we are at most cases in California and around the country is non-neglect. So we kind of wish we can do embedded evaluation or evaluation with more rigor there. We had a hard time because in our complex system there is different community standards for neglect and it's very difficult to get 58 counties doing, bringing about what neglect is even though there is common law about it in kind of regulation patterns. So there is, so we test on physical abuse for our embedded evaluations with the scenarios and people filling out sheets in the classroom.

Leslie Zeitler: Sexual abuse is well identified...

Barry Johnson: And sexual abuse is the other one. And then we have the new risk and safety assessment that has them actually look at the risk, the safety assessments tool to the Structured Decision Making tool. And are you going to talk about how that sort of works or should I talk about that now?

Leslie Zeitler: You can talk about it now.

Barry Johnson: So the embedded evaluations with the scenarios, one key thing is about them is that we debrief them in class but we collect the data prior to debriefing them. So there is four scenarios. Two are abuse, two are not abuse in the child mental treatment identification trainings and they determine whether they are going to substantiate or not. So it's a right or wrong answer. There is also right or wrong answers based on our sort of expert opinions about the different elements that they are supposed to consider. Then we have a good old fashioned carbon paper in that, the test sheets are on and they fill those out and they rip it off. They give us one that has their code on it that's linked to their demographic and they send it into, centrally to CalSWEC for us to analyze that data. But prior to leaving the classroom the trainer debriefs the right answers on that. So we know that while we are getting the evaluation data about how well the trainers, training is working we are also, they are not leaving without getting what the right answer is. So that's a way to sort of balance the, balance, getting some evaluation data that's meaningful, but also assuring quality.

Leslie Zeitler: Can I add one more...

Barry Johnson: Sure.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Leslie Zeitler: Comment about that, so I will say that one of the challenges in having the embedded evaluations is it does take a chunk of time during the training day and we have gotten push back from our trainers who have said, you know, this is a lot of time for an evaluation and what we've said is when you debrief the answers you are reinforcing the training and it really is because I've also sat in on those courses and you can tell that something is happening with the trainees are saying oh, okay I forgot that or I didn't think about that and the trainer is going through all of the correct answers, so that's really been a very useful piece of the embedded evaluation as a, both an evaluation tool and a training reinforcement tool.

Barry Johnson: And I would also say that it's the purpose of our evaluation is improving the training. That has given us the richest amount of information about changing the curriculum and updating the curriculum of what's working. I mean you'd be amazed at how many errors you have in your, I mean we go through our curriculum and our evaluation so rigorously and so I guarantee you that you have this going on in your curriculum out there and we will get it and we can't figure out why it's networking and we will figure out oh, you know, we were using the wrong answer for the key or oh, we didn't teach that. You know, yeah...

Leslie Zeitler: One of my old time favorite, the trainer didn't know how to do the exercise so they skipped it.

Barry Johnson: Yes, exactly. We, so there is all kinds of interesting information you find when you start looking at these results overtime. We will talk more about that.

Leslie Zeitler: Yeah.

Barry Johnson: This is just a graphic display of our different levels of standardization in California. And disseminations results, briefly we realized that we have multiple audiences for our results and we really wanted the information to be used by people, and we wanted to take, you know, for a number of reasons. Our trainers putting a lot of time for evaluation now that they are, it takes up time in their training classroom and we wanted to be able to show them results that we are meaningful to them and we wanted to show our academy partners how to look at their results and evaluate how they might improve their training or how they might diagnose what might have gone wrong or gone right in that training and also we wanted administrators to see what's, what the results for our training evaluation. I mean it's amazing by putting some graphs in front of people how much more satisfied they are about your training and about the money they are spending on it. I think that's a big impact like I said before.

So we have multiple reports. We have reports for the region that we issue separately. We have aggregate reports for the state that we give to the state Department of Social Services in CalSWEC reviews and then we go over the results carefully. Every six months we have recycle and we go over the results with them and try to figure out what's

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

going on in the curriculum and we use then in an ongoing basis to revise our curriculum. And we issue reports which are also on our website that for these different audiences.

