

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Panelists:

Kristine Frerer
Jordan Horowitz

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

Kristine Frerer: Good morning. How is everybody doing? Yeah, okay. My name is Kris Frerer. I am a, Associate Research Specialist at the University of California, Berkeley Center for Social Services and Research and I am here with Jordan Horowitz from...

Jordan Horowitz: I am with the Institute for Evidence-Based Change, not for Profit Educational Research Group in Southern California, throughout California. My office is in Southern California.

Kristine Frerer: Okay. And we have, we are going to present on two efforts...

Jordan Horowitz: And we are much closer than this seems.

Kristine Frerer: Yes. We are going to talk about two efforts that took place in California and that we will link to Child Welfare and Educational Data and looking at different opportunities to improve outcomes for Foster Youth and Education, so...

Jordan Horowitz: And it didn't say it in the conference abstract. They took out two words. I am actually going to demonstrate this, web based tool for accessing the data. So, you will be able to see exactly how it works and what we've been to accomplish.

Kristine Frerer: Okay. All right, then we'll start. Just my overview, I am going to talk about a pilot project that we conducted, the UCB Cal-PASS pilot project and I'll talk about overview and challenges. We had quite a few challenges trying to link these two data sets. I am just going to give you some selective outcomes. We kind of mined the data for all it was worth and so it's just a long report. I'll talk about lessons learned and actually looking at what we are doing right now, we are moving forward with the statewide initiative. So, there, and we're called the California Ready to Succeed Pilot Project. So, a little bit of background, you guys all know there is over 400,000 Youth and Foster Care in the United States and California has a lion's share of those youth, over 15 percent are in the State of California and most of those youth are school aged.

We know or we think we know that affects the mental treatment, abuse and neglect can compromises the use ability to learn. And in terms of some outcomes what we know is that compared to the peers, foster youth are more likely to have a lower achievement test scores, they perform below a great level and they are twice as likely to leave high school before completion. Oh, by the way if you have questions, just interrupt me. So, there is

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

not that many of us, so you know, we will discuss things as they come up. So, although a lot of entities identify foster youth as an educationally vulnerable at-risk group. This distinction does not exist in federal law. In fact the government law of no child left behind identifies groups at risk for academic failure. We have certain ethnicities of poor children low socioeconomic status. English language learners and students with disabilities and it doesn't really delineate foster youth as a group.

But what we do know is that there is a considerable overlap between children identified at risk and those in foster care. We know that kids in foster care are disproportionately poor. There are disproportionately children of color. We know that once in foster youth or foster care they are disproportionately placed in the special education classes. And something else that we don't know that much about, but, there was one study that Smith Gull did are regarding poor quality schools that while foster youth represent less than one percent of the student population, they are often concentrated in the worst performing schools. In California we do have some data challenges. At child welfare, we have a statewide data system, CWS/CMS and that captures all data for our youth that come into contact with Child Welfare System. In terms of education while we have the capacity to collect educational data on where more kids go to school, grades things like that, the data are woefully incomplete that the state mandates that social workers create a Health and Education Passport. But, really in terms of the quality of the data that goes into that Health and Education Passport is not monitored and sometimes it just comes down to supervisors and social workers that just want to keep capture that data.

So there is really, it's just, we can't really use it the way that we want to use. It's also, the educational data is only for children with an open case that they are in foster youth or they are in a family maintenance program. Once youth exit foster care, we have no idea what happens. Where they go, there is just no way to track them. California at the educational front also has some data challenges. Currently we do not have an operating statewide educational data system that the state is working on a system called CALPADS, which has the ability to be longitudinal. We are having some little fiscal problems in the State of California, our budgets are being slashed and the budgets to actually get CALPADS, up and running is gone right now. We are hoping some time it's going to be reinstated, but, right now it's a possibility in the future. It is actually going to be focused really on this K through 12 progression with the capacity to expand to post-secondary later on in the future community colleges, CSU's, the state universities as well as the university, the universities.

So right now each school and each district keeps its own data and student data and reports aggregate numbers to the state. So, there is actually no dataset that we can tap into and if we want to start linking child welfare data to educational data, it's district-by-district-district and sometimes school. It just becomes very, very hard. There is also a confusion over who holds the educational rights for a foster youth. You know, and that's still unclear whether the parents retain those rights, whether the foster parent has those rights, whether the social worker can somehow kind of intervene for the children and so what happens is a lot of times nothing happens, that no one takes responsibility for the educational health of a child. And then lastly, and not least, local jurisdictions, they very

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

often *[indiscernible]* *[00:06:31]* information sharing and interpretation of this, the federal confidentiality laws, which is FERPA, its the Federal Educational Rights and Privacy Act, which dictates how educational information can be shared with, who it can be shared with and I will talk a little bit about that coming up. So, to do this we do have some relevant policy.

We, at this, the federal and state level, there is no child left behind and even though the foster youth are not really delineated specifically we do know that they overlap and so it behooves us to actually look at educational outcomes for foster youth. The child welfare side, we do have the Child and Family Services, review why youth are in care, there was a wellbeing measure that, I think it was kind of abstract but, that foster youth receive the educational services they need, something to that tone, but, there is a directive to look at educational outcomes for foster youth. And also there is the Chafee Foster Care Independence Program that amended states to, actually look at educational outcomes for youth who age out or emancipate from the foster care system. So, here is, our collaborative partners, to actually try to fill these, this gap, in knowledge regarding educational outcomes for foster youth. Some agencies got together and one of them is the one that I work for is the University of Berkeley, Center for Social Services Research.

That we, the children's archive at CSSR receives quarterly extracts from the Child Welfare System we have some existing interagency agreement with the California Department of Social Services, we have an existing MOU to receive these extracts and that one of the things the archive does is reports the, this CSFSR outcome measures on the website. We have a Child Welfare Dynamic Reporting System, also anything you want to know about child welfare statistics in the State of California is actually on this website. It's a fabulous website. So, another partner is the California Partnership for Achieving Student Success, which is Cal-PASS and Jordan actually works for Cal-PASS, which is now called the Institute for Evidence-Based Change he shall tell you all about that in his presentation. So, the Cal-PASS is the only system that collects data about student success and transition from every segment level of Education in California. Do you want to talk a little bit about Cal-PASS or just...?

Jordan Horowitz: I will get to my presentation and I will talk about the details, what we collect and how, so...

Kristine Frerer: Okay. And then lastly but, not least, The Stewart Foundation, which is our very generous and patient supporters, financial supporters that The Stewart Foundation has funded not only the pilot project, but, also statewide progression, a 100 percent funded us. So, we appreciate them very, very much. So, the pilot project, so in 2009 these are collaborative partners, the project team we started plants to link data from CWS and Statewide Data System and to the Cal-PASS educational data. For the pilot project we decided to only go with four California counties. Three became, our project was actually under a larger group of projects sponsored by The Stewart Foundation called the Ready to Succeed Projects. And so we chose three of the California counties for the pilot project, because they were already doing other things with the Ready to Succeed, they were in other different projects that Stewart was doing. And the fourth county that

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

we picked was because they had a very strong relationship with Cal-PASS and so there was four counties.

