SOCIAL SECURITY ADMINISTRATION

OCCUPATIONAL INFORMATION DEVELOPMENT

ADVISORY PANEL QUARTERLY MEETING

DECEMBER 8, 2010

RADISSON PLAZA LORD BALTIMORE HOTEL

BALTIMORE, MARYLAND

* * * * *

DR. MARY BARROS-BAILEY

CHAIR

S R C REPORTERS
(301)645-2677

M E M B E R S

MARY BARROS-BAILEY, Ph.D., CHAIR

ROBERT T. FRASER, Ph.D.

SHANAN GWALTNEY GIBSON, Ph.D.

THOMAS A. HARDY, J.D.

JANINE S. HOLLOMAN

H. ALLAN HUNT, Ph.D.

SYLVIA E. KARMAN

DEBORAH E. LECHNER

ABIGAIL T. PANTER, Ph.D.

DAVID J. SCHRETLEN, M.D.

C O N T E N T S

C O N T E N T S (CON'T)

P R O C E E D I N G S

MS. TIDWELL-PETERS: Good morning. My name is Debra Tidwell-Peters, and I'm the Designated Federal Officer for the Occupational Information Development Advisory Panel. Welcome to the December 2010 quarterly Panel meeting.

I will now turn the meeting over to Dr. Mary Barros-Bailey, the Panel Chair. Mary.

DR. BARROS-BAILEY: Thank you, Debra.

Good morning, everybody, this chilly December morning. I would like to thank you all for your attendance, live or telephonically, to the second quarterly meeting of the OIDAP for this fiscal year. It's nice to be in Baltimore. We have been all over the country, and it's good to be here and see a lot of the faces from Social Security among the audience.

If you hear a voice from on high, that is Dr. Gunnar Andersson, one of our panel members who could not be here live, but will be joining us telephonically. And if you are joining us telephonically, you can go to our web site, which is

Social Security "dot" gov, forward "slash" OIDAP, to follow along in terms of our agenda. You can look at agendas from past meetings, materials, technical working papers, formal correspondence. The two reports we have issued are up on line. One is a September report from last year. The National Academy of Science, our review of the -- their review of the O*Net.

And shortly, there will be a couple of other pieces of information up on the web site. One of them will be a summary of the public comment that we had for nine months, and we delivered that report to the Commissioner yesterday. Another one will be the recommendation that we finalized in November. The general recommendation number eight, also known as the Occupational Information System Planning Recommendation.

As we indicate at the start of each meeting -- and I think it's important to reiterate each time -- our charter is to provide Social Security Administration with independent advice and recommendations for the development of an

Occupational Information System to replace the Dictionary of Occupational Titles in disability adjudication.

Our task is not to develop the OIS itself. As our name implies, it is to provide advisory recommendations to Social Security as it develops the OIS. Again, it's to provide advice and recommendations to the Social Security as it develops the OIS.

This meeting is particularly special. December 9th, 2008 -- so 2 years ago tomorrow -- the Panel was chartered. And we are going to be honored this morning to have Commissioner Astrue with us to follow our Project Director's report, and also Deb Lechner's presentation. We're going to be a little bit flexible on the agenda this morning. It's my understanding he may be coming a little bit early, because he might have to leave a little early. So we will go with the flow with that.

The proposed rule for the revised medical criteria for evaluation of medical disorders in terms of the Federal Register notice had two

projects that SSA is currently undertaking that might have some significance to that. One of them is our project. Another one was the project of the National Institutes of Health with SSA. I had requested at the last Panel meeting that we hear what that other project was about. And we are going to be having a presentation from that project this afternoon. I think it's important to understand that sometimes projects are different and don't go to the same purpose. So it will be interesting to see what that project entails.

So after the break we will go into the routine items of our agenda, including Panel discussion and deliberations. We're going to have the opportunity for public comment. We're going to have the administrative meeting consisting of the review of the last teleconference Minutes, and discussion of dates for the 2011 meetings, as well as the March agenda.

Now, I would like to past the meeting on to our Project Director for the Social Security Administration development of the OIS, Sylvia

Karman.

MS. KARMAN: Good morning.

DR. BARROS-BAILEY: Good morning.

MS. KARMAN: All right. Just as an overview, we are going to cover just a few things that have transpired since September. One of them was the creation of an office or unit dedicated to this project. I also will provide some status on our activities, and what some of our next steps are in the next few months.

So as of October 1st, the Office of Program Development and Research in Social Security established a new unit dedicated to OIS development. The Occupational Information Development Project Team has become the Office of Vocational Resources Development. So you will hear us referring to OVRD, perhaps, a little more frequently.

OVRD is charged with directing and conducting research and development of SSA's Occupational Information System. It consists of two branches. One dedicated to the research design and development, scientific standards, testing, data

collection, and data analysis. When I say data collection I don't mean that literally we're going to have all of the individuals that might be necessary to gather the data, but rather that we have oversight of that activity when it takes place. The other branch is dedicated to program integration to ensure that the OIS is developed in a way that meets SSA's legal program and operational needs.

Provide a little bit of update on our Occupational Medical Vocational study. Again, the objective of that study -- we presented the early results of -- on the initial review in September. It's basically to identify the primary occupational functional and vocational characteristics of Title II and Title 16 disability applicants whose claims were approved or denied at steps four or five of the sequential evaluation process. And those decisions having been made at either the initial or hearings level.

We have completed the reviews of initial level cases at the end of July, which, of course, you all are already aware of. We presented those

highlights to the Panel late in August. We call it a September meeting, since it was almost September when we met.

Also, we have completed the quality reviews of the initial level cases. We're developing -- we have developed a hearings level data collection instrument a bit different from the initial level data collection instrument, and the protocol methods are a little bit different, because we had to ask reviewers questions in a different way to get at some of the information as it's presented in the appellate files. We are going to begin pretesting of that data collection instrument later in December, and we hope to begin reviewing those cases in January.

Another activity that we have had under way -- some of you have been aware of for a while -- is we have been examining or investigating the types of occupational classification systems that exist internationally. What we have done in beginning to develop our overall comprehensive plan is to recognize that it will be very helpful for us if we

also combine into that study information about any national occupational classification system. Not so much because we are needing data, but because the actual methods of this classification may be of value to us, certainly in terms of the design decisions that may have gone into them.

So what we are planning to do is to combine the international investigation with work that SSA has conducted over the last few years either on its own or with a contractor to examine other classification systems, and as well we're going to look at other classification systems that exist within the federal government, such as those in the Department of Defense, even in the Office of Personnel Management that may be very helpful to us to determine, based on the purposes of those classification systems. They may very well have had to make design decisions that may be of value to the types of design decisions that we need to make. And we also want to explore what methods they have used.

So another area -- sorry, I'm having trouble with my computer here.

All right. So one of the things that I think would be very interesting for us, as Mary mentioned, we will be hearing from the researchers from the National Institutes of Health, as well as Boston University. They are working with the Office of Disability Programs at Social Security; and I think that that presentation will be very interesting for us.

Our staff has been meeting with the Office of Disability Programs regarding the work that NIH and Boston University are doing. Largely, we are very interested in their methodology at this stage. So we have also begun conversations with NIH and Boston University. One of the things that I think is going to be of interest for us, our collaboration at this point consist of sharing information regarding the methods that NIH and BU applied to identify the content from various sources and to develop items for their questionnaires, which they will be describing.

The projects are separate within Social Security, and they are not dependent on each other.

Also, they are quite different in scope. ODP's project, or the Office of Disability Programs' project with NIH and Boston University is exploratory, investigating ways in which we can obtain in a more effective manner better information from claimants and medical sources regarding the claimant's function. And obviously, as you all know, the OIS development has more immediate or applied mission and focus that is aimed at replacing a data source that SSA now uses.

However, SSA is aware of the need to provide the adjudicator with better functional information about the claimant on one hand, and better information about the advanced support on the other. Adjudicators must be able to associate the information we obtain about the claimant's function with information about work. Working on ways to provide adjudicators with better information at both the front end and back end is the subject matter nexus of these two projects. And as I mentioned before, OVRD is interested in the methods of ODP's exploratory work within NIH.

We also brought back together the OISD development workgroup. The workgroup is -- at this point has been briefed on the status of the project, the new organizational layout, and also discussing future activities of the workgroup, which includes establishing a formal charter. So as the workgroup is meeting in December, they will be working on that.

The other aspect or other thing that our staff has been working on since September is the OIS Research and Development Plan. And we have completed an initial draft of that Plan. When I say that it's comprehensive, I'm going to describe some of the components that we're anticipating having in that Plan.

We -- so, actually, the Plan right now is currently in review in our office, and will go into our associate commissioner as soon as I complete my review of it. We also continue to conduct work that was underway before we began the Plan.

Some examples of components that we intend to include in the OIS R & D Plan are certainly scope

of the research and development for OIS, as well a business process for all of the OIS activities. And an entire section naturally devoted to research design, which should include for each of the major activities, goals, and objectives of those key activities, research questions, and evaluation criteria.

The actual tasks and activities that would need to be accomplished in order to get at those research questions and by what methods we anticipate doing that, and perhaps, in some cases identifying, perhaps, more than one series of methods or options that need to be considered.

Also, any dependencies that we are aware of at this time, relevant sources and literature that we have examined or believe that we need to examine. Also, what resources we believe are needed in terms of is there certain expertise that might be needed; or even, you know, down to the level of what kind of software might be, you know, important in a particular investigation.

Also, I think it's important for us to

identify any risks or threats to validity that's at issue; and we're also planning to include communication strategy, time lines, information about budget, and other things that are necessary for an overall plan. So we're looking forward to being able to produce a draft of that.

We're also working right now on the content model and we are drafting -- sorry, I'm really having trouble here with my --

So we are drafting methodology. At this stage we are in the preliminary development phase under our business process. We are engaged in information gathering and laying out a conceptual plan for the methodology to conduct content model development.

We are also in the phase of initial consultation, and that basically involves a discussion within Social Security on certain -- among members of my team; and, you know, with others who may have information that can be very helpful in helping us develop the study design, which is the second phase under the business process. So at this

stage we do have research questions developed. We certainly have identified the goals. We are looking forward to completing that, and putting more meat on the bones of the methodology there.

Another thing that we are aware of is the need to, you know, be sure that when we're setting up the rules and ways in which we determine what information is included in the content model for testing purposes that we have a way of checking back to see -- you know, in terms of inter regalia agreements, whether people using the same rules come up with the same types of information or types of data element. And also to look at areas of disagreement, and why, in fact, there may have been disagreement.

So we have really heard a number of the questions and comments that the Panel gave us and others as well about the content model. So again, something we're looking forward to producing a draft of shortly.

Another area that we are working on is -- and this -- we're working on a number of standards.

One involves legal standards. Another screen here I'm going to get to shortly about scientific standards.

One of the major activities that our research design identifies is OIS requirements. And under that, one of the first activities will be to establish legal standards, and to articulate what they are. And in doing so also examine scientific standards. Right now OVRD is working with Social Security's Office of General Counsel. Plan to consult with them next week. In fact, begin our consultation with them on legal framework that is needed to support the OIS. Also, initial research that has been -- we are providing them with initial investigations that we have done -- what we think are key legal issues and discuss with them what things they believe we should be focused on given the scientific standards.