Word about resources, you can see we spent a lot of time and effort on this, so it's not a small commitment. Some of that time and resources is because we are California to be honest about that if you are in a simpler less complex system it's probably easier to do. I am also amazed I go places with smaller systems and I talk to them and give them, you know, technical assistance and then I care about things that we are planning and trying to do and then I talk to them a year later and they have done it and we haven't managed it yet, because we have a very politicized and a complex system. But you can see the different areas that, you know, we have, there is trainer time, there is curriculum writing time and curriculum rigor, there is our staff time, our consultant time for analyzing and designing our evaluations for helping us to design those evaluations. We have a data, a specialist centrally at our office who usually is a grad student and they were overwhelmed with data. And so we have a data specialist now whose job is to get all the data in and clean it and get into our consultants into our specialists so that we can analyze it. And we have a number of, you know, sort of equipment purchases. We use unique identifier codes on them and people write them in pen, so we have a scan system that allows, that reads people's writing not well always, we are just kind of the case.

Cindy Parry: Still an art.

Barry Johnson: Yes.

Cindy Parry: Collecting a hand written information.

Barry Johnson: But we do have, but you do need to invest in that, in that kind of stuff too. Okay, alright.

Cindy Parry: So I get to talk to you a little bit about the fruits of the labors that we have just been describing to you. So at this point we've been collecting demographic data since 2005 and we have over 5,000 new child welfare social workers and over 663 supervisors, child welfare supervisors who've participated in our common core process. And this data is fresh. We just through June 30th, okay. And we continue to capture demographic data, it's an ongoing process. At our Level-2 again that's our formative evaluations for all of our standardized curricula, I mean, that's for standardization at the content level and at delivery. We have revised those curricula and it's based on a process where we have someone go and observe or more than one or two people observe from the different regions, from a statewide prospective. They give feedback. We've also gotten feedback from trainees. If we have a little bit of time at the end of the training day we will say, you know what, this curriculum is going to go under revision and if you like to give some feedback let's just talk about what worked, what do you think would be different, what would you like to have in this curriculum as a brand new worker, what would be more helpful and that's been very fruitful to get that feedback.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

At Level-3, that's the satisfaction level and we do not collect that at a statewide level. We leave that to the regions and the counties. So Level-4, that's our knowledge tests and so there are three topics for which we collect both pre and posttest level data. That's our child and youth development course, our family engagement, case planning and case management and our permanency and placement curricula. And basically trainees and these are brand new child welfare workers have improved significantly from pre to post during these training courses over the past several years. We are finding that as a pretty consistent statistic.

Okay. Now we also as part of our collection of demographic data we look at who are our 4E trainees are. We've gone through the MSW 4E preparation program and, so we found that, oh yeah, go ahead.

Yes, so that is a stipend program and I don't know if it's in all states, but it's in several states where if you are interested in getting an MSW you can apply for this stipend program, but it requires that, if they give you the stipend I think it's 18.5 each of, for each of two years at this point that you owe two years of time to the Department of Social Services in your state or your county social services in your state.

Barry Johnson: The amounts vary from state-to-state, but that's the California amount.

Cindy Parry: Yeah. Yes, go ahead.

Questioner-1: Yes. I just wanted to remind everyone just for you are reporting.

Cindy Parry: Yes.

Questioner-1: Can you please ask the questions if you like, just a friendly reminder.

Cindy Parry: Yeah, thank you. Yeah. And so what we are doing, and to come back to the demographic data as seeing if there is any differences between MSW 4E stipend program participants and those who have not participated in the 4E program.

Barry Johnson: So our 4E program I guess it will add some interest to it. Our 4E program is very large. It's also had Cal, administered by CalSWEC. We have 21 schools in social work that participate providing stipends to students. It also provides support for specialized education in those schools, competency based education in the schools. The master contract titled 4E contract is the largest contract at UC Berkley and it's about \$32 million. So people are interested in knowing whether they come out knowing more or not.