Audience: What are those four counties?

Kristine Frerer: They are Fresno, Sacramento, San Diego and then San Bernardino where the four all together, so kind of midsized nice grouping. From those four counties we identified over a 150 secondary and post-secondary districts. And so our goals on the pilot project were actually kind of backwards if I must say so, that we really, were just trying to see can we do it. Other people have tried before, California has been trying for at least a decade to link Child Welfare Education Data and no one has really been successful. So, we were just trying to see can we do it. Can we figure out the mechanics of actually doing it and more importantly can we get through this political kind of landmine and get the actual permissions to link these data. And then lastly we were going to look at analysis, but, that we didn't really, we weren't really thinking about it. We just didn't, we were just like can we do it and that actually became the thrust of the pilot project.

We started out with some permission's IRB CSSR is we have an MOU with the California Department of Social Services to receive these quarterly extracts and to use these data, the Child Welfare Data. And so we had permission, verbal permission from the head of CDSS. Though they don't actually own the data, that State of California owns the data, so we actually, we had to submit an IRB to the State of California to use the data and our, we have some question very, very broad. It was so broad, where we didn't really know what we were going to find. But, we also had to submit an IRB to the University of California at Berkeley because CSSR, we are actually an entity of UCB. So, that wasn't so bad, so we had two major IRB protocols submitted and once they were approved we can actually move on with the project. Two permissions from the school districts, which got a little bit dicey and Lauren Senko who is actually manages the project at Cal-PASS did an incredible job getting permissions from the school districts. Cal-PASS, do you want to talk about the volumes or anything...

Jordan Horowitz: Yeah. So, briefly, Cal-PASS, we are not a state agency. We are a voluntary system. It, that began in 1998 when our Executive Director was Director of Institutional Research for Community College District VAS the accrediting agency was coming back and the faculty came too it and said we told VAS when they were here four years ago that we would track our students from the Community College District who went to San Diego State University to see what happens to them and we have no way of doing it. So, that's when the, our data encryption and the way we linked data across the segments was created. Then the local K-12 said, hey, well we weren't in on this and it became a regional data sharing consortium that slowly expanded to statewide.

About six years ago we started to get some state funding, but, it's a fully voluntary system. So, we have about three fourths of the K-12 districts that have high schools in them giving us data, I'll get into details about the data in my presentation. 18 of the 23 California State University Campuses and nine of the 11 University of California

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Campuses and all 112 of the Community College Campuses give us their student transcript level data and other data. But, the permission, the ownership of the data resides with the districts, we don't have any ownership. So, when we get involved in a research project we need to go to each district that we want to involve in the project and get an okay from them to use their data.

So, and even though it's fully anonymous all the data are encrypted, they are all, student identifiers are stripped out, so we cannot tell or go back and figure out, which student links to, which data. We still have to get permissions to use the data from any district who want to involve. So, my team had to go back, Lauren Senko as Chris mentioned led the team and get permissions from each of these 150 districts to include their anonymous data in the project, so...

Audience: And because it was kind of the first time that you guys have taken on a project that, but, this manages right... That they decided that an active consent or response was required so, wasn't everyone were trying to get on the phone or get an email or like an, someone say yes on the phone and so that actually kind of group in terms of how...

Jordan Horowitz: This project, this project causes, led to us changing our MOU's with the districts and we now use a passive consent for involvement in research. We got rulings from attorneys and of course and so on that we could do that because it's all anonymous data. So, we've now gone to passive consent, yeah.

Audience: Yeah, right. The passive consent, do you send out letters to all the parents and schools?

Jordan Horowitz: Oh, because its anonymous data, we don't need parental consent, we just need district consent, because they own the data.

Audience: So, is there what's the...?

Jordan Horowitz: So, we said we have, in our MOU there is a point person identified at each district who gives that consent, they get an email from us saying here is a description of the project by this date. If you do not want your data to be included then we need to hear from you and we use a reverse conformation to know that they have received the email and that's how we work it.

Audience: *[Indiscernible]* [00:16:19]...

Jordan Horowitz: Oh, 150 K-12 and 470 districts.

Audience: Have you used, are they employed managers or...?

Jordan Horowitz: No, they are not. In fact, some of the districts that require IRB protocols, for example Fresno, San Diego both require, school districts required IRB's and no each district has their own approval process as well as Berkeley's and the states.

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Most of the districts were fine, the state and Cal, both approved if with their IRB's then they are confident, but, the occasional district wanted us to go through their IRB's and that led to, you know, longer timeframes and their IRB, oh, well, gee, our IRB's doesn't meet until June. It's February, so that's four months more added to the project before we could get started and so on and so...

Kristine Frerer: It was a lengthy process that's just about it, so, yeah.

Audience: So, to track the students you encrypt those student identities...?

Jordan Horowitz: I will get into that when I do my presentations. Exactly what we do to encrypt and our data security procedures, so...

Kristine Frerer: Okay. But, anyway, so that we have this active consent and Laurent's on the phone or on a computer trying to get out the school districts to respond and unfortunately that started, that kind of coincided with the start of California's fiscal crises and educational moneys were being taken out of the budget and so even for some districts even just responding became overwhelming and, you know, one patiently just kind of badgered and hammered at them and we just talked about the IRB's. Well ultimately because of this effort we got 91 districts to agree to participate and 91 districts were actually represented over, I think 65 percent of the student population within our four counties, so we got the right districts so...

Okay, so and then we had a couple of challenges. We had more challenges actually and one of the things is we overlooked permissions from an important stakeholder. We had permission from CDSS, we had permission from the state to do this, but, we didn't have permission from the CDSS Research Board and it was daunting but, we needed that apparently, because we needed to get an MOU between Cal-PASS and CDSS, Berkeley we already have done. And so we overlooked the CDSS Research Board and when they kind of got wind of the project, they were not happy to say the least and so they called a meeting and Jordan and I, we are fortunate to be able to get to Sacramento to attend that meeting in person and then my boss Dr. Barbara Needell has a really incredible relationship with the people at CDSS and made sure that people supportive of the project where in that room and on the phone. And so what we actually thought was going to be the death knell of the project turned out to be really good so, yeah...

Jordan Horowitz: So, I must say Social Services Research Committee has more attorneys than researchers on that. That says it all.

Kristine Frerer: But, so that was kind of fright and mode picture of that well. MOU was in place and so now we have a full support, we didn't miss anybody. A more important challenge actually occurred in late 2009 and the federal government released re-interpreted FERPA guidelines, which before it was de-identified and they kind of quasi about who can merge and you cannot merge and the new guidelines basically said it has to be anonymous that if a children can't be, they can't be identified anywhere. They can't, the kids can't be linked back to any system, so, and on top of that the original plan

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

really was to upload Child Welfare Data to the Cal-PASS, they have a secure server, Jordan is going to talk about that. It's actually a very secure process. But, we have, at CSSR we have an existing MOU with CDSS about how we can transfer data. It's antiquated, it doesn't really make any sense but, what happened is we had to follow it.