We also have underway right now what we're calling a BPA, a blanket purchase agreement. It's basically a contract with ICF international. And I'm just going to say this because we -- our team

frequently confuses the -- people confuse it when they hear it. There is the -- you will be hearing later from NIH and Boston University about the ICF, which is the international classification function. ICF International, however, is a contractor, and they are working with us on this BPA to establish for us a business process for recruiting, training, and certifying job analysts.

We have awarded that contract in September, and ICF is now conducting focus groups and literature reviews to identify and benchmark a variety of job analyses methodologies across various disciplines. So what is accomplished or done currently in the area of voc rehab, perhaps, insurance companies, perhaps -- you know, what else is being done in terms of job analyses in the industrial occupational psychology field. So they will be using that information to document the current trends also in terms of recruitment, training, and certification of these individuals who conduct job analyses.

We plan to have a draft report on the

training and certification recruitment due to us in April of 2011; and a draft report on job analyses methodologies, which will be due to us in June of 2011.

We are also, as I mentioned earlier, part of our plan is a communications strategy. So we are working on a detailed communications plan that will be in the overall comprehensive plan. We're working with a number of offices within Social Security to define and identify future web based requirements for us so that we can establish some online communities with a variety of professional communities that our stakeholders have interest in the work we're doing, as well as researchers. And also just to have a way for us to communicate with users.

So we're taking steps to explore the requirements of this type of a web based approach, you know, interactive, and also allows for Panel and SSA staff to have placed -- post documents and information for each other.

Here is the screen on the scientific

standards. We have developed a method for locating the relevant scientific standards that can be applied to various OIS project activities with the primary goal of establishing requirements for the Occupational Information System that would lead to scientific defensibility. We want to develop prototype instruments that may be very useful to the collection of the occupational information.

Also, reviewing the OIS research and design plan that's underway to identify both major and lesser tasks that would require scientific guidance and try to connect when, in fact, those tasks might be requiring certain types of standards. So we want to be able to identify what those standards are as a part of the criteria.

The next step that we have in 2011, as I mentioned, is to complete the OIS research and design plan, to develop the job analyst recruitment training and certification plans; and we want to be able to prepare and implement a study design for content model.

We're also examining the methods for other

OI -- Occupational Information System, as I explained earlier. And we intend to follow-up with the Census Bureau. This is something that was left over from mid-summer when Dr. Allan Hunt and I, and several staff went to meet with the Census Bureau to examine the type of data that they collect in the American Community Survey. To see to what extent that information people report about their work might be something that we could use to help us identify certainly kinds of jobs in the economy. That's it.

DR. BARROS-BAILEY: Thank you, Sylvia.

I would like to open it up to the Panel to see if there are any questions. Shanan.

DR. GIBSON: I don't have a question. I do have a comment. I just want to say that I was gratified to hear, actually, about the development of the legal standards working paper. I think for a long time we have struggled with the question, what are the legal program and technical issues that have to be identified at the outset, and that has to also drive the project that you discussed. So I'm very

gratified to see that included and know that progress has happened there. I look forward to seeing it.

DR. BARROS-BAILEY: Okay. Any other questions? Any other comments?

Sylvia, thank you for a very comprehensive report. I appreciate that.

Next, I would like to ask the Chair of Job Analyst Ad Hoc Group to present on their findings.

Just to give a little bit of background, as many people know who may have been following this Panel, we're a Panel that is very versed in many ways; and one of the things that we learned very early on is that we might be using the same word to convey different concepts. And one of the things about the Job Analyst Ad Hoc Group is that it's composed of three members. One is a physical therapist, one is an industrial occupational psychologist, and the other one is voc rehab background, rehab psychology background; all who have experience in job analysis in terms of different methodologies.

So I asked the group to take on an experiential exercise in terms of actually doing a job analysis to see what were some lessons learned that might be helpful in our advice and recommendations to SSA as it looks at data it collects through job analysts. So at this point I would like to turn the meeting over to Deb Lechner who will be giving a summary of their findings.

MS. LECHNER: Good morning, and I would like to thank everyone for giving us the opportunity to do this demonstration project, because I thought it was very instructive. I thought we all learned a lot, and it was an enjoyable process. I would like to thank Shanan and Robert for participating with me in this project and providing valuable information.

DR. BARROS-BAILEY: Can I interrupt real quickly. Your report and these Power Points slides are in the three ring binder under tab number two, the third red section back if anybody would like to follow along.

MS. LECHNER: Thank you, Mary.

Just to give you an overview of what we

are going to talk about this morning to begin talking about the project overview, the administrative set-up that went into arranging these job analyses. We included that information because we think it's instructive into the type of set up that will have to occur as the Social Security Administration endeavors to perform job analyses.

We used a couple of different protocols in this project that we will both -- we will all three explain our processes and what we did; and then we will compare and contrast these two approaches. And then also compare these two approaches to what we believe are Social Security needs, and then provide a quick summary.

The overall purpose of the project was to demonstrate two examples of job analysis protocol. And I think it's very important that we do make a bit of a disclaimer at this point about what was not the purpose of the project; and it was in no way to be construed as any kind of formal research project. This was just a demonstration project. And we're not advocating the use of either of those protocols

for Social Security Administration purposes. They were just merely used as a -- examples of convenience.

A protocol that -- I used a protocol that I'm familiar with. Shanan and Bob used a different protocol that Bob has had some input into developing in the past. So we just want to make those two important disclaimers before we start.

We had three specific goals under that overall purpose. One was to illustrate the protocols that are performed by three different types of professionals, as Mary mentioned, physical therapy, industrial and organizational psychologists, and vocational rehab professionals. And again, this is not to suggest that job analysis is only performed or should be only performed by these professionals. That just happens to be our background.

We compared and contrasted those methodologies, approaches and the reports that they generate; and then we compared them to SSA's expected data collection needs.

The process that I used was -- it's called quantitative job demands analysis or abbreviated QJDA. It's a process that I developed for work that I do in the field of job analysis; and it's primarily used to perform job analysis for the purpose of developing specific post-offer/pre-hire and return-to-work functional testing. So I'm using the results of the job analysis typically to develop functional testing. And it's focused exclusively on the physical, the sensory, and environmental demands of work.

The process that Shanan and Bob used was developed by King County, and the purpose of it is to provide a basic understanding of job requirements and job -- and to perform job matching, and to provide information for job accommodation. It has a more broad focus that includes physical, cognitive, behavioral, and sensory demands that are often used -- and the information from that type of analysis is often used in worker reassignment.

The administrative set up that Social Security performed -- and we were very grateful for

all the work that went into this -- they arranged -- they assumed the primary responsibility for making all the arrangements. They arranged for the three of us to visit local grocery store chains to perform an analysis of the cashier.

The staff reached out to various stakeholders in the Boston area for assistance in identifying these grocery store chains. And our contact held conversations with the store executives explaining the process, introducing to them the goals of what we were trying to achieve; and they also requested job descriptions.

As a starting point the three of us all used whatever existing job analyses there are as a starting point for our analysis. And again, we don't want to imply that this would be the process, the administrative process that would be used by Social Security Administration. We fully expect that there will be a very formal sampling methodology that is used when -- when SSA goes out to collect data; but this was just, you know, a process that we used for convenience sake to get

this little demonstration project done.

The process -- the QJDA process begins with reviewing existing job descriptions; and from that review developing a preliminary task list. Once I was on site I reviewed that preliminary list with input from supervisors and incumbent employees that were actually doing the job. And then revised that task list according to their input.

And then I proceeded to videotape and measure -- videotape all the tasks that were identified, or all the job functions that were identified, and measured the maximum forces exerted by the employees. I also measured distances and heights that these manual material handling demands occurred. And then we -- I also documented the environmental conditions, personal protective equipment, tools used and operated as well.

And then once off site, I reviewed the videotape and the measures, and entered all of that into the generated software. And then the software calculates the frequency and duration of each job task, and uses that data to calculate the percent of

day each of the tasks are performed.

Then I reviewed each videotape task to determine the percent of each task that is spent in each of the physical demands; and then the demands -- demand percent from all the tasks are combined to calculate the total percent of the day spent of the jobs in each of the physical demands.

The report that was created includes percent of day that job task was performed. It also documents an overall level of work that's defined by the DOT as either sedentary, light, medium, or heavy. And that, again, is not to imply that this will be the calculation system going forward; but we were all operating on the current DOT classification system.

We documented environmental conditions, personal protective equipment; percent of the job spent in each physical demand, the heaviest weight handled, and any manual finger dexterity requirement and coordination requirement. And I will turn it over to Bob and Shanan to describe the King County process.

DR. SCHRETLEN: Before you go on, do you want us to hold questions and comments until the end? How do you want to do that?

MS. LECHNER: What's your preference, Mary?

DR. BARROS-BAILEY: Let's go ahead and continue with the report and hold them until the end.

DR. GIBSON: When choosing what methodology I was going to utilize for doing this process, I always say that the nature of the job helps identify the type of job analysis or job analyst you will be using. Given that my typical reason for doing job analyses is to facilitate human resources functions within an organization, such as selection and training, I looked at existing instruments that were available for what I would normally use. And one of the instruments I had been introduced to was the King County instrument. I had never utilized it before. But what struck me about it was that it was composed of what I would typically call generalized work behaviors. And

since we have talked a lot about generalized work behaviors, it seem to be a natural form to adopt for this purpose. So that's how I came to use this form.

When we examine the King County forms that are available, it actually turned out that there were several King County job analytic forms that they have. There is a physical form. There is what they call the behavioral and cognitive forms. Then there were two addendums, which went into significantly more detail. There was a hand usage addendum, and a sensory addendum.

So I took all four of the forms and combined them in one very long analytic form. I maintained the exact scales and measures that King County utilizes, and we didn't move from that at all. We kept that form as theirs. Each form also included environmental factors and contextual factors; and we included those as well, as part of the analytic process. So that's where the form came to be that both Bob and I ultimately utilized.

In terms of the process I presumed -- I

did it just like I normally would if I was working with an organization. I had a contact name of the human resources manager for the organization; and I contacted this individual and requested a copy of the job description, just as Deborah did; and received back information that actually there was no existing job description for the job of cashier.

There was, however, a list of seven or eight general required physical dimensions that they use for the labor contract negotiations. That's all I had to go on from the actual organization at the outset. So that's what I began with.

As I would do before I went on site anywhere, I did go online and do some research. I looked at job postings and job descriptions for cashier across the web site, monster.com, anywhere I could find one to help key me into what I would be looking at when I got on site.

So I went on site. I did not have the advantage, as I said, of an on site human resources person, nor did I have the advantage of an on site floor manager. I point this out because there are

great differences in how businesses runs the job of cashier. And all three of us experienced something very different.

I heard somebody say this morning I had the classic 23 year old young man who was probably a cashier six months before, and because he showed up for work everyday on time now he was my floor manager. And he said, hi, nice to meet you. Here is your cashier. I have better things to do. I was told to go stand there.

With me was a SSA staff person. Claire actually did the exact same form I did, but worked with another cashier with the idea being this would be a model of multiple analysts on site, both conducting an analysis so that later you can come together and confirm, and find where you get discrepancies and consistencies in the data you collected at this site.

So I stood with the cashier. I actually literally stood behind the cashier -- I probably got in her elbow room -- as she did her job for better than an hour.

I watched. I marked. I filled out the form. And as I progressed through the form, I noticed that there were several things that I was not seeing her do, and so I couldn't answer and complete the form adequately. So I conferred with her, did you do this? Have you done that? Would you do this? How would you do this? When would this come up? Because, again, I'm dealing with a generalized work behavior analysis instrument, which wouldn't necessarily use the exact language she would in doing her job. That worked out fairly well for me.