Cindy Parry: And the other thing that we probably should mention too when is that the learning objectives that the curriculum is geared toward and that the testing geared toward and the same learning objectives that are developed for the 4E program. So there is a close connection between the learning objectives for the professional education that happens in the 4E program.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Leslie Zeitler: Yeah.

Cindy Parry: The in-service training and then the testing that Leslie is giving you the thoughts about.

Barry Johnson: And the other thing that we have always heard since we had both the training program and 4E education program is bitter, bitter complaints from the 4E students about how they don't need the training, because they just went to graduate school and got all this training and now we have something to say to them besides be quiet.

Cindy Parry: And these results tend to be consistent as well. As that the 4E trainees tend to come in higher at pretest, they tend to leave higher at posttest and the posttest score differences have been statistically significant for all of our models over the past two fiscal years, and for two out of the three models that are pre and post testing since January of '07. And trainees, the 4E trainees, they tend to achieve significant gains from pre to posttest, but that's not happened every single testing cycle. Has that happened every single testing cycle?

Leslie Zeitler: Yeah. I think they have.

Cindy Parry: I am thinking of the second bullet.

Leslie Zeitler: Yeah. You are thinking of the second one.

Cindy Parry: I am thinking of second one.

Leslie Zeitler: What it, what basically the pattern is as they come in higher they need higher, but they also achieve significant gains in between and the gains that they achieve don't differ significantly from the gains achieved by other people with other educational backgrounds. So that's what Barry was saying when he was saying now we have something to say to them because...

Cindy Parry: Yes.

Leslie Zeitler: It's like, yes, you do need to be here.

Cindy Parry: [Overlapping conversation] [00:52:35] something.

Barry Johnson: Right. Well, and the other piece of that is that's a very, that was a very big relief because it makes, it validates both of our programs because it says that they are getting something, they are coming in higher so they are getting something out of the program that out of their educational program you are going for, that's bringing them in a higher level but they are still benefiting from the training. So we are pleased.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Cindy Parry: Yes.

Leslie Zeitler: Okay. Okay, still at the knowledge test level, so there is one topic that we only assess at the end of training, posttest only training which is our critical thinking in child welfare assessment and that's assessing for safety risk and protective capacity. It's a one day training. And the reason that we only do the posttest is because it would probably take a little bit too much time out of the training day to do a pre and a posttest if you've got a six hour training. So what we have found though in looking just at the posttest results is that even though there is no formal standard that's been established it's serves you as a yardstick with a mastery. Our data shows that trainees leave with a substantial level of knowledge related to our learning objectives. Did you want to add anything Cindy?

Cindy Parry: Yeah. I can help, but I am sorry. This one, this class drives me crazy as an evaluator because I don't have the yard stick. I don't have a criterion established here to say yes, they learned, yes they learned a lot, did they learn enough, how much did they know coming in versus what they know going out. This has been sort of an issue for the group that came up as along the way I thought us endive evaluator will be a simple thing to sit down and talk to the group about okay, we need to a standard. It was politically a hot potato because people had a really problem with establishing a standard because someone might fail and what can you do with someone who fails when we have a system of id's that people, that are self-generated. We don't identify people. They self-identify and we link the pieces using that id, but we don't know who they are. So if someone could theoretically fail this and we wouldn't know who they were and that has been a real issue. And that's why as I kind of quickly said before we are coming back around to some additional discussion about what we want to do with these results and I think it's a kind of thing that evolves and when you first make this decision before you really get into it and before you get to the point where you are struggling with some of these messy issues, you may make one decision and then later on you may have to come back and cycle around and revisit that and see you are still comfortable with that and that's kind of where we are at right now.

Barry Johnson: And you know it really is a, it's a liability issue especially since we publicly report this and there has been more legal action that talks about the training. You know they, it's been involved in legal cases recently in California and you sort of are forced to acknowledge once you set the standard that you are sending a proportion of people out who don't meet your standard and you are not doing anything about it except improving the curriculum next time. So that's some of the policies behind that, yeah.