So, the team went back almost to the join board to figure out what or how we are going to do this, what is our process, how we are going to actually link these two datasets and get it back to where it makes sense. We can do some analysis and so what we came up with was this. So, just talking about that, start up at the top CSSR that's me, extract the Child Welfare Data from CWS and we looked at unique children that were in foster care from 1988 to 2008 and we identified over 90,000 are school age kids that who were six and older. So, pulled those kids, pull every child over there above that we said that we are going to look at as well. Assigned a unique project id, which cannot be linked back to our dataset or Cal-PASS's dataset, put that all under secure server, which is where I pulled sensitive data, which names, Social Security Numbers, dates of birth, all that kind of information. And then took that information and encrypted it and Jordan is going to talk about they're encrypted, how they encrypt.

But, basically we created two encrypted ids. One is an encrypted social security number, the other is kind of a, based on a quasi probabilistic matching schema that actually combines variables, first and last names, date of birth, all that kind of stuff. It's called a Drive Key. So, that, those two encrypted id's as well as the unique project ID were loaded on a external hard drive and when I got into my car and I drove it up to Sacramento, I, live Berkeley, drove it up to Sacramento and handed it off to a representative of CDSS, this is antiquated data exchange system. So, and luckily Debra Williams, we thank her and her staff so much for actually being our hand off person. So, she takes, Debra takes the flash, the hard drive for me and kind of holds it till someone from Cal-PASS actually, Alex from Cal-PASS, drives up there and picks it up. During that time he has actually taken all of the educational data that we might need from the Cal-PASS server and placed it on a non-networked, even though it's a non-that's just way it has to be a non-network hard drive.

There he does the joints that based on these encrypted id's he joins it to the Cal-PASS dataset, which results in multiple academic records we had multiple files depending on what the educational variable is, California Standards Test's the CAHSEE Award for secondary, for foster youth and we were also able to pull a comparison group. So, we have all the information loading back on to the external hard drive. We go drive back up to Berkeley where it matched and, you know, the time we get back to Berkeley and we match it back to the Child Welfare Data. There is, it's only the project id, all the encrypted ids have been stripped. So, it's ultimately a dead file, it's an anonymous file, it cannot be updated. So, you know, getting back we figured out the process, its FERPA complaint. But, if we want sort of to recreate this data, we have to run through this whole process again, so...

Audience: How did you create your conversing group?

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Kristine Frerer: You know Nathan, actually Nathan, the researcher at Cal-PASS did that in and basically there were different levels of how he pulled the data. But, the basic although was gender and ethnicity and age. So, we tried kind of an aging schema, like within six months of the date of birth and so that was it and I will tell you how he matched them. But, it just basically what he pulled was this really nice generous pool of comparison kids that we could actually do more exact matches to, so... So, here is our real match statistics. We started out with potential pool of 90, over 90,000 unique foster youth school aged. At the K through 12, level we matched 39,000 somewhere in the Cal-PASS dataset and I will talk a little bit about the challenges about in a second. Matched over 7,000 in the community college level and 400 at the university level and so across the three school segments we matched, we had a 47 percent match rate. Right, okay, slightly funny so, let's talk a little, I just want to talk a little bit about the complexities of the Cal-PASS dataset and do you want me to talk about this or do you want to talk about this?

Jordan Horowitz: Go ahead. I will step in.

Kristine Frerer: Okay. That because our Cal-PASS is a voluntary consortium where the district has voluntarily agreed to upload data that the, what types of data they upload is kind of up to them. Now they have a guideline of what they want to see, how many years of data and types of data, but, it's really up to the district what they want to do and so for some districts they upload everything. Five years of data, we have the California Standards Test's, we have the CAHSEE, the excellent High School Exit Exam, the board course file post-secondary, you know, so for some districts we have everything and others, not so much, right. And so some districts only have one year of viable data, other ones have like Fresno, Fresno Unified was just a, we thank them so much, they had so much great data that we could use and so, but, it is kind of complex about what we can say and how we can actually pull these data into something that make sense.

So, it's kind of a hodgepodge a little bit sometimes. Did I, anything else? Okay, all right. So, talk about selected findings. For the most part we were just kind of, like I said before that we really were working on this mechanics and we didn't figure out our analysis. So, once we actually linked the data I'm like, oh, we've got so much data what do we do with it. So, we went back to it with our stake holders like what do we do and know what we didn't really decide anything, so what we decided to do is just kind explore. What do we have, how can we use it. And so for the majority of it, based also on the complexities of it, the education data, we did these cross sectional analysis. So, we didn't, really do a longitudinal work. And this is just the K through 12, level we also did work at the post-secondary community college university and the CSU level. So, overall we had about 4,200 unique youth with a history of foster care placement and so we looked at two primary examples for each measure.

[Informal talk]

So, we looked at two primary examples. One, we match foster youths with comparison students and we matched them very, very closely. We matched them by age and grade

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

level, school year, gender and ethnicity and we also matched them by a series of educational risk variables. We looked at ELL status and in California ELL status was four levels anywhere from English fluent to, you know, speaks it completely speaks another language. There's, four levels and for the matching part we matched them exactly to the level at they were, but, for analysis we condensed it down to either they were ELL or they weren't ELL. We also matched exactly by primary disability, which is kind of special education status, that's just the name of the variable primary disability and once again for analysis purpose we reduced that down to yes or no. For matching purposes if that foster student had a hearing disability that comparison student had to have that hearing disability, they were matched exactly by disability. Free lunch participation, in the free reduced lunch and then we also matched by school district and school. If school was not possible, every one's matched by district, but, if school was not possible, we matched by a measure of a school quality is what we came up with. In California, California department of Education ranks the schools each year based on a variety of measures. One of them is their API and then there is a kind of a black box mechanism that goes in and the number of percent of, you know poor kids there is a lot of variables that go into this black box and what happens is they spit out a rank between one and ten. So, we use that rank, one being the lowest ranked school and ten being the highest ranked school, the highest performing school. So, we use that, we condense that down to four categories and matched by school rank.

And then for analysis purpose we actually used school quality as a measure and we reduced it down to just really poor schools and all other schools. We also looked at just the foster youth only group and that's where we explored some child welfare variables. So, our independent variables were the demographic education risk factors, which we kind of talked about gender, ethnicity, English language learner, disability, free reduced lunch, school quality and at the post-secondary level we have financial aid. Child welfare factors age at entrance, whatever reason, length of state, number of placements, episode or spell total, basically is it their first entry, did they reenter and then we looked at exit type outcomes. We looked at a lot of things at the K through 12, level we looked at course work in English and math. We looked at California Standards Test performance level, which is our standardized test that they give to all kids in grades two to 11 on a yearly basis. It is age or grade norm referenced. We'll talk a little bit about that. There is a moment to talk about it right now.