There were things that I encountered, because I was there for an extended length of time that only happened once. If I hadn't been there at the right moment at the right time as an analyst I would have missed it. The example I have in this case was, you know, traveling and carrying weights at the same time. Walking some distance, carrying. For 90 percent of the time we were together -- 99 percent of the time she was confined to a very small space behind the cash register. And then one person

came through her line with vouchers to receive infant formula.

Infant formula is kept in a secure, locked area behind the customer service desk. Then she had to leave, travel by foot, go into the storage room and came out carrying a case, which was significantly more heavy than anything I had seen her carrying, and walking and traveling. So I am just trying to emphasize that part of the analytic process you have to allow for these things that happen infrequently, but which are certainly documentable parts of the job.

So for anything I didn't observe I talked with them. At the conclusion I went through the form with her, asked her if she actually agreed with my ratings, and discussed it. One of the best things about finding -- and I find when you work with people, is to understand why they are doing this. And I, unlike Deborah, didn't feel confident in believing that the store manager had explained the reason I was there, and what I was doing, and why I was working with her. So I made certain at

the outset that she understood, that there was nothing secret in my ratings, and tried to be open.

In an idea situation Claire and I then would have gotten together and actually compared our results; but, unfortunately, we never did that. We could have looked for consistencies, inconsistencies; and, perhaps, completed our final report and list some things that we saw that were different. And there were differences. My person was at a speedy check-out; and she did her own bagging. And her person wasn't, and had a bagger. That's the type of job title variability, which could be very common and is relevant in this type of job analysis. So that was my experience.

MS. LECHNER: Thanks, Shanan.

Bob, can you give a little bit about your process.

DR. FRASER: Yes, I managed it from Shanan putting several of these forms together. This was also kind of enjoyable for me, because I had some input into the cognitive, behavioral requirements many years ago; and I never knew what happened until

I was focusing on the form.

This store had a very sophisticated job description, which is a great, you know, platform for beginning things; and it was also current. There is a lot of jobs description here that were not current. She confirmed that this was current. That HR was on this every year. That was great.

I also had a nice sit-down review of the position with the manager who was very articulate, a college grad that was there also. I believe she is in graduate school. She was a really kind of on the spot, very clear, decisive in her responses across the range of physical, as well as cognitive-behavioral requirements.

Then I had time, maybe half hour, 45 minutes sitting down with the store manager. Then I had time to observe the cashier. And I was just across the aisles from her, not behind her; but just across the aisle from her. And was able to observe her for an hour or so. And then I also looked for discrepancies between what I saw and the store manager's input.

And then when that was over it was over, I was able to go through the ratings and confirm them with the cashier/checker; and then ask if there is things that I'm missing. And she pointed out several things, which would not certainly be essential functions of the job. But once a day you have to go up on the elevator on a pallet and pick up bound, heavier bags that might weigh up to 40 to 50 pounds, put them on a pallet, push it on the elevator downstairs, and then stack these bound bags behind the counter.

So again, maybe under four minutes of actual lifting, probably lifting on, lifting off; but in a nonessential function of the job kicks the job -- those requirements were a medium level of lifting.

She also pointed out that there was some crouching involved. She come around the corner if it was a an elderly person or person with a disability who was having trouble, kind of arranging things on the cart at a low level. So some minor points like that.

But I think that between the manager, which is a definite benefit for me, and the job description, and the input from the cashier/checker, it was good information and I enjoyed the process very much.

DR. GIBSON: One more thing I think I will point out. Since this was on a form I hadn't utilized before, and wasn't a form I would necessarily use for my own purposes, one thing that was important for me was actually training for myself on the verbiage on the form. I'm not a physical therapist or an occupational therapist. So, as I said, there was a whole addendum on hand usage. So it was necessary for me and Claire to sit down before we went on site with Mary and Bob and Deb and say, please explain so I understand the differences in these terminologies so that I can adequately assess them.

I think that's just indicative of something we need to be cognizant of, and SSA needs to be cognizant of. As they plan their training for their job analysts, there is going to be terminology

that is new and distinct, and that everybody needs to have the same working definition of these movement categories and other aspects of it. So even as a job analysts there was training involved for me prior to doing this.

MS. LECHNER: Thank you, Bob and Shanan.

One thing that we did not include on the slides was the approximate time frame that it took. All three of us spent one and a half to two hours on site. Then all three of us spent roughly couple hours preparing our report. So just to give you an idea of the time frame; and this is a fairly repetitive job.

And a comment that I want to add to Shanan and Bob's comments about these things that occur irregularly. I ran into that situation as well, and ended up having the cashier simulate a couple of things for me that she was not doing in the course of my regular -- you know, her regular job, as it was being performed while I was there right that minute.

So I had her simulate some things that she

had to do in terms of restocking some of the supplies. So you are going -- you know, typically you do run into things that are not being performed the day or the hours that you are on site. And some of these things you are going to have to describe or explain or simulate it.

We are going to take a break and allow the Commissioner to come in and speak to us; and then we will resume our presentation at the conclusion. Thank you.

DR. BARROS-BAILEY: Thank you. Let's go ahead and take maybe a five minute break, and resume with the Commissioner, who is here. As I indicated at the start of the meeting, the Commissioner has to leave a little bit earlier. So we will be finishing this presentation after the Commissioner leaves. Although, if you are ready to go now, we can take a break after.

COMMISSIONER ASTRUE: It's your choice. Whatever you want to do.

DR. BARROS-BAILEY: Let's go ahead. We just started not that long ago. Sorry about that.

Thank you, Commissioner, for coming to meet with us today.

As I said at the start of the meeting, this is a pretty important meeting for us in terms of significance. It's been two years since the Panel was chartered. And I would like to say on occasion I have referred back to the Commissioner's words to us at the inaugural meeting. I wasn't able to be there when he swore the Panel in at the beginning of the meeting.

And there are some words that he had in his opening comments to us that I think have kind of resonated with us over the last two years. And he says -- he said that OIS would take expertise, persistence, and creativity to replace the important part of SSA's process; and to do it in a way that is more thoughtful, will help SSA make more accurate decisions, faster decisions, and be as user friendly for SSA employees and for the public to use as possible.

Over the last two years, there has been quite a bit of movement within SSA with this

project. And we recognize this is not a program improvement project, that it is a mandatory project for the disability programs.

In developing an OIS, or an occupational database to meet SSA's needs, there are three criteria identified by SSA that we constantly heard about that this has never been done ever. Ever.

Those three criteria are that the databases are representative of the national economy -- the work of the national economy; that it's based on physical and mental-cognitive, human function; and that it meets SSA's burden of proof; that is that it's forensically defensible.

Although, I list that third criteria as last, what underlies the defensibility, such as the scientific rigor; on behalf of the Panel we are unanimous in our belief that ensuring scientific rigor means making sure that the skill set process and plans are in place to deliver that essential criteria.

We understand that the DOT is not defensible, that O*Net is neither usable nor

defensible for the same or other reasons; and neither system was created for SSA. There is no other system out there to meet SSA's needs. There was a report that came out from the Office of the Inspector General last week that reiterated these themes.

Again, we provided summary of the public comment report yesterday to the Commissioner of the nine months of collection of individual comments. The public comment responses called for high quality data. Data that was empirically established, valid, reproducible. I think Sylvia in her Project Director's report gave a really good overview of SSA's efforts going forward to meet those needs.

A variety of terminology from the public comment, obviously, assumes that in order for that outcome to happen with the database, the requisite framework needs to be in place.

So I just wanted to reemphasize three areas that we discussed that we, as a Panel that is made up of practitioners and scientists, believe are essential. And sometimes it's hard to get a Panel

like this to this place, but we are unanimous in this.

And we were happy to hear about the development of the Office of Vocational Resources Development. This is huge in terms of SSA's commitment to this project. It signals that commitment. The structure is in place, and we believe that the hiring of a lead scientist would compliment that skill set; and would be beneficial not just for research and development, but in the long term maintenance of the OIS.

We also believe that -- a business process. We had a very good meeting yesterday in terms of an introduction to such a process that would not only just lay the foundation within which programmatic and scientific staff within OVRD -- but also within the Agency -- could understand what happens when, and help the Panel now and after we are gone anybody within the Agency could understand what's going on. But in -- such a process would speed things up in the long term.

The last thing that we are -- that we

believe are essential to this process is the plan general recommendation number eight in terms of the overarching plan for this. Spending a little bit of time and resources in these three areas in the short-term will ultimately speed things up in the long term. And also, we believe, underlie the scientific rigor of this project.

On behalf of the Panel, I would like to state that we are committed to the vision and to the charter outlined two years ago. I would like to personally thank you for the opportunity to work with this group of very distinguished individuals. And when I say "this group," I'm talking about not just the Panel, but also SSA. This has been an incredible experience. I would like to turn over the floor to you.

COMMISSIONER ASTRUE: Thank you, Mary.

Let me start with some thanks myself. First of all, thank you for accommodating my schedule. I have a call to my oversight board at noon, so I do appreciate the flexibility. Thank you all for the meeting yesterday and for the gift of

the original copy of the 1939 Dictionary of Occupational Titles, which I did take a peek at.

You know, it's an interesting thing. They must have started this right about the same time that the Social Security Act was passed. We're celebrating our 75th anniversary this year.

And from 1935, 1939 it actually is an extraordinary document representing an enormous amount of work and the best thinking of the time; but things have changed. And I think that what you're doing to help us move from the best thinking of 1939 to the best thinking of 2010 is extraordinarily important for the Agency; and I think, you know, perhaps, for -- more for some of the others who are watching this project.

One of the things I would say is, don't look at it just in isolation. That this is part of a broader effort to overhaul our disability process and have it be across the board entirely state of the art. I think a lot of people will recall in 2007 the backlogs, the quality of the process overall was under assault.

It's actually remarkable that in some ways that this was part of the process that was relatively ignored; but we have tried to look at it from top to bottom and say going forward over the long haul, where do we want to be? What do we want to look like? And it's a little difficult for me on certain days because a lot of the things that we're doing that are so important I will never actually see the benefit of those. My gift to my successor, and my successor's successors. But I think if you care about an institution, that's what we do.

So we're doing it in the IT area where we have been operating for far two long with 54 separate Cobalt based systems. Many of which have common spine involvement adapting. So every time we want to make improvement for the public, it's a very long, slow difficult process, because we have to find Cobalt programmers to modify 54 separate systems.

So the good news is that we are moving slowly, because it is the federal government. We are moving to a common state of the art IT system.

I think we're dealing with the usual bid protest now; but we're hoping to have the beta for that up, I think probably late summer is the reasonable expectation of next year. It will probably take us a couple years to get that up and running; but I think that's an important piece of it.

A more visible piece was the medical part of evaluations. So four years ago, a substantial number of our regulations hadn't been updated since the 1970's to 1980's. So for instance, we have known for 15 years that, you know, contrary to what my mother always warned me, that being type A, liking spicy food caused ulcers, that they are actually caused by bacteria easily treated by antibiotics. We went 15 years not recognizing that, and many, many similar things in our Regulations.

It is just critically important that when we're making these decisions that are so important to individuals, not only -- in most cases not so much for the cash benefit, which is often important; but much more so the Gateway to Medicare and Medicaid, and medical treatment for the disability.

It's important to the individuals. It's also a big financial commitment by the taxpayers to these people. It's important that we try to make the right decisions as often as we can, and as quickly as we can; and having state of the art medical information is important.