Leslie Zeitler: Okay. So Level-5, that's again where we assess for skill in the classroom, the embedded evaluations and that's when Barry is talking a little bit earlier about child mal treatment and identification parts one and two. There are four scenarios for each of those, where there is two are abuse is the correct answer and two are not abuse. And so when they talk about at least 87% and then most of the years 90% or more of new child welfare workers make three out of four correct decisions and whether or not abuse occurred and that's a pretty good number, we are able to make the correct decision. And

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

again for this one we've also struggled with not having a passing score for the same reasons that Cindy had talked about earlier, so that's why we reported out as making three out of four decisions correct given the test design.

Okay, Level-6. This is transfer of learning in the field. We did complete some regional studies on transfer in field training with a, for a couple of our regions and if you want more information about that you can look at our 2009 whitepaper which is on our website and I think it's noted, the link is noted at the end of the PowerPoint.

Level-7 which is the outcomes level, again we are focusing on developing our building blocks to get to or develop the chain of evidence to get to outcomes. And so that's why we continue to do and looking our strategic planning process. And in the next strategic plan we will be looking at kind of putting in more parts of the building blocks that go from transfer of learning to outcomes, I am trying to figure out how to do that in a rigorous kind of way.

Okay, so did most of you get the handout, there is handout on the chair.

Barry Johnson: Anyone not given?

Leslie Zeitler: I think most of them...

Cindy Parry: I don't think people would reach handout and...

Leslie Zeitler: This is the handout. It looks like -- this has got our logo on top left hand corner. It's our statewide report for trainers and county administrators, analysis of common core data of June, 2011.

Okay, so I am briefly going to go over this. This is an example of one of our reports that Barry had talked about before we have a variety of constituents who want to know what is happening in training. And, so I am going to go over a little bit about this report. So first, this one is for our trainers and county administrators. We do have another set of reports that goes to our board to our academies that are a little bit more descriptive and we wanted to be very mindful in terms of what we report out that it not get misinterpreted. We want to be as clear as possible in what we share and how we share it. So the structure of this particular report is we give a little bit of background, that's the first page. A little further down, it's a brief overview of the results for a six month period and in the box at the very bottom that is if nobody looks at anything else except this you will get the crux of what happened during that six month period. And so for this, this is where we talk about knowledge from pre to posttest increasing at a statistically significant level and talking also not just about knowledge test, but our embedded evaluations.

If you go to the next page, page-2, we talk a little bit about how the evaluation findings are used including for instance for curriculum improvement and also to help improve the testing process and then you move down and we look at results for select common core

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

topics. So we've got family engagement and case planning and case management here. You've got the graph on the left and the textbox on the right in terms of identifying how many participants and whether or not the learning of gains were statistically significant. And if you go to the next page we cover child and youth development, the permanency and placement. So these are all the ones that have the pre and posttests. If you move on...

Barry Johnson: And at the graphs when we show, when we issue reports to our training partners they can look at their results and compare it to the statewide results. So it allows them to look and see where we didn't know as well and why in a particular area.

Leslie Zeitler: Right. If you go to page-4, then we start looking at the results for the child mental treatment identification, parts one and two embedded evaluation. So you have the graph on the left is we do say three out of four pass fail there. And then we do look at the right hand side graph is looking at the percentage of correct responses for abuse versus non-abuse scenarios. So they are getting both abuse scenarios where they are getting both non-abuse scenarios right and they would do that for both CMI-1 and CMI-2.

And then we look at some select demographic information, because where both the statewide reps and our regional reps are interested in who is coming through training. We will look at the age of our trainees, whether or not English is a second language, race and ethnicity, whether well time worked in child welfare part of core position, their education level, whether or not they participate in the stipend program. And then at the very end we do a little bit of a discussion of the relationship between MSW Title 4E preparation and their test scores and the differences possibly between 4E prepared trainees and those who are not 4E prepared. Any questions about that?

Cindy Parry: I just want to add...

Leslie Zeitler: Go ahead.

Cindy Parry: The version of this report that goes to our macro evaluation team, just sort of our advisory and decision making body that the training administrators from the various regions has a little more elaborate structure to it and we do give, do a multiple regression analysis and that one to look at the influence of some of these demographics that are just presented descriptively here on the test scores. So that's how we can kind of monitor whether all different trainees are learning approximately this is the same amount of whether we have some issues with differential response that we want to delve into and explore. For example, in a few models here and there we have found differences based on race occasionally on age so that's what triggered the differential item functioning analysis to try to weed out any items that were problematic for a particular group.