There is five levels ranging from far below basic to advanced or proficient and advanced and so we actually reduced that to whether they were proficient or not, which is, if the California Permit Education Standard. We also looked at CAHSEE, the California High School Exit Exam if they passed it by the end of tenth grade. We looked at high school award at the post-secondary level. We looked at entrance whether they entered community college or university, persistence whether one year after they entered, where they still there, looked at basic skills, English and math and post-secondary award, so we looked at a lot. We had a lot of different outcome measures. And so our analysis approach for everything was pretty much the same. We did a scripted analysis and we did some bi-variant work mainly with the foster youth comparison group, with the ten percents and differences between percents or links between groups. And then we ran a

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

higher level model, we run a multivariate logistic regression model, I am not going to go into it, but, we actually used the modified persona approach as outlined by Zoe *[phonetic]* [00:31:04].

And so here is an example, with sample one, this is a foster youth comparison group looking at the differences between the groups, the percent proficient on the California Standards Test is the first one. So, we have grade eight, grade nine, grade ten and eleven and that is the percent proficient. So, just talking about grade eight at the CST English Language Arts, 22 percent of the foster youth were proficient in grade eight compared to 29 percent of the comparison group and there is a significant difference between those two percents or two groups. And what we found was in English, there was significant difference in grade eight, grade nine, grade ten and grade eleven, so there were some differences. The second set of numbers is actually course, whether in grade nine they were taking an appropriate level English course, same MO, differences between percent, those are also all significant. And then we looked at the passage of the CAHSEE, the English language arts at the end of tenth grade and there was whether less foster youth passed by the end of the tenth grade and the differences between the percents is not so delicate.

So moving on to the example of our higher level model, multivariate model and so I am just going to walk you through it. These are all the same measures I just talked about. I am going to focus on the California Standards Test's English Language Arts. We have the grade level and then the same size and our model controlled for gender, ethnicity. Our reference group for ethnicity was white foster youth or white students. English Language Learner, whether they are, the reference group was, they were not ELL students, free reduced lunch once again they were not participating, disability the same thing. We looked at bottom ranked school and then finally when we looked at foster students the reference group was comparison students. So, even controlling for all these variables within our model, these educational risk variables that foster youth were still significantly less likely to be proficient in the California Standards Test. Same you see in the English course. They were certainly less likely to achieve a passing grade, which we define as C or above or in terms of the CAHSEE there was no difference between the two groups when controlling for the other variables.

Example of English, this is our sample two, this is foster youth. The model controlled for a lot not going to go into it, but, what we really kind of found was that the educational risk factors were significantly less so the child welfare variables I mean where there was some support for exit type, kids that were unified as adopted and placement numbers but, overall, it was these educational risk factors that were more salient. So, findings of overall foster youth were certainly less likely to have positive age outcomes compared to the closed matched peers, once again the education at-risk factors were more salient predictors than child welfare variables and especially kids with disabilities really, really doing poorly, special aid kids and I just talked about that. The child welfare factors really not a lot of consistency across the board, but, our analysis was limited to cross-sectional, so just to be, keep that in mind. The full report is actually The Stewart, is available on The Stewart Foundation. Last time I checked they still hadn't, posted it but, it's ready to

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

go, so if anyone is interested. So, limitations, definitely administrative data were missing a lot of variables. We do with what we have. It's a snapshot, our analysis, cross sectional analysis is really a snapshot of how youth performed on a given education outcome in, you know, one school year, which is going to be misleading.

Risk and outcome are measured simultaneously and basically it's really hard to make practice for policy recommendations. It really was just kind of a, gave us idea of, if we move forward what would we look at. I am going to go through this really quickly, this is longitude analysis, actually I used, my dissertation I used the dataset from my dissertation at UC Berkeley and so I did some longitude analysis for data and sample sizes were pretty small, 455 unique foster youth who entered foster care for the first time in grades four to eight. I matched once again I have two primary samples. I matched them to a comparison group at baseline. So, the entry years were grades four to eight baseline years were three to seven. So, additionally to all those factors that we've just talked about the cross sectional analysis, the only difference was I actually matched them exactly to the performance level baseline. So, all the kids based on the variables looked exactly the same before they entered foster care. I once again looked at foster youth only, independent variables are same, child welfare factor is a little bit different, majority placement type, we looked at the variables a little bit differently.

I also looked at an educational risk, some which, I summed ethnicity, black and Hispanic, I would have some Native Americans, but, there was in my sample, ELL disability, free/reduced lunch and poor quality schools, so those are quite possible. Also added a movement in year-one and so movement was to find, year-one was when foster youth actually entered foster care. So, they are entering and looked at further movement, a normative school transition, which is they moved from, you know, elementary school to Junior high. Looked at non-school, non-normative transfer where they moved in, you know, they moved schools and looked at if it was due to a placement change or not. Residential with a placement, change, exit from foster care and reentry and I also looked at kind of the cumulative impact and for, want of a better term total year-one changes, which is the sum of normative school transition. A transfer, a placement chance in exit and a reentry, so going through this outcome was performance level, the California Standards Test, English and math, once again those five levels.

So my analysis approach was different I, looked a group based analysis, which assumes a number of discrete classes, each having specific intercept slope, an estimated population prevalence and I used a sense of nominal distribution because there were five levels for the CST's and then I looked at multinomial logistic regression, which is predict group membership. So, just kind of looking, at this as the foster youth comparison group and there are six different groups, trajectory groups and, but, you can see really closely quite apparent is that youth do not change performance levels really readily, right. So, they are pretty, it's a pretty flat trend except for their one group, group two. And so this is the multinomial logistic results and so we had, I combined the last two groups because of low samples size, so we basically had give groups, reference group is the worst performing group. We have a below and improving, we have a below, a basic and then proficient advanced and here are the variables that were controlled for in the model and so basically

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

what it's saying is that there was no significant difference between foster youth comparison group. Foster youth statistics was a not predictor. When match closely to the comparison students at baseline. They were performing about the same. What did predict group membership really was ethnicity and disability.

Same thing with foster youth, we are seeing a little bit more fluctuation in Group-A, with the groups in terms of their trajectories and oh, this is just a big muddle, I am not going to go through it all. But, the first couple, the first couple of rows were really looking at these educational risk factors and then they looked at a number of child welfare variables and then also this movement in year-one. And so basically just to take away from this not going into it is that anything in bold is a significant predictor and there is education risk factors, a lot of those were in bold. And that's what's really kind of standing out in this chart. Also looked at cumulative risk and so this is based on mean risk and once again education risk was a sum of five and compared to the lowest performing group, each group, which was performing better on the CST's has significantly less education risks. They also significantly have less year-one changes, okay, and that's kind of the take away from there.

So, what is the finding? First of all the exploring kind of genet or variance within foster youth population and I think it's really important. We don't do that enough and, you know, I understand looking at foster youth as a population as a whole is really important for policy and practice implications and, actually, you know, making sure that we have funding for the group as a whole. But, we also know that not all foster youth perform poorly in school, some of them excel and to understand, you know, even in resilience I mean what makes them different, what do, they look like compared with other kids, okay. So, it's really important, I think we start working on and looking at variance within the population. Overall English and math trajectory is the foster youth comparison students were similarly and it really was these educational risk factors that increased vulnerability for foster youth, same one particularly the foster youth population in itself. Education was prior to entrance are more salient predictors of poor performing trajectories and the lower performing trajectory groups were really defined by increased number of education risks that were present prior to entering foster care as well as school changes in the first year of placement and that becomes, kind of set them up for a downward spiral, okay.