I think the most neglected part of the process is others are instrumental in being more salient as in the occupational part. And I think that we, in the Agency, have known that this was an issue for a long time. I think there was sort of a hope that there would be an update at some point. That the Department of Labor would all of a sudden have a revivable of interest, and all of a sudden we would have the instrument that we need. But at some point you have to say they officially gave it up in 1991.

The last update of any size was 1979; and even in 1979, we were talking about marginal changes with a paradigm that really didn't fit a more modern economy. I remember my first experience with the DOT in 1983, and thinking in sort of a starky young

person's way how updated the document was then.

So from my vantage point it's been tremendously important to try and figure out how to do this right. It was interesting when I came in that seemed to be among outside groups a pretty common theme; and they could speak up, I think, usually expecting that we wouldn't actually do anything about it.

So it's been interesting as we have tried to move forward, and we have tried to reach out and get the very best advice we can from all quarters -- and this Committee is the most important part of that process, but also reaching out to the Department of Labor, NIH, Census. We were discussing yesterday with the revival of the Administrative conference in the United States, they might also be able to help. National Academy of Sciences, Institute of Medicine.

I think that I speak not just for myself, but for all the senior people in the Agency, that we're very conscious that while we have substantial expertise in a lot of areas that relate to this

project, we're not under the illusion that we have the full knowledge base to make all the decisions. And it's important for the quality of document, important for the credibility of the document that this continue to be a very open process where we try to seek the very best thinking of wherever that may be.

So we're committed to doing that. And I think that some of the anxieties that have been expressed in the last year I do really think are misplaced. Because I think if you look where we are, I don't think any rational person could look at it and say, yeah, that's fine; you should just stick with that. I think anybody that looks at it says it -- should be saying that we need to embrace change. And I think that your efforts have already been extraordinarily helpful in that regard.

I think that for a number of the key policy makers in Washington, particularly in the Congress, I think that your 2009 recommendations and your Panel findings on the NAS review of O*Net, I think those kind of things, I think, get people to

pay some attention, but not as close attention as others the assurance that, you know, this is -- this is the type of effort that we need to embark on. That we need -- that we can't rely on the old document. And there is no other easy solution that's out there. There is a little bit of facile discussion that, oh, you know, the Agency is making a horrible decision just not shifting to O*Net. I think having a Panel such as yours has not only given us the independent opinion, but evaluated what others have said. I think it is tremendously important and has laid a lot of the framework for moving forward.

One of the things we discussed yesterday is I think that at this two year anniversary, there are starting to be a shift of priorities. I think we needed to go through a process for a while, think about conceptually what we needed to do, how to do it, what the options were. And I think the pull now is to actually getting on with the job, figuring out the important details of what the process should look like, how we continue to build a process that

continues to get the highest quality input; and also defines the task in an efficient way.

It would be nice if we could start using some of the new -- some of the material from the new instrument before the instrument is completed all together. And I also think that it's important for us in the Agency not to in subtle ways get bound by the old paradigm. You know, we work with it so long, it is so much a part of the architecture of how we think about this, that some time I think it's hard for us to look at it in a fair way and say, you know, there are other ways to doing it.

In particular, I think one of the enormous achievements of the original document was going through a lot of the detail of what was substantially a blue collar economy. Of course, for us when we first started the disability program in the 1950's it was really set up to deal with those type of injuries. So it was a fairly natural marriage between the original Title II disability program and this document.

But the economy has changed enormously. I

think that, you know, we look at most of the jobs that are in the economy today, and the whole taxonomy of the DOT just doesn't fit very well. We have to think about different ways of describing the jobs that are in the economy.

The other thing that's changed radically is our notion of disability generally in the country has changed and expanded. And even though we deal with a much more limited statutorily specific definition of disability in Title II and Title XVI, even within that narrower framework, the notion of who we're serving has changed very dramatically. We're serving children now. We see an increasing number of people with developmental or other intellectual limitations. A number of things that -- a number of conditions that were not considered disabling in 1957 are fairly common. Basically, it's more disability today.

So I think it's some -- tremendously important for the outside input to help us make sure that, in addition to getting state of the art information and evidence and data in the document,

that as we come up with a new general structure, a new way of thinking about this, that we're not bound by the old paradigm, because I think that would be very easy to do. I think that's one of the reasons why it's tremendously critical for us to continue to get outside support.

So I'm going to stop and leave a little room for questions. I would be derelict if I didn't thank you all for your service. It's been long. You have been very dedicated. This is a very hard project. And you know, it is certainly not that the compensation is high, the glamor is high. It's like a lot of Federal service, I think you have stuck with it, and have done a very nice job. I'm very appreciate for the sacrifices that this group has made.

DR. BARROS-BAILEY: Thank you, Commissioner, for your words.

I would like to maybe open it up to the Panel to see if anybody has any comments; any words. Allan.

DR. HUNT: Could you expand on your

statement about doing this in an efficient way, and give us a little bit better idea of how you see that?

COMMISSIONER ASTRUE: So what I would like -- I think what I'm suggesting -- it's not what I would call directive for an expectation, but I do expect from the point where we are now to when we say, okay, we're done, we have moved from creation to regular maintenance will be a very long time. I'm not knowledgeable enough to even give a good guess of how long it's going to be; but it's going to be a long time. And they will be sending me post cards to tell me how that's working out.

So one of the things I said is, you know, we do have a tendency, I think, like all federal agencies on these big projects, to think in a Manhattan Project style; and that nothing in the world will change until we get to the end. What I don't know is even possible -- I continue to ask the staff to think about it; I ask you to think about it -- is if as we go through our Manhattan Project here, if there are discrete parts that are easier

that we could segregate out and use earlier than when the whole project is completed, I think that would be a great blessing to the Agency and for the people that we serve.

Whether that is even doable, I don't know. But it's the type of question that I hope that my staff will continue to ask, and that you will continue to help them as they ask that question.

DR. BARROS-BAILEY: Abigail.

DR. PANTER: That's actually very consistent with the information, thinking about the nature of the scientific process. So I am happy to hear that you think this way. This is how we work. Usually it is gathering data, and in a fair and balanced way drawing some conclusion and updating. So appreciate the comments.

DR. BARROS-BAILEY: Okay. Do we have any other comments, questions?

Okay. We thank you for -- sorry about that. I didn't realize they shut each other down. We thank you for your time here. We appreciate -- we know that you were coming from D.C. to Baltimore.

We know you have another commitment that you have to run off to. So thank you for your time yesterday. Thank you for your words and your time today.

COMMISSIONER ASTRUE: Again, I just want to say we're incredibly grateful for everything that you do. This is a very important project. And we're just thrilled that you are still working at it, still giving us so much help; and we're going to continue to need it for the foreseeable future. So thank you very much.

DR. BARROS-BAILEY: Thank you.

At this point let's go ahead and take a 20 minute break. Thank you.

(Whereupon, a recess was taken.)

DR. BARROS-BAILEY: I think we have almost all Panel members seated to be able to resume our meeting.

MS. TIDWELL-PETERS: Ladies and gentlemen, if you could please take your seats. We're about to begin the meeting again. Thank you. Please take your seats.

DR. BARROS-BAILEY: Thank you, everybody,

for coming back from the break. I would like to put Deb and her Ad Hoc Group back on the stand, basically -- you are not on the hot seat. I testify too many times.

Deb, if you would go ahead and continue with your presentation. Thank you.

MS. LECHNER: Sure, Mary; thanks.

One of the things that as a Panel we bumped into from time to time is just coming from different disciplines, doing job analysis with different approaches, we have often encountered terminology differences. And it's caused some confusion in some of our discussions that we end up having to spend some time to straighten out. So one of the things we look at, you know, between these two different job analyses processes, how is the terminology different?

And one of the things that we found was that the terminology is very similar in both approaches and both reports. And they are the top terminology, because they were -- both the processes were developed when the use -- under the use of the

DOT. They are based on DOT terms. Again, not to imply that we think that's how it should be going forward, but that's just the two instruments that we were using.

We did notice a couple of minor variations in the QJDA. The job is broken down into essential tasks; and in King County, essential functions.

Demands are referred to in things like sitting, standing, walking. The QJDA process refers to them as demands or physical demands. The King County process refers to those exact same things as requirements. And so a couple of minor variations; But if two different professionals are talking about -- one is talking about essential functions; one talking essential tasks; and one talking demands; one talking requirements, you can see how that will create some confusion.

And as Shanan has mentioned earlier, one of the things that all three of us feel very strongly about is this whole issue of operational definitions, and what an important role that needs to have in whatever job analysis process is

developed in the future. We really feel like having those definitions in whatever procedure manuals are developed, having them emphasized and taught in the training, and making sure that certified analysts are competent in recognizing those -- that terminology; as well as, perhaps, even having some of this terminology -- as an addendum to every report there may be a glossary. So especially as this is being rolled out nationally, it will -- something like that could facilitate communication. So that's one of the things SSA may want to consider.

And then protocol comparisons. The QJDA, as we have already mentioned, only include tasks that have physical or psychophysical demands. Whereas, the King County process includes tasks with cognitive, behavioral, physical, and sensory demands.

The QJDA documents the percent of day that tasks are performed. The King County just indicates that a function is performed, but does not indicate what percent of day. And then in the physical

demands or requirements section of the instrument, the QJDA indicates the exact percent of day that the demands are performed. Whereas, the King County form indicates a range of percent of demands -- or a range of the day that the job or the tasks are performed -- demands are performed.

The QJDA doesn't include the work surface that the employee stands on. While that's something that is included in King County, the QJDA does not include maximum continuous due to duration; whereas, the King County form does.

The QJDA doesn't indicate what the most common -- what the person is most commonly doing when that physical demand is performed. The King County does include that information.

The QJDA doesn't include bending the neck, while King County does. QJDA breaks reaching into two different levels. King County breaks it into four. And QJDA doesn't include ratings of cognition and behavior; whereas, King County does.

QJDA documents only the four steps required for pinch; but the King County breaks it

down into three different types of pinching.

So I bring these up, because this is really down in the weeds, the detail between these two protocols, but we felt it was important to highlight this. Not because these are huge important differences, but these are things that as the protocol is developed, SSA and the contracting company that develops it will have to reach some decision about the level of detail in a variety of different ways.

The environment, the QJDA has a more open ended descriptive approach to the environmental factors. The King County has a more inclusive list of environmental factors that the analyst chooses from. Again, this could be important in terms of the level of consistency that's seen between the different analysts. So if you have the longer, more standardized pick lists, then you probably will be more likely to get consistent data; but you may lose some of the richness of the descriptive data. So there is pros and cons of different methods of recording data.

And then the data collection methodologies. One -- QJDA uses videotaping and actually measures forces and distances with a four stage tape measure. The King County relies more on observation and some self-report, but self-report that is validated through observation without weighing of forces and weights, and measuring forces and distances.

And then after we compare and contrast these two different methodologies to one another, we started thinking a little about well, how would either of these methodologies compare with what Social Security needs in terms of the information from job analysis. And our conclusion was basically that neither of these processes would likely meet all the needs of SSA. And I think that's probably going to be true as the contracting Agency does the literature review of the existing protocols out there, none of which have been developed specifically for the use of SSA.

I think we're going to see -- I think they are going to turn up a lot of really good protocols

that are very good for specific purposes, but none that would meet all the needs of Social Security.