Leslie Zeitler: Yeah.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Barry Johnson: And you will notice, well also keep in mind that the demographics. Well they seem like very basic demographics, this is the only sort of statewide demographic thing that we have really. We do a workforce study about every five years that's a point in time look, but this is, this shows what the workforce in the different regions is like over the new workforce over a period of time and its new information, it's not information that is collected by the state otherwise. So it's useful in that way and if you look at the race and ethnicity you can see that there are differences by region that are quite remark, quite marked differences pronounced yes and there is also differences in educational level from region which allows people to sort of interpret the results differently.

Cindy Parry: I would say also another point about the reports that go to our oversight community, the macro evolve team. Bless you, is some other reports for instance for the child maltreatment identification ones, where we look at the individual elements in how the percentage of trainees who've got certain elements correct, the red flags if you will. That's helped us figure out, that's part of where we figure out did they really get this element, did we teach that, cover that well enough in the curriculum, was this because there was a point in time where the instructions were not clear enough, because there was some consistency across regions that people were getting certain things wrong. So it's been really helping us to refine not just the testing instrument in the way we are designing the test, but also it's helping us to go back to the curriculum and say okay, we really need to be beef this up, we need to add more content about this, and we've also seen regional differences in the, some of the responses from trainees and so then we will go back to the regions and say hey, we notice that you had a lower percentages for all, you know, the three last scenarios out of the four, what happened in the training room. And we'll, and they'll come back and say oh we ran out of time, we used a lot more examples in the other end of the training in the beginning, end of the training and we'll rectify that. Yeah, go ahead.

Questioner-2: In this California is such a large and the first thing I am sure is about some of the regional variability, are there cultural influences that you, against the influence of the culture that you are getting that kind of feedback and being able to make changing you know to evaluate the curriculum or meaning it...

Cindy Parry: We've gotten some feedback about English as a second language for some of our embedded evaluations and just needing more time. And I will say our scenarios are a couple of pages long. There can be depending on the region trainers are trying to address cultural differences within that region and being mindful of that it's been a struggle I would say.

Barry Johnson: Well we also have as one of the themes that runs through our entire common core. What we've called fairness and equity and that deals with disparities and also cultural issues. And so there are little icons and they are encouraged to, I mean the content is the same, but it doesn't mean that you can't use a local example, right.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Questioner-2: And we are thinking of the urban or an having really rural, you know, something California with some of those things where urban or suburban's say we need a chance, but they are not really rural, should be a real problem here.

Barry Johnson: Yes, alright. We are not being good about reminding people to use the mics.

Questioner-1: To be honest [indiscernible] [01:05:38].

Barry Johnson: So do you want to just repeat that briefly?

Questioner-2: Yeah. Well the cultural piece that I was really speaking of the, was the rural urban diversity there and are you, well and just the question to you Leslie was does that show up? Do you see that kind of regional variability that way?

Leslie Zeitler: We have a, as that specifically on the demographic forum. So I wouldn't say we could actually break it down that way based on the trainees. Can we? I mean it's...

Cindy Parry: Well no. We know what county they are from and so we can do it.

Leslie Zeitler: We know what county they are from, yeah. That's as close as I think we can get.

Cindy Parry: And where it has surface I think is when we've been discussing in the child mental treatment identification discussing neglect. And even amongst the macro evaluation committee people just identify very different standards for neglect.

Leslie Zeitler: Yes.

Cindy Parry: And we are still, we haven't given up on trying to asses it but we need to get some kind of consensus about what it is when we see it first.

Leslie Zeitler: Right. And I'll talk a little bit later about us trying to take a look, one of the things that we have found for some of our test results is differences by gender, ethnicity or race and ESL and that's part of the analysis that we've been doing and trying to improve the testing items and questions and try to eliminate some of the differences too. Okay, so did you want to cover this briefly and I think I need to...