Limitations in this study was really exploratory once again the dataset was limited, administrative data, sample size, some of the sample size was really small and therefore not weighted to the general population and so that was that. We are almost done. So, pilot projects lesson learned. What did we learn? We learned how to not exclude state course that, you know, that, it's just easier to kind of make sure that everybody is onboard and move forward as oppose to trying to go backwards to get their permissions. We really understand that in terms of the political claimant of workload constraints, that's really important and in fact that we are moving statewide, Jordan was talking about that they switched their MOU into a passive as well as instead of an active consent and that's actually was much, very helpful as we moved statewide with getting on the district permissions. We've learned to stay flexible, creative and patient. We had to redesign many parts of our process a couple of times. This study was supposed to only take one

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

year, but, it was two years from the start to finish and once again we thank our funders were being flexible and patient.

One of the things we did not do, because we are so, just focused on how to link these two datasets is we didn't create a formal analysis plan at the beginning and that really actually became very complicated at the end, because by that time no one could agree about what to do. And so, this time around as we moved statewide, that's the first thing that we did, because we already figured out the mechanics, we started with the analysis plan, what do we want to do. In pilot we learned so much from the pilot project and I was so far through the statewide linking effort it's, in terms of the mechanisms it's been so much easier. But, we wouldn't have known had we not started small. And even just in terms of the complexity of the Cal-PASS dataset that, you know, we know the strengths and challenges and what we can do and that's just everything was such a great learning experience.

So moving forward we successfully linked all 58 counties in the State of California to Child Welfare Data. We are actually working with the data right now. We are going to be using weighted examples to make it normative of the child welfare population in the State of California. We plan to follow youth overtime. We are going to actually only do longitudinal analysis for this wave. We are going to look at youth as they move from the high school to post-secondary, we are going to follow them and look at, try to identify potential critical junctures were will their pathways diverge from comparisons students. But, also looking for programs or anything that help facilitate this process for foster youth, one of the things we found at the post-secondary level in the pilot project was financial aid. Foster youth who had financial aid were significantly more likely to persist in community college and that was a big finding. So, we also planned to investigate residential school movement. We have multiple educational child welfare indicators and one of the beauties of the Cal-PASS dataset we do have multiple years of education data.

So our focus is really this change overtime. We are going to try to identify patterns, identify timing of patterns and hopefully it would allow us to make strong policy and practice recommendations and we are really, you know, even though it's just administrative data and, we're limed to what we have, we really want to use the data to try and tell story about the foster youth educational journey, educational journey while they are in care though, even though they have exited the care. One of the strengths of our dataset is really, we can look at outcomes for kids that have exited long time ago, so that's it end notes. Any questions, yeah.

Audience: When you announced the longitudinal analysis you did for your dissertation, what did the classification table look like in terms of the prediction. You said one day you looked at predicting the group assignment? How did it look, how it worked, how do you predict it? I saw the pseudo r-square was pretty well at the point of four, but...

Kristine Frerer: That's a little bit different in terms of the model. And so, yeah just...

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Audience: Just try to interpret it with same way it is for traditional logistic regression as traditional normally regression as, you know, proportion of variability explained by the model, but... Did you find that you were correctly predicting the membership as much of the time?

Kristine Frerer: Yes, as best as we can. In terms of, there was a different, there is a whole process and I want to tell you the whole process, but, we looked at, I looked at by variant analysis as well and I am just looking at the differences between the group and the significant difference with FTES and then kind of looking at the model and they were pretty similar and in terms of the, if the model itself holds that's kind of a longer discussion and what I ended up doing was stepping it out using group based analysis only to assign the groups and then using multinomial logistic regression in a different platform to look at the predictors and that's because there is a glitch in terms of the algorithm of how the group based analysis is done and I am quite sure that's, which you are talking about kind of, okay. Sort of, but, yeah it was exploratory, when I am planning I am actually recreating it, which a much bigger sample size. It will be much more of a confirmatory approach and so I will be able to see if that is true with a large sample size, same years of data though, so...

Jordan Horowitz: I would like to point out in case you didn't notice. In the program it says Chris Frerer MA since we submitted the proposal, its Dr. Frerer, so congratulations Chris.

Kristine Frerer: Thanks to Cal-PASS pilot project.

Jordan Horowitz: I think there will be people in the room who can appreciate that journey. I am going to talk about the second, the other side of the effort that The Stewart Foundation has funded for this Ready to Succeed Effort for us and that is the development of a web based tool that folks who run foster youth programs and support foster youth support programs on post-secondary campuses, a way they can access information about their foster youth on the campuses both while they are in campus and before the, what happened to them before they got there using a web based tool and I will get into this. I will start with a discussion of kind of who we are, I will talk about the projects and then I will actually demonstrate this web based tool for you. So, the Institute for Evidence-Based Change, our mission as it says to help student succeed in school and beyond by empowering organizations to use data to improve practice and outcomes Cal-PASS is, as I mentioned is voluntary data system.

We were all employees as we were hired of the community college district where this started. Overtime with efforts like this and other efforts we really outgrew the community college district. We were doing a lot of work outside of California. For example, we are running Hawaii's Statewide Inter-segmental Educational Data System for them and so we formed our own 501C3 and pulled everything out from the community college district with their blessing. I mean we really were overburdening their business office. Our private funding accounted for 90 percent of all the private funding coming through the district at that point and they kind of couldn't handle it and

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

so we formed the institute, the chance throughout the district is on our board and so we have a very good relationship spill. The Cal-PASS remains one of our flagship efforts.

So Cal-PASS links primary, secondary and post-secondary institutions on regional basis. We track students from one segment to the next. Our focus really is on the transition issues. We bring faculty together locally across the segments to talk about issues because we can look at course level tracking such as why do 43 percent of our students who finish Algebra two in high school end up having to retake Algebra two when they get to college or university. We have over 7,000 K-12 schools, community colleges and universities submitting their data and our dataset now includes about 480 million student records and we have those image in public and private funding. We do create a pseudo ID number, which is our primary number for doing our matches across the segments. So, we take a series of encrypted data elements and use that to create this pseudo ID number locally our match with the pseudo ID number locally being saved from a high school to their local community college districts and on to university and we find about 95 percent to 98 percent match with this pseudo ID number, statewide well it's about 92 percent to 95 percent, which is close enough for the kind of work that we engage in.

As Chris mentioned and referred to, we have a number of uploads that schools give us. We have the student characteristic student file, we have courses, we have awards we have their high school exit exam data and their California State Test data. Those last two are the most difficult for us to get primarily because they are given to the districts, the K-12 districts on a disc each year, which means to do their upload they need to figure out, which secretary's drawer the disc was last left in. So, pretty much that, that's the challenge there. Our student file includes a series of demographic and status information, gender, ethnicity, date of birth, English learner status, disability status, which includes special Ed through physical disabilities as well as their free and reduced lunch status, which is a challenge because free and reduced lunch status are not education data, these are data held by the U. S. Department of Agriculture. We have a couple of districts that will not give us free and reduced lunch status. One of them actually had, been involved in a research project six years ago now and the project was published and the district actually got a cease-and-desist letter from the U. S. Department of Agriculture saying these are not your data, these are our data and you cannot use these data without permission from us and they got very, you know, very cautious.