The QJDA focuses primarily on the physical and psychophysical job demands, and environmental factors, where SSA really does need the cognitive and behavioral components; and we know that's an important piece of information. The King County includes all of that, so that it would come closer to meeting SSA's needs from a comprehensive perspective; but, you know, on the flip side of that, the measurements are not as specific as the QJDA. And what level of specificity is going to be right for SSA, I don't think, has been determined.

And then the cognitive and behavioral aspects of the King County approach are not -- the terminology is not really consistent with the MRFC. And so as that component of the job analysis process is developed, there will have to be some decisions about what parts of the MRFC need to be kept in tact. If they are, what -- how open is SSA to new terminology in that arena? Because I think that will be much more so than the physical area, the

mental cognitive area will introduce new terminology to the process if it's expanded in the way that we believe SSA has expressed that it needs to be.

Both formats also include details that may not be necessary for SSA's adjudicative process. For example, the exact percent of the day that a job task is performed. Is that performance important? Or the surface on which the employee sits and stands, is that important?

Both formats also exclude some of the details that may be important to SSA. For example, what is the grade of the walking surface? Is there a required pace or speed of ambulation? So the different kinds of questions that -- that need to be answered.

And then, you know, really the level of detail that's necessary for SSA's adjudicative process we really feel like that, to some extent, will drive the process that's used for job analysis. We have to know that -- the Administration needs to know what data we need to collect in order to develop a real -- really effective instrument.

Both approaches were designed with specific training that teaches the process, the data collection process, report generating process. The QJDA approach requires training, and a written and practical exam in order to achieve certification.

The King County approach initially has some formal training, but through conversations with some of their representatives lately, I think the training that was originally associated with it has sort of become over the years less standardized. That training has a little bit of decreased emphasis as of late. And so that example really speaks to the importance of ongoing maintenance of the training process and holding the standards over the years. Because I think some of us who have spoken to the original job analysts that were with the Department of Labor, I think what we heard from some of those folks was that that process initially began fairly standardized and as time went on standardization deteriorated.

So I think that's something that SSA will need to guard against, and have procedures and

policies in place that make sure that the initial high quality standards of the training and certification process are upheld throughout the years as, you know, initially we certainly will do more attempt data collection, but over the years as new occupations emerge, data collection will need to continue; and existing occupations will need to be revised as technology and other changes dictate.

We also had some discussion as a team about videotaping versus observation. The videotaping adds an element of time and expense. It provides additional validation and legal defensibility if the purposes are to develop functional testing, or to develop ergonomic countermeasures; and neither of those will be the real purpose of SSA's job analysis.

So SSA will need to decide whether videotaping is desirable. Whether, you know, this will be instruments that measure the forces that for each of the QJDA processes are -- the instrument have to be calibrated periodically. So all those are operational decisions that the Administration

will need to deal with.

The self-report of weights and forces, where weights are handled and forces are exerted, things in the environment like the grocery store chain we were in, and maybe a warehouse environment where the weights are actually labeled on the material that is handled, observations works out pretty well. But then it becomes in other environments where equipments, weights are not documented by the organization, just observation alone becomes a little bit more problematic.

It is also problematic for tasks that involve machine pulling, because the force -- for example, if you have a rolling cart, couple hundred pounds on a cart, that doesn't mean that it takes 200 pounds of force to push it. And the amount of force it takes to push it depends on really the friction between the surface and the cart being pushed on. So it might only take 30 to 40 pounds to push a rolling cart with 200-pounds of force. So we have to keep those kinds of things in mind as we move forward. And so self-report may not be optimal

in settings where the weights and forces are unknown in pushing and pulling. But making those measurements does require more time and investment in equipment.

Neither of the reports had operational definitions that were embedded in the reports, and that's something that SSA may want to look at. And what the King County process didn't really provide these formal operational definitions of the requirements. As Shanan has already eluded to, that that creates some challenges in communication among analysts and -- between analysts and the actual folks that are doing the job. So we really feel like that -- that operational definitions are important. And I have already mentioned the issue with glossary.

So in summary, our project provided some insight into similarities and differences between a couple of different approaches to job analysis. We found a lot more similarities than differences. And many aspects of either approach provide information important to SSA's.

There may be issues important to SSA that are not covered with those two processes. And SSA needs to determine the content model of a new OIS before a formal job analysis system can be developed. The rating system -- the OIS items, the rating systems will dictate the job analysis process to a great extent. And that's all we have for our report. Thank you.

Can we open it up for some questions?

DR. BARROS-BAILEY: Absolutely. Bob.

DR. FRASER: I just want to -- as I look at the job analysis that I did at King County forms, I think the cognitive behavioral is actually a lot closer than I initially thought. A lot of it just could be wordsmithing. For example, attending is one of our criteria in our cognitive behavior list of things; and they have something called working with heightened awareness and scanning. Some of these things can be pretty close.

The other thing about the King County approach is it really goes beyond -- below the task level. Like neck bending would be part of a task.

So they have a lot of things that are really down at what's called an element level, an element of a task. A group of tasks would be a domain of function or category of function. Just some comment.

DR. BARROS-BAILEY: Tom.

MR. HARDY: I have a couple quick questions, and one theoretical question, which is not quick. I was curious about the detail on page four about how the one system is able to calculate percentage of time. Can you give us a little more information on how that's done for the QJDA process? That's question one.

MS. LECHNER: We start with an interview of the incumbent, supervisors to get how often a particular job function or job task is performed. So let's say it's a housekeeping job, and the housekeepers have to go around and mop hospital rooms. How many times a day do you do that? And how long does it typically take to you to do that? Then knowing the total shift length we calculate the percent of day that that job function is performed.

And then we go videotape each job function, and we document the percent of that job function spent in all the different physical demands. So, you know, if they are mopping, what percent of the day -- what percent of the mopping task is spent bending, stooping, squatting, reaching. So we would add up all those percentages. Then those two percentages are factored together so that we get a weighted percent. That's all added together at the end of analysis.

Does that make sense?

MR. HARDY: It makes sense. I'm just trying to figure out in kind of a broader sense on page ten of my book you have both coordinates with detail that may not be necessary for adjudication, such as exact percentage of the job -- the time that job task was performed.

I'm curious because if we're looking at trying to establish -- they're many different ways of trying to meet that information. I'm curious why you said that, because say you have a blended job, and you need to know how much time is done for one

kind of task for one part of the job, and how much time for another. Of if you are trying to come up with a final exertional level, you know. Why does that statement seem a little bit counter intuitive?

MS. LECHNER: Well, you know, in my experience with job analysis I find that that information is very helpful and useful; but there is a time element involved in doing this calculation and doing a more quantitive analysis. And I don't feel like I have a really good sense yet exactly what would be important to SSA.

So I just wanted to point out that that level of detail very well may be what SSA needs; but on the other hand, it may not. So I'm just trying to be open minded at this point and not say this is absolutely what SSA is going to need, because I don't feel like I know yet.

MR. HARDY: I just have one more quick question.

You said that the three of you varied between two and a half to three hours to do the job analysis. Did you guys break that down in how much

time was spent in observation; how much was prep time prior; and how much was compilation thereafter, or is that kind of a chunk?

MS. LECHNER: We each spent about an hour and a half to two hours on site; and then off site we spent about that same amount of time, you know, entering data, creating reports. And then I can't really speak to how much time was spent in preparation. I don't know if someone from SSA's staff is here that can speak to that, you know. I would imagine there were quite a few phone calls back and forth. So I can imagine that several hours were spent in making all the arrangements.

MR. HARDY: Did either of these formats have a specified amount of time in your directions as to how long you should spend in observation, or is that decided individually?

MS. LECHNER: I didn't see any specified time frame that -- and in my experience the time frame is typically dictated by the variability of the job. So if it's a job -- let's say if it's a manufacturing position and it's a line worker, and

they are repeating things over and over and over again the exact same way every 60 seconds. You spend a lot less time either observing or videotaping that job than you would, say, if you analyze a job for a maintenance position where their job is so variable that one day they may come in and spend the entire day working on the HVAC system. The next day they are coming in, and, you know, repairing, replacing light bulbs in the ceiling, those kind of things.

So those kinds of jobs sometimes even require multiple days on site to be able to capture the variability. So the length of time required varies greatly according to the variability of the job.

DR. FRASER: This format in King County was used for the King County executive and was also used for meter reader. So the variability is --

MR. HARDY: That takes me to my last question, which is sort of broad and theoretical. Shanan, you said something about you think the job analysis utilized varies by what it is you are going

to be exerting. We're talking about trying to come up with something, I think, that would cover everything.

DR. GIBSON: What I said was that job analysis varies with the technique. The forms you utilize varies based on your purpose. That's actually different. For example, if I'm doing a job analysis and I know the organization's purpose -- and I wouldn't do them without identifying the purpose with them first -- is to utilize this to develop a structured interview for selection purposes. Then I'm going to be focusing on identifying things at a level that allows me to develop good, behavioral questions.

If they are doing it to develop their training program, I'm going to seek out a job analytic method or I'm going to take it probably from a task level analysis, because I need to be able to train at the task level. So yes, for SSA to do disability adjudication, everything they do has to be based on that knowledge of what they need to know.

So when we talk about developing the items for their content model, at what level do they need to measure that? What type of scales are they going to utilize? What frequency matters for the percent of time? And what this one measures -- for how many minutes or hours at a time. For how many cumulative minutes or hours, total work shift. That type of philosophical question they may have to answer at the outset before they could develop their measure, I think. So that's what I was eluding to.

DR. BARROS-BAILEY: Abigail.

DR. PANTER: I very much appreciated the methodological points that were brought up, and some of the things we need to be thinking about. I also would like to make a point that it's important to formalize the idea of having multiple raters with the same information, and also multiple performance about a job. We saw cross examples that you could sometimes have a job description and sometimes not. Sometimes have a person that's on site, and sometimes not. And you could have a supervisor, and sometimes the supervisor is more articulate than

another supervisor. So these are all forms of information that come into -- the information that will be important to our values.

DR. BARROS-BAILEY: Janine.

MS. HOLLOMAN: I'm just curious as to whether either of these instruments identifies something that comes up very often in the disability determination and in appeals, and that is the ability to do the job with a sit/stand option regardless of the weight; and whether the job can be done one handed. Did either one of these instruments identify those components or potential job components?

DR. FRASER: Definitely not sit/stand option.

DR. GIBSON: But maybe the one handed. I'm looking it up in the items.

MS. LECHNER: The answer to that question with the QJDA is typically our goal is documenting the job as it exist at the present, and then if you are asked by the organization to make recommendations in terms of ergonomic modifications

and changes then there is a section to the report where we identify the hazards of the job and propose counter measures; you know, ergonomic changes that could address that.

So we're not coming at it with a standard addressing of that particular question, although, it could become clear that things could come up as part of the ergonomic assessment piece of it.

DR. FRASER: The one hand isn't in the King County form.

DR. GIBSON: I actually have to concur with Deb as well, in the fact that when I'm doing job analysis it is for documenting the job as it occurs. To ask me could it be done becomes a very subjective measure, which sometimes puts you in a problematic position, because I am supposed to be documenting what is verifiable. So that becomes the opinion. If I were to inquire of five different incumbents I might get three who said yes, and two who said no. So that's a very hard position to take as a analyst to say "yes" or "no," it could or could not.

MS. LECHNER: You know, and I agree with Shanan. Once you get into the speculation of what could be it gets a little bit gray, and a little less subjective. Although, I know in many testimonies from end users within SSA, those two issues come up over and over again so that, you know, the job analysis process should, perhaps, attempt to address those issues; and that SSA should be attempting to address those issues, I think, is the question.