Questioner-2: I think very briefly.

Leslie Zeitler: Hold this up. Hold on just a second.

Questioner-2: In fact maybe I will just sit here.

Barry Johnson: Yeah, please sit there.

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

Leslie Zeitler: Okay.

Questioner-2: Well I can have to look at whatever on it.

Leslie Zeitler: I need to pull it up for the, there it is. Okay.

Barry Johnson: Just there we go.

Leslie Zeitler: So what I am opening up here...

Barry Johnson: Yeah, I think you should...

Leslie Zeitler: Did I just reopen it there?

Barry Johnson: Yeah.

Leslie Zeitler: Oops, sorry.

Cindy Parry: Maybe we will just talk about it.

Leslie Zeitler: Okay.

Barry Johnson: They have it.

Leslie Zeitler: Okay.

Cindy Parry: Yeah. You do have this as a handout, but unfortunately we discovered that when it got converted to PDF it cut some things off. So if anybody is really interested in getting a copy of this, we certainly give to you or is it on the website?

Barry Johnson: It is on the website.

Cindy Parry: It is on the website.

Barry Johnson: We still have a problem with it, but, I, there is just a couple of things with it.

Leslie Zeitler: Yeah.

Cindy Parry: But we talked about, you know, customized products and one of the things people had asked for and this is an example was when they looked at their regional reports and they saw a difference between their regional reports and they saw a difference between their region and the statewide figures on that bar graph. If they didn't do as well as everyone else did statewide what did that all mean, so well, what this is if we try to put together a flowchart for people to look at and say okay, if I see that we didn't do as well,

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

is it something we need to address, is it something that needs to be addressed elsewhere and this flowchart is just sort of a decision tree that we made for them. So we gave it you in case that's helpful to you. If anybody has feedback on it we'd certainly love to get that too and you can get this on the website.

Leslie Zeitler: We'll get there. Okay, so the next one. We are going to be talking about where we were going in our strategic plan, our second strategic plan. So at this point we are now in the middle of our second strategic plan. We are getting ready to embark on our third. We are going to, some of this will stay the same, our efforts as with our first strategic plan. We will continue to collect demographic profiles for both our line workers and our supervisors and do analysis of line worker core test data. We will continue with that and also related analysis of our supervisor core test data. We didn't talk about this very much, but there is an embedded evaluation for our supervisors and in the, during the first strategic plan it underwent several iterations and so we didn't have a large enough and to really be able to make any kind of, we couldn't say this was significant or this wasn't significant or this is how the supervisors were doing.

Now during this second strategic plan we are, we've been working closely with our curriculum specialist and it's undergoing a major revision and we've piloted it now to determine whether or not both the curriculum and the embedded evaluation seem to work well together and how trainees respond to it. We will also continue to do analysis of 4E trainees and their test data.

Level-2, we will continue our formative evaluations and we've made new forms so that we divide people's assessments, observer's assessments into looking at content and also looking at delivery of the training. And we hope eventually having e-learning platform and we develop, we will plan to develop a formative evaluation for that as well.

Level-3, again that satisfaction level. Again this will only be evaluated at the regional county levels. Well, we won't be doing that at the state level.

Level-4, further time being. We are going to continue doing our knowledge tests, our multiple item, multiple choice test questions for the same curricula that are currently evaluated at this level. We are, as Barry said or as Cindy said earlier, we are moving towards more diagnostic testing and what that means is our statewide oversight group has said, you know, for this curriculum let's say its permanency and placement. We would like to emphasize these five learning objectives and that's where it gets the most time or coverage during the training day. So now we are tailoring the test to cover those five learning objectives, is that make sense. And that way we can really hone in on whether or not trainees are seem to be really getting this content or not and that way we can also make targeted revisions start training based on the evolve data.

Cindy Parry: Yeah. As yet when we talk about diagnostic we still haven't tackled the idea whether we want to give individual feedback. So at this point we are just talking about trying to almost build a little subject structure so that we can at least give general

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

feedback about what areas people are doing better and worse than. So it looks like we are getting close to the end. Should we open it up for questions or do you have left any?