So, and we have a couple of districts who don't give us their free and reduced lunch status. Everything is linked to county district and school and everything is also linked to term. So, if a student has a course record from a community college say and their high schooling for the same term, we know we are looking at a dual involvement situation if we had different course in uploads from multiple districts we know students have moved during the term or through the school year. We have every course the student takes and their grade in that course, so we have transcript level data. We have all the California State Test Scores for every year, every test and high school exit exam scores. Students take the high school exit exam in California for the first time in their sophomore in high school. There is English and math, if they fail one of those or both of those sections they can retake that section up to five times before they get their high school, before their time

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

to graduate to get their high school diploma. So, because everything is into term, we can also track high school exit exam attempts and successes over time.

The award file we have diplomas, degrees and certificates, these are community college certificates in career or a vocational certificates. These aren't, you know, good student certificates from the local high school or attendance certificates and so on. Speaking of attendance that is, because this is transcript information, that is a gap in our dataset. We do not get attendance information. We also do not get discipline information. However, we have hundreds of blank fields and schools can upload that information to our system if they want to. We also use these blank fields and this is what we, I will talk about as I get into the project. For projects or schools or districts to flag students who are in certain programs. We use, the fill out an Excel Spreadsheet, the first six or seven columns are the data elements we need to create the pseudo ID number and then after that they can use columns to upload whatever they want with a coding structure that we agree on. So, for example, in the foster youth for this we have the post-secondary institutions to flag the foster youth in the Cal-PASS dataset, they give us the first six data elements that we need and then the next column is this foster youth, simply a foster youth on their campus. When I said foster youth, I am really talking about former foster youth when we take post-secondary, okay, primarily.

So, is this is a foster youth on campus they know off, is this a foster youth in their program, so they know the service, the student has received or is this a foster youth on campus who they know who is receiving services from some other program. So, we have that designation is that when a foster youth project director on campus goes to use this tool, they will be able to compare those three groups as well as to the general population on their campus, you know, is there someone in my program as a group or it's been so on, we are fully FERPA compliant as I mentioned all data are anonymous, all personal identifiers are either removed or encrypted. We've worked with most of district information systems that the vendors that sell to the student, to the districts and so that Banner, Eagle, Aeries, SASSY *[phonetic]* *[00:56:45]* PeopleSoft. They actually have Cal-PASS extracts built into their systems for us now and each year we update those for them. So, a K-12 district that's running a package system....

That's running a package system can just click on Cal-PASS extract and it will run this whole thing and pull the data encrypted, actually it does a little bit of validation before the encryption. It's not real validation, just kind of make sure the date fields updates in, that there is nothing odd and a report goes back to the person at the district or a post-secondary institution saying you might want to look at this. But, it doesn't recruit them from uploading anyway. But, for example, we had a community college district in Northern California that's '98 percent wide and every student was coded African-American, you know, so we were able to tell them this looks kind of strange and that they went and corrected and realized there was a region there, a coding system. So, the data are encrypted and then uploaded. Our servers under the secured level for datacenter control access, I visited for the first time actually just last year and it's pretty intimidating. They use IRIS and full palm scans with doors and armed guards everywhere plus everything is up of the floor, you know, chilled and, you know, it's

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

really quite a facility that's unmarked and you don't know what this huge building is, you know, and with all these office buildings.

Kristine Frerer: Is it the one at Sacramento?

Jordan Horowitz: It's outside Sacramento if I tell you where, I'd have to shoot you *[overlapping conversation] [00:58:34]...*

Kristine Frerer: You have to shoot me yeah, that's right.

Jordan Horowitz: We, and there is this, you know, they said that this, it's like basic control of balloon launches in the late '60's, you know, there is screens of these guys. It's an, nobody is smoking. You know, I was doing a presentation about our work to, at a tech high school and I was talking about this and I said, you know, it's like, you know, actually hold a chrono-type machine called, you know, when the moon launches I had this blank stares, nobody knew it. I felt very old. We are very transparent about our security. All of our security documentation is available on the Cal-PASS website. So all, of our data element dictionary, so if you want to see exactly what we collect, how it's coded, it's all at calpass.org. So, this project as, so you can tell we are very thankful to The Stewart Foundation. They...

Kristine Frerer: We love The Stewart Foundation.

Jordan Horowitz: They really have been great. They under, our project officers Michele Franzwa, her boss Terry Cook really understood the nature of this work and we are flexible enough to allow it to happen. We piloted this in three regions, Sacramento, Los Angeles and Fresno. The schools identify the foster youth and former foster youth at their institutions and in programs with the custom fields as I mentioned. I will show you how that happens. We developed web based queries for direct access to this 480 million record dataset. As I mentioned we have about 75 percent of the K-12 districts with high schools in them submitting data to us. But, that represents a little over 90 percent of the students in California because we have the largest districts submitting data. So, LA Unified, San Diego Unified, Sacramento, Oakland, San Francisco and Stockton, they are all in the Cal-PASS population of districts. In fact I don't know how familiar you are with education data LA Unified has stopped data to the National Student Clearinghouse because of the cause of doing it and are relying on us to look at their student transitions for program improvement purposes. They figured students who are going outside of California to universities are succeeding, they are not really, and since the students who are staying in California that tend to be their lower performers, so we have that data, so that's all they need.

Audience: Did you say you had Oakland Unified?

Jordan Horowitz: Excuse me?

Audience: Did you say you had Oakland Unified?

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Jordan Horowitz: Oakland Unified is in, yeah. The fact that we have LA Unified kind of gets rid of all the security and privacy issues from anywhere else in the state. I was at West-Ed for 16 years. We had a contract to evaluate LA Unified's Health and Prevention Programs for them and I couldn't get the data. So, once LA Unified wasn't giving us their data, anyone else has any compliance and then the ideas that once we piloted in these regions and the regions were selected because the, similar reasons to the data linking effort. These were post-secondary institutions that are already involved with projects for their former foster youth funded by The Stewart Foundation. So, we included a consortium in Sacramento of Sacramento State University, University of California, Davis and local community college district, Los Angeles City College and Fresno State University. Those were the post-secondary institutions involved in the pilot.

The goals would establish a custom system to provide reports from our database to projects Stakeholders College Pathways is in the name of this projects that Stewart funded up the institutions. To support evidence-based decision making, continuous improvement efforts for education of foster youth as the program coordinator at Fresno State said, you know, this is great because in addition when I mentioned all these other data elements, they wanted to put in type of service and dosage for youth that were provided to each of the foster youth on their campus. So, he said this is great. For the first time I will be able to know if my Summer Bridge Program is making a difference in educational outcomes, because if it isn't I could use that money for services and efforts that I can see are making a difference in educational outcomes for these students on my campus. And also to ensure that foster youth had the same opportunities to succeed in education as their peers, there is, as you can tell a strong social justice spent to our work and it comes through in this effort as well.