MS. HOLLOMAN: And I agree with that wholeheartedly. Thank you.

DR. BARROS-BAILEY: I wanted to add to that, because I think we're talking about, does it exist or can it exist? And from doing job analyses in potato warehouses in Idaho, I can tell you that in a lot of those potato warehouses that the potato sorters are given stools. If I observe that they can do it sitting or standing, it is part of the job.

So that option isn't a theoretical option, it exist. It is part of the way the job is done.

So I think that sometimes we need to consider is it a part of the job, and there are some jobs it may be.

A couple of things -- a couple of observations that I wanted to make in terms of the comparison terminology. I think everybody is -- understood that I am kind of hyper concerned about terminology. I think what I heard the Ad Hoc group say was that terminology was important in terms of roll out. What I would probably encourage is that terminology is important right now. And I know that Sylvia talked earlier about the R & D Plan; and I would highly encourage that terminology to be developed as soon as possible. So everybody along this process has a common language to go from, and not as things get rolled out later.

A couple of things that were interesting to me and part of that terminology -- going back to that slide -- that QJDA and the King County used either essential tasks or essential functions. Just understanding that in terms of ADA language we're not dealing with that in terms of SSA looking at

core issues.

A couple things that I was interested in as well was, Deb, you were talking about looking at job descriptions, coming up with tasks.

Shanan, you were talking about the King County; and the King County, the thing that struck you was that generalized work behaviors is generalized work activities. I know sometimes those two terms among us have almost caused a little bit of conflict. Did you find that you were actually talking more about the same thing rather than different things when you were talking tasks and generalized work activities or work behaviors? I would love to hear both of your input on that.

DR. GIBSON: I'm sitting here looking at the things that were actually measured between the different things, sitting, standing, walking; those are very common terms. So I think we have made this point several times, the delineation between what is a generalized work behavior, and what is a task is not always clear. They are frequently the exact same thing.

I actually googled a moment ago smiling, what Bob said, things like bending your neck are almost subtasks or an element; or it can be higher than a task, because it's something that transcends many tasks. So it's all in how you look at it in a hierarchy.

This one has lots of things. One of my favorites was pinching, all right. Pinching could be seen as a subpart of picking up marbles, if picking up marbles is your task; or it could be seen as something that transcends multiple things that you do, picking up marbles, picking up widgets, picking up -- so that's not always clear. That language we use can vary somewhat, the terms.

DR. BARROS-BAILEY: I think because they are so similar, it becomes a little bit confusing in whether it's a work demand that I see in that slide -- QJDA talks about work demands. What is a work demand, as opposed to King County? What is it that work requires, which is more person centered? What is it that the person has to be able to bring in -- it meant the same thing; it is just the

terminology. So it really was demands instead of requirements.

DR. GIBSON: I think you are using a connotation of the word "requirement;" assume it's a human requirement instead of a work requirement.

DR. BARROS-BAILEY: Okay. Back to terminology. My point is made with myself.

Any other comments, any other questions for the group? Go ahead, Deb.

MS. LECHNER: You know, in terms of this whole terminology issue, I think that an important take home from this is that the exact words used aren't really important. It is just that the landing on -- everybody agreeing, okay, these are the words we're going to use and these are the operational definitions of those words. And I think that's sort of step one in developing a good content model. And will be critical for the instrument is that -- you know, who cares what we call it, let's just all agree, and agree on a definition.

DR. BARROS-BAILEY: Dave.

DR. SCHRETLEN: Thank you. You know,

first of all, just an editorial. I want to thank you guys for going out and completing that exercise. I'm sure it was a lot of work. But -- one of the things that I like about this, we have been doing so much work over the last couple of years on this Panel, this feels like kind of the first fore to actually going out and seeing what data looks like. So far it's been much more restricted to literature reviews, and it's been a lot more abstract.

So notwithstanding Deb's disclaimer about the fact that this is just a demonstration project and nothing more, I am very excited about it. It brings up for me a ton of questions, more than we're going to be able to answer before 11:00 o'clock when we are scheduled to adjourn for lunch.

But one of things that I have just a broad muster of question is what is the universe of job analysis instruments that is out there? Are these two examples of a couple dozen that are out there, or scores, or hundreds, or just a handful?

DR. GIBSON: I won't try to tackle that with an authoritative answer, but I will say there

are dozen of commercially available instruments out there. There are also -- many times when there is no instrument used -- if I'm going in to do a task based analysis, I don't have an instrument. I have a series of scales I know I'm going to utilize. What are my frequency scales? What are my repetition, my importance, whatever it is. When I'm going in I am starting with a scratch list, a sheet of blank paper.

DR. SCHRETLEN: So when you and Mark describe these taxonomies, there are lots of taxonomic systems out there; but they are not all tied to instruments that have been used to sort of support those taxonomies, is that correct?

DR. GIBSON: What was presented in the work analysis subcommittee's report was every -- was a list, if you will, of all the empirically derived work dimensions or associated work analysis instruments. Are there others out there that, perhaps, have not been empirically verified and published? Yes. But we limited ourselves to those because we felt like those were the most defensible

place to begin for SSA if they are going to start creating an universe of elements for consideration.

MS. LECHNER: I have a little more to add to that. I think there is the universe of what -- of instruments that are out there that have published research and documented reliability, validity, or that are in the public domain. And, you know, certainly I can recall, you know, 15 or so years ago doing a grant proposal where we looked at the instruments that were in the literature that were focused primarily on ergonomic type job analysis. And that ended up being, you know, somewhere in the neighborhood of 15 to 20 instruments.

I think there are probably double that many out in the -- being used in the world of job analysis that are not -- you know, that are commercially available to use, but no -- not without any research behind them. And then there are literally hundreds were clinicians -- maybe even thousands were clinicians who they want to get into job analysis, they put together their own form.

So you see that whole spectrum of research based instruments that are standardized and have evidence of reliability and validity. There are commercially available instruments that are standardized without research. Then there are, you know, sort of the hodgepodge, home grown instruments. So I think there is a huge variety in that.

DR. SCHRETLEN: So in a sense you guys did a little experiment in which you sampled two of the universe of instruments. Is this just a sample of convenience, or things you knew, or were there any particular -- you don't see these as necessarily the best instruments or anything like that?

MS. LECHNER: These were absolutely samples of convenience. I used what I developed and know; and then Bob had input into the King County cognitive elements. So he was familiar with that instrument. So we thought, great, these are two that we can just div up. By no means have we did a literature review and choose the best one.

I think part of the ICF work that we will

hear about in the coming months, they are going to conduct formal literature reviews, looking at instruments out there; and they will have probably some recommendations and thoughts based on that literature review.

I think that will be a very, very exciting piece of work. It's a big job to look at all that's out there. I'm looking forward to hearing that information.

DR. BARROS-BAILEY: Sylvia.

MS. KARMAN: I just wanted to point out that the experience that the ad hoc subcommittee had with doing the job analyses, at least from our perspective, is really about the process of conducting job analysis, and what that entails. So the instrument, while that's certainly a feature of what the protocol then is, when you show up on site, and whom you are going to talk with, what are you going to ask them, and how long that may take, because of the questions -- the instruments themselves, as I understood it, was not the significant aspect there that was being examined.

So I just thought I would put that out there.

DR. SCHRETLEN: I do appreciate that. I don't want to overly focus on the instrumentation, but it just seems to me like this might be foreshadowing of what we will see that, in fact, at the end of the day will be necessary to cannibalizing various systems that were not like the system deeply suited as the existing form.

But if we're not going to focus on instrumentation, I would like to return for a moment to Tom's question about how do you extrapolate from your observations? Because in both of these systems you talk to other people, and you talk to the worker, and you observe.

So in a general sense, in the world of job analysis, how do you resolve differences, when the worker says, oh, I do this 50 percent of the time; but your observation, and you think it's a typical one, shows that it's 10 percent of the time?

DR. GIBSON: I hate to say we rely to some degree, not entirely, on the law of averages. There is some. It goes back to what Abigail said a moment

ago about multiple raters, multiple subjects. So you need the wisdom -- the combined wisdom of several analysts looking at the job of cashier. You need those several analysts each looking at several cashiers.

And then sometimes -- what I'm faced with personally where there is disagreement, and where I see disagreements some time is what the management says and what the incumbent tells me. If we have to sit around the table and you have to play good lead facilitator and try to come up with, well, is the manager telling you what he wishes was required of the job? Is the person telling you what they have gotten to after six years of experience and they are cutting corners on the job. Sometimes it just requires some digging.

DR. SCHRETLEN: You just described two very, very different methods. One is, you say the law of averages. The supervisor says it is 10 percent of the time, and the worker says it is 20 percent of the time, and your observation is 40 percent of the time. You could take the average

of those three, and then you are just assigning equal weight to all three sources of information. You are, in doing that, saying that your observation is no better or no worse than the worker's estimate or the supervisor's estimate.

Alternatively, you said that you could use judgment. The judgment of the trained job analyst who listens carefully and appreciates the particular kind of slant that the worker or the supervise might put on it, or the fact that your observation might be at a time of day that is less typical of what the person says. Is that correct?

DR. GIBSON: First of all, it is not -- perhaps I misspoke, but I didn't think I was going to be taken literally when I said the law of averages. I didn't literally mean you take A, B, and C, and do it by three's. What I meant was that you have multiple sources of data. And yes, you would rely, I think, more heavily on the trained analyst to be the tie breaker, if you will.

But once again, it is not just three sources; it's many sources. As with any data, you

have to go and look at the data as a researcher and decide, are their outliers here? If there are outliers actually involved it becomes even tougher, because is that outlier a legitimate source within job variability? Because if it is within job variability, you may have feed back data; it might be legitimate data. And that's where your analysts have insight, because they should have seen multiple cashiers across multiple locations.

And I personally am always careful when I find myself talking about clinical judgment, because we know that the research on clinical judgment versus empirically derived evidence isn't flattering on judgment typically. However, I do think we should be cognizant of the fact that a trained analyst playing a role here should have significant insight.

DR. SCHRETLEN: I didn't mean to be concrete about -- when I said the law of averages. In fact, I -- I distinctly heard Abigail say that's it's important to have multiple raters. So it wouldn't be three. It would be a larger number; but

it's still a number that we're talking about. Actually, I'm not meaning to be critical of that. I think that the truth may, in fact, emerge from multiple observations and you identify the central tendency. That's a very reasonable scientific approach to things.

But it does sort of underscore this thought, this presentation, the distinction between observations and reports that you get, and there are going to be inferences. And I just don't -- this is not my area, so I don't know how this is resolved.

It makes sense to me to use one's judgment, but I appreciate that in the end you might have a combination. Each job analyst using his or her best judgment enters data about the amount of time or amount of effort on a particular kind of generalized work activity, and then we take the average of those, or we use some representative measure of those.

MS. LECHNER: David, I think, this is one of the areas where I do think the videotaping approach gives a little bit more objective, concrete

information. Because when we sit down and do our initial interviews we say, okay, so how long does -- three times a day do you do this mopping task, and how long does it take you? And they give us a number, and we go and videotape it. Let's say they say 15 minutes, and we can see, wow, it takes 40 minutes.

So, you know, obviously, you can only do "X" number of those in an eight hour day. So we have a more quantitative mathematical approach that we can compare the self-report. Then when we generate our reports and say, you know, we changed this. We changed it because -- or it -- was this an exception? Was this -- you know, was there any reason that what we saw and videotaped was an exception?