Leslie Zeitler: I will just do these last ones really quick.

Cindy Parry: Okay.

Leslie Zeitler: Okay. So again for knowledge test level we'll continue to do the differential functioning analysis, something that we are looking at right now. We are having a study in our southern region to look at the possible effect of stereotype threat in trainee test performance and I don't know how many of you are familiar with stereotype threat. Is anybody know? Okay. So the definition of it is being at risk of confirming a self characteristic, a negative stereotype about once group. This is steel and earns in 1995. It's a situational threat and it means that it's not dependant on the internalized belief at a given stereotype. It's just a fear that one will be judged by others through the lens of the given negative stereotype. So if we are asking about demographic information, sensitive information from our trainees and then we are administering tests later on in core, it might have an effect. We are asking about their demographic information, for instance the race or ethnicity, gender, ESL and things like that. So what we are doing in our southern region is first, some groups and this was, we had worked together, I work with Cindy and one of our southern reps to determine which cohorts, trained cohorts would get there at demographic form at the beginning the of cohort and then which cohorts of trainees would get there at demographic forms at the end of course after they have taken all the tests and see if there is a difference in the results and how trainees respond to the tests questions. Is there anything you want to add quickly about that?

Cindy Parry: No, not really. I think that we are just trying to do everything we possibly can to eliminate any, in this case based on, differences based on very specifically.

Leslie Zeitler: Yeah.

Cindy Parry: We've done everything we can to eliminate potentially biased items. We've just determined that every one of our subgroups or race and ethnicity subgroups achieve significant gains from pre to post, but there is still is a difference in the final posttest level that we would like not to have there. So this is another thing that we are trying to see if we can totally eradicate some of these non-test or non-curriculum related differences, potentially non-curriculum related differences.

Leslie Zeitler: I am going to move us along. Level-5, this is our embedded evaluations. We will continue to asses for differential performance by demographic groups. Now we've piloted the embedded evaluation of our SDM, Structured Decision Making version of our critical thinking curriculum. And what I would say that's been interesting about piloting that curriculum is there the correct answer in terms of, I don't know how many of you are familiar with Structured Decision Making, but there is a section of the forum, safety and risk assessment forums where you have to determine whether or not they are

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

override, so are you going to override some of the decisions. And so what we found in the, just piloting this is that one region got it all correct and another region, they had more than a handful of people choose to override out of a small training class. And so we went back to the region and said hey, what's going on, because if we didn't have the embedded evaluation we wouldn't know that it happened and they might go into the field and do these discretionary overrides when it's not appropriate. So that's one of the uses also of collecting this information and as I said before we are revising that it evolve for our supervisor model. And we hope to actually pilot the neglect scenario as a part of an embedded evolve.

We are in the process now of conducting a feasibility study for transfer of learning. We are trying to work with some of our partners to do that. And Level-7 outcomes continue building our chain of evidence and linked outcomes. And then we are, we have other training evaluation projects outside of our levels and really quickly the one that I want to highlight or focus on is our attitudes or values of evaluation regarding the impact of attitudes towards sexual, child sexual abuse disclosures and this is a collaboration with the UNC School of Medicine. And then we also have quality assurance where we are trying to see about the fidelity of the training and how, and this is for a curricula that are undergoing revision in a given fiscal year. And just seeing can, is there too much content in a given training, is this taking too long and that also helps us do better in terms of making sure that the training serves our trainees and that we are improving curricula. So, turn over to Barry.

Barry Johnson: Alright. So do we have, and we'll put it up, turn over to you if you all have questions. We are almost out of time. Sorry, we didn't manage to do this. You can do this on your own, think on your own of this questions, but we will answer any.

Questioner-3: Did you, sort of work out any, how you are evaluating the critical thinking piece? I am not sure I really captured what is involved. Is that still a multiple choice test? Are you sure it's a, we are writing sample, what does that consist of?