So we identified the pilot sites, then we had the input from our funders regarding the data elements to track and they wanted to know the number of students being served, the number of credits involved in and completed each semester, grade point average, degree and certificate completion for persistence the number students who begin each semester, how many year involved with the end of the semester and how many returned the following semester. A, input secondary subscriptions called retention. Unlike retaining a student in elementary school and high school, which you don't want to do, in California's post-secondary system, retention is the proportion of students who involve in a course, who are retained through the semester and complete the course. So, retention is a positive aspect of the data in post-secondary. For the four year institutions, the number of students who complete upper division courses is of course who that are, upper division. Now for the community colleges, the number of students who complete transfer level courses, so these are courses that are for credit and can be transferred to the four year institutions if the student transitions on fund of the two year community college and the number of students receiving financial aid. Then we have input from the pilot sites about what they wanted to know. So, what are the demographics of the foster youth in their program and at their institution, gender, ethnicity, the high school from, which they graduated, the number of high schools attended, prior post-secondary enrollments, age at

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

enrollment, high school outcome data for their foster youth, the highest math course completed successfully in high school.

Successful as C are better, by our terms D is passing, but, not considered to success. Highest English course, their 11th grade state test scores, their, because math course is not linked to grade, the highest math course they completed. We also wanted to know the initial math and English placement where they in medium courses or college level courses. We also wanted to know the units attended by term and cumulative persistence, the next term to the next year and continued persistence, the movement to transfer level courses from medium and below transfer level and degrees institute gets awarded and then time to degree. So, there was a lot of information folks wanted out of the system. We had to develop a method for flagging the foster youth in the program and on campus and that's I described our custom flagging procedure. It's important to note though that these flags are not universal flags, so if the math department chair at a university wanted to compare foster youth, non-foster youth in the math department, they do not have access to that.

That permission is granted by the foster youth project coordinator on the campuses. So, if somebody in the math department do this comparison and try to access this tool, they would not have access and I will show you how we control access. They wanted that, then we have to send an email to the project coordinator saying this person wants access to foster youth data, he said okay for them to have it. I like to say we want to target foster youth for services, we don't want to pay the target on their backs. Oh, so what are the data, we have the data elements required to flag them in the database and this is what we need to create our pseudo ID number and there is the program status and then these project coordinators wanting to include the five data elements that were added to the federal financial aid forms two years ago and that is, has this student been in foster care since age 13, are they, or have they been dependent over the core, are they have been emancipated minor, where they under legal guardianship while in foster care and are they homeless. Yeah.

Audience: Did they have to be in the same removal episode since they were 13 going forward or could they have is that basically like the entry into care measure?

Jordan Horowitz: That's an entry into care measure, I believe. I would have to check on that. If that's entry since age 13 or if they were, they might, could have entered before 13. I am not sure how the Feds define for that for the financial aid form.

Kristine Frerer: Yeah since, makes it seem like a long-term foster care placement of some sort.

Jordan Horowitz: Yeah, yeah. And then they also want to get the student receiving financial aid, housing assistance and their employment status. California has a separate program for housing foster youth, providing housing assistance, I mean including providing housing over summer and holiday breaks where foster youth might not have a place to go home to or former foster youth.

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Audience: Is that something that the college provides or the state?

Jordan Horowitz: It's provided by the college with fund, with state funding. It's a IOP program that's through, it depends on the campus, some it's through the financial aid office, some it's through the housing office, some it's to the foster youth program of it barrage.

Kristine Frerer: Transitional living and transition...

Jordan Horowitz: Transitional living programs and so on. And finally after we know what folks want out of the system, how to identify the foster youth in the system, how to link all this other non-core academic data to the student's core academic record, we could develop our SMART Tool, SMART stands for Standardized Metrics for Analysis Reporting and Tracking, one of our better acronyms.

Audience: It's a pretty good one.

Jordan Horowitz: Yeah. So, I am going to demonstrate the web based reporting system to you. It's, you will see its question driven. The first tool that we developed was for a project called ASPIRE Project from the Irvine Foundation. It was student's supports on community colleges and we developed this tool, you will see the tool and we have all these data on it. So, we just dragged that into columns and rows that creates these tables for you and I had all the funders in, I had representatives from Sacramento assembly people. We had a table setup for each college with computers and their teams and it was like grade, you know, go at it, show us, you know, be interesting and they are like, oh, look, 47 percent of our students in the program are female. Well this isn't exactly what we are hoping for and we realize and I should have known this, because when I was director of evaluation research at West-Ed, I would tell my team, don't expect faculties to be the researchers, you are the researchers and we realized we have to find out what they want to know and make this question driven and I will show you how that ended up so you a list of questions and you click on the question and what's your next question about the students in your program and then you can drill down if you need to, you know, what about Hispanic students versus Non-Hispanic students and so on.

It has been updated annually with the new updates every, by the way we get our data every October we open a window and the data for the prior academic year get uploaded by all the districts and post-secondary institutions. So, we are about to open the window for the 2010-011 academic year that ended last year. And we need, it needs to be able to support both the reporting needs of these projects as well as their own internal uses. Okay, so let's get to the demonstration. Let's see if, okay. We are actually accessing this 480 millionth record dataset. It happens pretty fast, it's a little slower here because of the band, wireless bandwidth. So, it'd be even faster if you are at your college with, you know, a line. I am going to make the size of this smallest that we can get it all on one screen, okay. So, this is our list of questions and the question that it opened with was

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

“What was the last high school English course taken by our former foster youth?” You can see we have a list of, the questions were created with the project folks.

So, we can go to what are the demographics of the youth in your program, click on the question and here are the demographics gender by ethnicity. These are our column measures and these are our row measures. So, the first thing I want to do is I am going to clean this up a bit, let's get rid of the unknowns, unknown genders and let's go with these. Let's go with African-American, Native American, Asian, Black, Hispanic and White. We will stick with these kind of basic categories so that it simplifies what we are looking at. If you are a graph person as opposed to a table person, you can click on chart. Oh, this is going to BNR Chart because there is so few of these and so many these. But, this is, you can easily get a chart instead of a grid. So let's say I am interested in how these students performed in, let's go back to the English question, how these students performed in these English courses.

On my measures I can select the success rate, which will show me the proportion who earned C are better in these courses and now, and these numbers are pretty low because we are working with one college that gave us permission to use their data, but, you can see the success rate in English courses by students. If I want to, I can also bring in, you know, if I want to look at this by gender, I can take my gender variable and just drag it over here and it will add gender to my columns. So, now I can see my gender breakdown for the high school English courses that my students took. Now these high school English courses are by name from any district that these students might have come into. What we are doing is when we roll this out and it will be rolled out in next month actually for full use. We are going back to our course descriptions that we get from the districts and we will be collapsing these so that it's easier to see 9th grade, tenth grade or 11th grade, 12th grade English and then other specific English courses like Mythology or so on.