Then the other point I want to make is that I think SSA has some very important decisions to make in how they approach job analysis when we approach it from a perspective of developing a post offer, preemployment testing we're going into the work site and looking for worse case scenario.

What's the most weight you would ever have to lift? Because the job applicants are going to have to come in and do that worse case scenario. And if this is for the purposes of preventing injury, then you want to make sure that the applicants can do the worse case scenarios.

SSA may have a very different perspective on that and be looking at, okay, I just want to know the minimum of weight that has to be handled. Or we may, you know, end up deciding let's take an average, you know, that -- because when we're looking at jobs we're trying to create occupational categories that occur within the national economy.

And so -- or we can look at because we have access to such sophisticated IT approaches to these questions now that we didn't have when the DOT was developed, we could also be looking at, okay, frequency. There are a hundred of these cashier jobs, you know; 50 percent of them require 20 pounds of lifting. The other 50 percent require 100 pounds of lifting. So we can look at frequencies and compare frequencies to even geographic areas, if

that becomes an issue.

So you know, I think we -- the power that we have to analyze -- collect, analyze, group data is an advantage. We have just got to decide -- or SSA has to decide how far do you want to go into these issues? And how much and how long you want to spend on data collection? All those are sort of scientific questions related to policy and operation.

DR. BARROS-BAILEY: Tom.

MR. HARDY: Yes, I got another question. I recognize the goal of this was just sort of to look at protocols, formats, and get out there and give a start to this great exercise. But for me part of the job analysis also entails some of the things that seem to be very physically focused. Was there any evaluation of time and efficiency, work complexity, anything like that to see if there is agreement or disagreement, or how you approached it? I'm just curious.

MS. LECHNER: There was not that in my QJDA process. I will let Shanan and Bob speak to

the King County.

DR. FRASER: Tom, there are about three and a half pages of cognitive behavior -- work behaviors in there. And it's kind of similar to what we currently have listed. It's just more detailed. For example, there are five or six kinds of remembering. You know, remembering auditory information, video information, idiosyncratic details, and stuff like. So it's pretty comprehensive.

Just one comment from David's earlier point. If we don't videotape, the quality of the training for these job analysts will be very, very important. For example, you talk to the manager there is no meeting level listening, okay. You talk to a cashier, a checker, and she says oh, no, 20 minutes a day I'm lifting at the medium level.

Then the trained analyst would probe, well, are you lifting for 20 minutes?

No, I'm actually lifting for two minutes. Well, actually maybe between one and two minutes, then I am unloading one or two minutes.

So if I'm not videotaping, I think the quality of the training has to be very good, so the analyst pursue appropriate data.

MS. KARMAN: I just wanted to bring up I think that this discussion that we have had about the questions that Dave raised is really important for us. I would like to see that highlighted in the report from this exercise. Because how we go about addressing things like what do you do with inconsistencies in the reporting? I mean, we're going to have to deal with that, not just with this instrument, but with a lot of the work that we're going to do internally and in development.

Two things, one, the way in which we determine what types of ways that we go about resolving this will have to come from our requirements, the legal standards that we come up with, and our scientific standards we come up with. So, obviously, we are not going to resolve this today. Just really excited to hear this kind of discussion.

Also, it sounds to me like we're talking

about -- you know, one of the problems I have with central tendency issues involve, well, now, I'm sorry we brought that up. That may very well mask information and details that we may very well need.

On the other hand, you know, we do need to get to resolution with things. We have to have a protocol for that. And I do think that, you know, possibilities may include reporting ranges, as Deb mentioned, you know, mixed methods. So that maybe we combine, you know, sounds, qualitative, quantitative that can help us we resolve that.

But truly I think that the impetus or the manner in which we decide we want -- how we want to resolve this, and how we want to build that into our protocol should be coming from our scientific legal standards. So I'm thinking others have something to say about this.

DR. PANTER: There is a very nice segment of literature on all the different ways that we can resolve this, the indices that will assess consistency in some way. So we will definitely want to look into that literature, and also think about

all the other ways, not just quantitative ways of approaches to use; but there is a good base for SSA.

MS. LECHNER: That's comforting.

DR. BARROS-BAILEY: Dave has one comment. We are over time. So I'm going to go ahead and defer to David, and then we will go to closing for lunch.

Okay. Go ahead, Dave.

DR. SCHRETLEN: So I appreciate that ultimately where we go with the job analysis may bear virtually no resemblance to either of these. And so, you know -- but I still think that just by talking about this experience, no matter what the instrument looks like at the end of the day, we can address many very important issues. And I could easily see spending a few hours just talking about the experience you guys had, and tabbing what you learned from this, so that we as a Panel could learn. And I'm just wondering if there is some mechanism whereby we can have more time to discuss this at some point, because I think this is a very useful exercise for this Panel to think about.

DR. BARROS-BAILEY: Thank you, Dave, for that suggestion. I have been writing as we have been going through this process, because I think that some of the -- of the benefits of this experience to the Panel moving forward.

We're about to break for lunch and come back at 1:00 o'clock. I do want to acknowledge something that I had heard -- a couple things that I heard this morning. A couple words. One of them was what benefits can we have from this process? What can we, as a Panel, bring to this? That might be more -- I would like you to kind of have that in the back of your mind when we go to some of the discussion and some of the deliberation this afternoon at the very mackerel level.

And also, I think, Tom, maybe one of the things I was hearing you say, or infer by the questions that you were asking, how long this took and that kind of thing, even at the micro level, what are some efficiencies that we might be -- that we might be able to recommend in some of our processes or thoughts and recommendations and advice

back to SSA.

Those might be good filters for us to consider as we go along this process.

So we are at -- I'm still on Idaho time -- 11:11 right now. We will be coming back at 1:00 o'clock. Have a good lunch. Thank you.

(Whereupon, a lunch recess was taken and the proceedings subsequently reconvened.)

DR. BARROS-BAILEY: Good afternoon. I would ask everybody to please take your seats. Thank you.

I would like to open the afternoon by saying that we're going to have a presentation that I'm looking forward to. I have heard about this project. I know that we have had somebody from this project at every meeting that we have had the Panel from the very beginning.

Before I introduce the project and the members presenting, I know that they are going to be talking about the ICF, International Classification of Function, and there might be some confusion with ICF International that is -- has the BPA in terms of

the field job analyst project.

So I'm going to defer to Sylvia real quickly to have her explain the distinction so there is no confusion that we're talking about the same thing.

MS. KARMAN: Thank you, Mary.

Yes, at the risk of my being perseverating on this, as I mentioned this morning, it is -- may be a point of confusion for individuals, especially people listening, that our project right now has a blanket purchase agreement, which is a contract with an organization called ICF International to assist us in developing a business process for job analysts -- for recruiting training and certifying job analyst. Of course, there is the International Classification Function, which we're now going to hear a presentation from NIH and Boston University, and that's certainly going to come up.

Also, while I'm on the subject of ICF International, I think that during our morning session when we were covering the work that was done by the Ad Hoc group for job analysis, I think there

was some discussion about instrumentation and the types of task that ICF International will be performing for Social Security under this BPA, and it does not involve developing instruments. It is only to assist us with benchmarking job analysis approaches and helping us develop a business process for recruiting, training, and certifying.

DR. BARROS-BAILEY: Thank you. At this time I would like to welcome our presenters -- our four presenters. As I mentioned this morning, I had requested that we have a presentation by the collaborative project between Social Security Administration and the National Institutes of Health that was mentioned along with the OIS project in the NPRM listing.

Part of the reason for my interest in this project was that very mention. Therefore, anything that we can learn that may lend ideas or data elements, scaling, research methodologies, anything about the project that might be helpful to our process it is of interest. I understand this is a very exploratory research project; whereas, the OIS

development is applied to the work side, and asks us to understand the different purposes of both projects.

We're going to have four presenters. I would like to welcome Mark Spencer. He is the Associate Commissioner of the Office of Disability Programs, ODP, in the Office of Retirement and Disability Policy.

And if you look at the fourth red divider in tab two, we have very detailed biographical sketches of each of the presenters.

I would also like to welcome Dr. Beth Rasch who has been part of our meetings here for a long time, a familiar face to us. She is chief of -- Staff Scientist and Chief of the Rehab Medicine Department, NIH Clinical Research Center.

We also have Dr. Stephen Haley, who is Associate Director of Health and Disability Research Institute at Boston University, School of Public Health.

And Beth Barfield. She is a Pre-Doctoral research fellow at the Health and Disability

Research Institute.

Welcome.

MR. SPENCER: Thank you very much. Again, my name is Art Spencer. I am the AC for the Office of Disability Programs. The best way to think about ODP is disability policy and DDS. My brother in arms is Richard Balkus. We work very closely with him on many of his efforts. He helps us greatly on many of ours. We are proud to be providing some staff support to your work, and we're going to continue to do that, of course.

In August of 2007 we sought advice from the National Institutes of Health on new technologies, diagnostic tools and models that might help inform the disability evaluation process. The NIH Rehab Medicine Department suggested that we look at innovative ways to assess functioning across the spectrum of abilities. This discussion led to two parallel tracks and an inter-Agency agreement with the NIH Rehab Medicine Department.

First, an analysis of the existing SSA data. And that's proven very helpful to us and

meaningful particularly, and that work continues. We're not going to spend a great deal of time on that that day.

Then, secondly, the assessment and feasibility of developing computer adaptive testing instruments or CAT instruments, that could be integrated into SSA's disability evaluation process.

The focus that we are looking at here is on function. In the DDS world and in the ODAR world, i.e, the adjudicative world, we need function to assess many of our medical listings. We need functional information to develop residual functional capacity. And both of those lead to -- eventually to a decision to allow or deny benefits.

The bad news is, is that there is no easy way to get that. The best way, or the way it's often done now is, "Ms. Harmon, what do you do during the day?" Then we transcribe what's said. There is very little else that we have. We take that and try and translate that through the medical impairment that they have to come to a residual functioning capacity, or to prove the existence of

those functional limitations that might demonstrate that a listing is met. But there is, at best, some pretty weak correlation.

So working here with this group and now we're entering -- our third year?

DR. RASCH: Yes.

MR. SPENCER: Our third year. I think you are going to see this afternoon some very interesting approaches, some of which -- and this is, of course, with our work with Sylvia and with Richard -- some of which might inform what you do; and also ODP would be more than happy -- if it does seem an interesting approach, we will be more than happy to try and find a way to expand our work so that we can support you even more fully.

At this point I'm going to step aside and let the experts talk; but, again, thank you very much for being some wonderful hosts to my people. They also participate in your work and have done so from the beginning. Please rest assured that help to you will continue for as long as you like it. Dr. Rasch.

DR. RASCH: Thank you. I will just make one minor correction. I'm the Chief of the Epidemiology and Biostatistics section within Rehab Medicine. The chief of Rehab Medicine is Dr. Leighton Chan, who I thought I saw here today.

Yes, there he is. He is holding his hand up.

I want to tell you that we are incredibly pleased to be here today. My colleague from NIH, Diane Brandt (phonetic), and I have attended every single meeting of the Panel since the very first meeting, because we feel your work is so related to ours. And we have been watching with great interest your activities and your deliberations, and we want to stay informed of your work. So thank you very much for the invitation to be here today.

I would like to just clarify, as Mary said, this is exploratory work being conducted by NIH and BU, examining the ways in which claimants and their health care providers can quickly and easily provide information about the claimant's functioning. SSA has not yet adopted or endorsed

this approach. I want to make that clear.