Barry Johnson: Yeah. It doesn't really test, it tests knowledge, it doesn't test critical thinking. I wish we didn't test at critical thinking, but I talked to Michelle Graef at University of Nebraska-Lincoln. They use to have a critical thinking test that they administered to select candidates for their agency which is I think the way to go with something like that, yeah.

Questioner-5: Is there any part of the evaluation of the trainings that require candidates to do practical skills in terms of doing a role-play and then they are evaluated at how well they do the role-play or is it mostly diagnostic questions and the more critical thinking parts?

Barry Johnson: And we are getting at that through our scenarios and the embedded evaluations, but we didn't, and we don't have the, it takes a lot of person power to do that. You know, you can train peers to evaluate but that's not very good and so I mean it

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

takes a lot time too, so no. We are trying to approximate that with our embedded evaluations and have them apply the knowledge in the classroom.

Questioner-5: Okay.

Barry Johnson: We do however, whenever we have an embedded evaluation, whenever we do, we try to get across the people in terms of skills that if you want people to get a skill they have to, you know, hear about it, learn about, practice it, perform it. So you know you can't have 15 skills in a training and if you are going to evaluate the skill, you better do all those things, so...

Questioner-5: Thank you.

Questioner-6: I was just curious how involved where the trainers in the process of this itself? When did they become more or less involved?

Leslie Zeitler: They were involved in development of the content of the curricula. We had subject matter experts participate in groups for each of the curriculum topics and then for designing the tests themselves or the tests questions. So for instance for the knowledge tests, we also had the subject matter of experts either participate in a formal group that we had scheduled over a period of time or they gave feedback electronically. And then for the embedded evaluations they were heavily involved in making sure that in terms of making sure we had the red flags correct and ask, how we ask the questions and whether or not they agreed with our, what we thought were the correct answers and they were, every step of the process they were involved.

Questioner-6: Before hand where they as involved, before you embarked on this?

Leslie Zeitler: I don't think in the same way. Barry?

Barry Johnson: No. I mean I would say you know not all of the trainers were involved in that process, but we did have trainers involved in that process. So it was sort of a legitimate in their eyes, because we used a lot of the sort of master trainers, the long-term trainers were the ones involved in it. I think that they, you know, part of doing this and doing with our standardization of the common core has made people much more aware of evaluation generally. So the trainers now, they don't, I don't think they complain as much about the time taken up with evaluation maybe as much.

Leslie Zeitler: I think it's still little.

Barry Johnson: Yeah.

Leslie Zeitler: Feedback, yeah. And I would say it's more systematic now. I think having this process in place whereas trainers could, would probably the master trainers could probably give their input from time-to-time when things came up. But this, the

Session 5.01 – Assuring That Training Has Impact: Evaluating a Large and Complex Training System

process of having the statewide process and having standardized curricula, it's more systematized to get their input. Any other questions?

Barry Johnson: Any other questions? Oh! Let me do one more plug before you go.

Leslie Zeitler: Oh, great.

Barry Johnson: We, CalSWEC sponsors and cosponsors the National Human Services Training Evaluation Symposium every year. I don't know if any of you have went to it. It is, it alternates now between being in Berkeley at Cal and in another site. It's our fiscal crisis in California cause this to alternate. But last, this year in June it was at Cornell, it was great. It is, you know, I encourage anyone who is interested in training evaluation to come. All of the proceedings are published and they are on our website, so you can see past programs and everything and it is in May.

Leslie Zeitler: Usually in May.

Barry Johnson: Yeah. You know, it has been the week, the few days the week before Memorial Day weekend in Berkeley. So we are planning on doing again this year and we will try to get information out. Everyone, but please do check into it. It's a great group of people who are interested in stuff in a very interactive environment.

Leslie Zeitler: Let's say that one of the benefits of this symposium that Barry is talking about is that people kind of share their dirty laundry about what their challenges are in their training systems and then they can also get feedback from colleagues in a supportive way on have you thought about this, have you thought about that and so it's meant as more of a supportive environment for training evaluators who are trying to deal with political systems and the rigorous of research and evaluation and balancing interests, so okay and that's it.

Barry Johnson: Thanks everybody.

Leslie Zeitler: Thank you.