Audience: So, if one were interested in, let's say in Kristine's dissertation and you are going to export this into that file and...

Jordan Horowitz: We do not cut data files for the researchers to use. If a researcher wants to access the Cal-PASS dataset, they work with our analysts to do their research together because of the way our MOU's the ownership has setup with districts universities, we don't cut datasets and hand them off.

Audience: So, where are these, as these data rely somewhere in a total format again is it related more or less as a row by column, you know...?

Jordan Horowitz: Oh, no, it's a relational database, so we have tables, relational database on our server. Our guys use Sequel to do this programming...

Kristine Frerer: The difference between that, that way, the pilot...

Jordan Horowitz: That perspective is not taken from this is a completely separate...

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Kristine Frerer: But, our data is actually under the providence of the UC Berkeley that we are the purveyor of building datasets. So, Barbra Dell is actually their PI on the project and so permission to use this link data set, I am have to go through UC Berkeley once we've linked the data. So, Cal-PASS that was just the way our IRB worked out.

Jordan Horowitz: This is not linked to the Child Welfare Services dataset, that's why we need the institutions to engage in the flagging process try to identify who the foster youth are on their campus. This is slowly the Cal-PASS educational dataset.

Kristine Frerer: In a so monitoring use the link, education Child Welfare Data, that's how the UCB then they would through UCB protocol. It's been available to do that, so...

Jordan Horowitz: For example we are working on a project in LA right now because LA County has a unified data system that holds data from probation and child welfare health services TANF or health outcomes or justice outcomes and they are doing a lot of, they are doing some research right now with Dennis Cohen from PEN on, for the youth who are in probation, youth who are in foster care and what their post service use of all these systems looks like and they didn't have education. So, we took, they hint us and we took our education data and linked that into all these other data, but, we are involved in doing the analysis with them. We could not just say, okay, here are the data, you know, go ahead and you can have it and do what you want with it or be certain. One of our researchers is on the research team that is doing the research. So, we don't create datasets and handle off for others to use.

This tool is for the use of foster youth project coordinators on college and university campuses in California, that's what this is. We have other tools, we have a tool that supports the accreditation requirements for the community colleges. There are certain questions they need to answer about their students when they go through accreditation. Instead of having to ask their institutional researcher to run this for them and make sure they are getting, you know, the data they need. We created a tool that answers the questions that they need to answer for their accreditation every four years that gets all of this for them. So, we have various tools and this tool is developed for foster youth project coordinators on post-secondary campuses to be able to know what happens to their students. So, this is not meant to be a research tool.

Kristine Frerer: The universe of your data is all students from participating schools districts.

Jordan Horowitz: Correct, correct. So, when somebody wants to ask a question like this, about what happened to the height of the students in high school before they got to me? We have data for most of the students in California that go back five to 11 years. Some of the Cal State University Campuses got a little competitive, which benefited us. We actually had one said what is the most years of DB you had from any campus and we said well, 11 and they said we will give you 12. Oh, great. So, that's what this is meant to be is to write data to project coordinators on the campuses. These data elements are all

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

things that they wanted to be able to drag and drop to cut and slice and dice this population by, you know, here, is employment status.

We can bring in employment status, sorry employment status I guess is, since it didn't show up it's not in this version of the tool yet. So, but, these will all be in there whether or not they receive mentoring program statuses in prog, is as appear is going to be in program and it will have and these are, our analysts are still developing this tool. But, that's where it will say on my campus, on my campus receiving services, on my campus in my program and I will be able to drag and drop that and be able to just show up this three columns, they will be able to compare those three populations on any of these questions and these data elements.

Kristine Frerer: So, it's really, it's used with a history of post replacement necessarily, I always see a couple of measures regarding exits and so that's kind of interesting. I am emancipated minor, a legal guardian, but, nothing...

Jordan Horowitz: Right. And those are the facts, emancipated minor whether or not the guardian are the five fab side on this. And that's what are showing up here. So, foster care since age 13, emancipated minor, Chaffey EOPS are financial aid types that are given nationally and in California. Ward of the court is another one of the FAFSA data elements so they will be able to do, compare students who check that on their financial aid reform with those who did not and so on. So, that's this tool, yeah.

Audience: So, there is nearly readied question. So, how long you have been working on this tool and what's your infrastructure, how many staff you are working on this?

Jordan Horowitz: We have one analyst creating this tool and she has been working for about 15 months and we will be rolling that next month. I have to say the biggest obstacle that we are dealing with was the campuses ability to identify foster youth. It is a real obstacle. We are convening an advisory group this fall to address that issue. The community college system in California has a new data element on their application form, that's not really the case. They are requiring the colleges to include the data element on their application form starting next year. It's been voluntary for two years and then for the 2012/13 academic year it's mandatory. California's 94 community college districts and 112 community college campuses do not have a unified application form. There are 97 different placement tests in use and cut off points among these community colleges there is no common placement test, nor cut off points for placement into remedial math and English. So, it's a very, dispersed local control issue system. So, we are looking at the ability to use the foster youth data element in the application process with the understanding that that also with self-report and we may not be capturing everybody. There are other ways though to capture students and there are—as Kris alluded to—of course, issues using financial aid form data for anything other than financial aid. So, we need to develop agreements around using that information for this tool, so yeah.

Session 7.02 – Collaborating To Improve Educational Outcomes for Foster Youth: Linking Anonymous Education and Child Welfare Data Sets in California

Audience: A question, do you have any information about IEP statuses for the elementary school?

Jordan Horowitz: We do not have IEP information. If a student has the special aid data field as positive or checked, we know they went through IEP because they had to. If they have any kind of disability they had to have an IEP, but, we don't have the IEP data.

Kristine Frerer: There is a, I am sorry but, there are, in the statewide we actually, there is a variable I forgot, which follows in or it's just as yes or not if that...

Audience: Is sort of what they have in between I guess and there is anyway, with things like that have you, you know, have to find the data or are you not sure what the research question is yet?

Jordan Horowitz: We, the data elements are defined by the submitting institutions. So, every year we have, we do all of this work with three researchers and three programmer analysts and that includes our out of state work that includes all of our work. We are a very tight efficient organization. But, those folks when each year we update all this information from all the districts also, so if they have new courses with new California basic education dataset numbers, because the California Government gives course numbers every course, new names. We get all that information and that's what we spend, one of our analyst spends the summer dealing is updating, getting everything set for the full submission. So, any of these data elements are defined either by the submitting institutions or we got a group and say okay, how should this be defined, that is, how should we operationalize this that make sense to you.

Kristine Frerer: We just, just jump on that just in terms of the course work that we looked at for the pilot projects and we are going to be doing A through Z through the statewide is that we had to query because each district and each school calls of course something different, but, it all goes to an element. We had a query over 45,000 different English and math, different titles just to come up with an English category and an math category. It was just insane. It took weeks to do, so...

Jordan Horowitz: Oh, right. He has left it. This is another reason why we need to have our researchers involved in research projects because less of, you know the new ounces of the dataset and how to get that what you need, it's very easy to go longer...