In August of 2007 SSA came to NIH seeking identification about new diagnostic tests and novel assessment approaches that might expedite allowances. The SSA paradigm at that time, just by virtue of the question that was asked, what new diagnostic tests might expedite allowances -- the paradigm was one of diagnosis or impairment relating to disability. And we suggested that examining the capabilities of individuals within the context of workplace demands might be a more truthful line of inquiry. So we entered into an inter-Agency agreement in February of 2008. We signed a new agreement this past February of 2010; and it's five year agreement. So we have work outlined through 2015.

We have two broad objectives, as Art mentioned, to improve the determination process, including data analysis and development of computer adaptive tests. We have been given access to an unprecedented volume of data from SSA. And that has allowed us to understand the basis of award

decisions; to understand the problems with the process, such as the large number of decision reversals that happened during appeals. And it has also allowed us to develop data driven systematic approaches that will allow SSA to inform their decision making process.

So in other words, we're delivering to SSA analytic models that they can use on their own data to help support programmatic decision making.

On the flip side -- NIH is taking responsibility for that portion of the work.

On the other side, Boston University is leading the effort on developing computer adaptive tools, which are essentially -- it's intelligent software. It's been applied in educational settings for a very long time. And more recently, they have been applied to the assessment of functioning; and Alan Jette and Steve Haley are really leaders in the field in this area of applying CAT methodology based on item response theory to measure functioning.

And Steve is going to be spending the majority of the time today talking about CAT

methodologies, and the steps involved in that process. So I won't spend more time on it today. But again, the thought is that we are focusing on function.

So when we started out thinking about this work, and I eluded to it before, we start with how disability is conceptualize. And this has really evolved in the last 50 years. So early models, often termed the medical model, attributes disability to an individual. It's a personal trait as if -- similar to hair color or gender. However, if you look at the person in the upper right slide he has limb loss. But I would propose that if you examine functioning, that is skiing in this particular slide, I would be the one disabled compared to him, because I absolutely could not do that, what is being depicted in the slide.

On the other extreme is the social model, which indicates that disability is a socially created problem. So in other words, if a person in a wheelchair encounters a flight of stairs, is the problem that they can't go up and down the stairs,

or that the stairs exist? The social models would say, it's because the stairs exist. They're architectural barriers, socially created problems that prevent people with disabilities from fully participating in every day life.

So contemporary models really integrate both of these perspectives, and they view disability as the outcome of the interaction of the capabilities of individuals in the context of environmental demands. And in this case it would be the context of workplace demands.

There has been a long history of development of these models starting with the work of Saad Nagi in the 1950's. The models have been iterative in some respects. In other words, one builds on the work of previous models. The most recent model is the World Health Organization International Classification of Functioning, Disability, and Health. That's the full title termed the ICF. It's precursor for the ICIDH and the ICIDH-2. This is just to say that there has been a lot of work done in this area of

characterizing disability and functioning that's taking place over probably the last 50 years.

So this is the International Classification of Functioning, and you can see that there are several major domains. So -- and I will just give you some examples to orient you. So a health condition would be, for instance, a stroke or a spinal cord injury that occurs at the cellular or tissue level. Body functions and structures, the negative aspect of that would be termed impairment. So it occurs at the body system level, and it would be things like weakness, cognitive problems, blindness; those types of things.

At the next level, the activity level, those are tasks or actions conducted by an individual. It takes place at the whole person level. So it would be things like standing, bending, stooping, walking, reaching, lifting, the types of things that you -- the Panel often discusses. How those activities combine with environmental factors, and personal factors yields participation. That's how the person operates at a

societal level. That would be work. That would be running a household and various roles that people hold in society. So think about work as being participation.

The SSA paradigm has primarily been to measure health conditions, and body functions and structures to make determinations about whether or not people can work. And NIH has proposed that the measurements should take place at the level of the whole person measuring activities in the context of workplace demands to understand capability to work.

So contemporary concepts of disability depicted as being interactive. Remember, I was just mentioning that it's the interaction of the individual in the environmental context, rather than being an individual attribute; therefore, diagnosis alone is not a good predictor of disability.

Now, there is some instances that it can. For instance, SSA compassionate allowance program includes a list of conditions where people are highly likely -- where there is early mortality, where they are highly likely to have rapid

functional decline; and there are certainly conditions for which this is true, where people are highly likely to be disabled or not to survive for a long period of time.

However, it's a very small proportion of people who apply for benefits. So the problem becomes one of this interaction, examining capabilities in the context of the workplace environment. Disabilities multi-dimensional instrument, as you just saw in the ICF model. So there are many different conceptual components that constitute the definition.

It occurs along a continuum. So, although, SSA makes a determination that somebody is or is not disabled, in fact, disability occurs along the continuum of functioning; and it's dynamic. Based on contemporary definitions, if the environment is part of the equation in supportive environments people may have no disability who have profound impairments. In a less supportive environment they can be disabled. And disability can also change over time. People recover and

decline, as you know.

So measuring such a concept poses a methodological challenge. And in order to operationalize the concept we need to measure both individual attributes and environmental features. And as was mentioned this morning, the operational definition depends on the purpose of data collection. So if you are collecting information about people with disabilities for the purpose of providing accessible housing, you are going to collect very detailed information about the types of limitations that individuals have, and the types of architectural features they may need in their home.

If you are collecting disability information for the purpose of equalization of opportunities for civil rights legislation, you are going to adopt a very broad definition of disability so that as many people as possible are included in the legislation. So we really need to think about the purpose of measurement when we operationalize the definition.

SSA's definition, based on statutory

regulations, has been to identify people who are unable to engage in substantial gainful employment due to medically determinable physical or mental impairments that is expected to result in death or last 12 months; and are expected to be of such severity that the individual cannot do their previous work, and can't do their work in the national economy. I know you are quite familiar with the definition.

So there is a -- in our view, there is a gap between contemporary notions of disability, and how SSA operationalizes its statutory definition. We are not suggesting that SSA should change its statutory definition. We are suggesting that the operationalization of that definition change. Because the current operationalization is focused on physical and mental impairments, which is -- harkens back to the old medical model. While contemporary models depict disability as the gap between what individuals are able to do and their environmental demands. So as I mentioned before, diagnosis and impairment may be poor predictors of work

disability.

So the classic nonexample is two people can have loss of a finger. One is a teacher and one is a concert pianist. For the teacher it has no effect whatsoever. For the concert pianist it's devasting.

When Art Spencer first heard me give this example, he said, are you familiar with the well known jazz pianist who was badly burned and was able to continue his job as a pianist, quite well known person.

I said that is a perfect illustration of the problem. Because people have amazing resilience. And they can have impairments and continue to function. So you want to look at function. You want to look at what people do at the whole person level. How they play the piano, not at the impairment, which would be loss of range of motion, that type of thing, or it could be dexterity.

So measurement. In order to measure whether or not -- the components of whether or not

people can work, we really need to examine activities at this whole person level in the context of workplace demands, and the aspects of the workplace environment to yield decisions about whether or not people can work.

So activities are, as I mentioned, things like bending, standing, stooping, whole person activities; and we're suggesting that measurement takes place at this level.

I am going to toggle back and forth between two slides.

So the ICF actually does not make a distinction between activities and participation; and that is because there was a huge disagreement among the committee that developed the ICF about where to draw this line. So there are nine domains that comprise activity and participation; and I am showing you six of them here.

So based on the definition of activity, and NIH's opinion we would consider these to be more oriented toward the person. Things that people do at the whole level, rather than the person in

society. So these are interpersonal interactions and relationships, mobility, learning and applying knowledge, communication, self care, and general task and demands, which are things like multitasking, or organizing time, materials, and space.

The other domains are much more oriented toward community and civic life, and working, and activities in the community, which we would consider participation.

So having said that, we have coded all of the information that SSA collects through their forms and evaluation processes and have found that use of the ones through the ICF and examining specifically the activity domain, the very limited information is captured on learning and applying knowledge, general tasks and demands, communication and interpersonal interactions and relationships. Yet, these are areas that are critical to work, and I know that the Panel has discussed this in previous meetings.

So NIH and BU prioritized two domains for

CAT development, interpersonal interactions and relationships and mobility. We did this for a number of reasons. The first is because in the mobility domain SSA already collected a substantial amount of information about mobility, and we were able to build on that. There has also been a substantial amount of work on CAT development in the mobility domain -- or the physical demand domain. So we felt that we would have the best shot at developing a tool that could be feasibly implemented in the mobility domain.

There has been far less work done in the personal interaction domain. Yet, it's of great value to SSA, because they have had difficulty adjudicating cases where people have mental health problems, and the applicant constituency has changed so that more and more people with mental health problems are applying for benefits. So we thought this would be of great value to SSA. So we choose to start with these two domains; although, our plans are to develop CATs for all six domains.

So we made several recommendations to SSA

at this point, and I will capture the major ones. We recommend the whole person approach to capture all the conditions that applicants report, because the sum total of the impact of these conditions on functioning is what is critical to work. So we know from national data that the majority of working age adults have one or more chronic conditions. The majority of working adults have one or more chronic conditions.

People continue to work as they accumulate chronic conditions. And at some point they develop a condition that causes them not to be able to work. When they come to SSA for benefits they have -- they may have functional limitations due to all of those conditions or some of those conditions. And it's the sum total of how those conditions affect the individual that really would allow them or prevent them from working. So our recommendation is to evaluate function comprehensively and to capture all of the conditions that people report when they apply for benefits.

We are also recommending a focus on

functioning. This is data from 2005 from what is called the disability waterfall. So this is SSA data indicating that in that year there were 2.6 million initial applications. Many of those denials went on to reconsideration or to the ALJ level of disposition. Of those who went to appeal, 62 percent were allowed.

So the concern is what's happening in that process to cause decision reversal. And while SSA certainly works very hard to collect information about functioning and to evaluate workplace demands and to look at that interaction, we feel that the most detailed information about how people function in the context of work comes to light at the appellate level. And that if we could collect comprehensive, uniform information about functioning early in the process, that it would allow SSA to make more informed decisions very early in the process. And it really hinges on having a type of assessment that's very quick, comprehensive, uniform.

So we feel that this work has usefulness

to SSA, because it could dramatically improve the breadth, the completeness, the uniformity, and the precision of the medical evidence. That we can collect data when it's really most useful for decision making. And really even small improvements in the process may lead to reduced processing times, improved accuracy, uniformity of decisions, and reduced blacklogs.

So, again, this is exploratory work. It has not yet been endorsed by SSA. And if you haven't gathered by now I'm really introducing the project, and then Steve Haley is going to talk in detail about the computer adapted testing. So I would be happy to take questions on this portion of the talk if that's appropriate.

DR. BARROS-BAILEY: Sure. I will go ahead and open it up to the Panel to see if anybody has any questions at this point.

DR. SCHRETLEN: Thank you, Beth. That was a wonderful overview. Are you or Dr. Haley -- are you going to address the kinds of domains that are suitable for CAT technologies? I'm wondering what

sort of abilities are suitable. I assume that you --

DR. HALEY: Well, we have tackled one of each in this first group. We think physical demands is going to be more concrete. Some of the social, cognitive or interpersonal interactions we're doing are more difficult to scale; but they have been done in the past. And we're going to make a good effort to make sure that they are acceptable for CAT use.

DR. BARROS-BAILEY: Great.

Thank you, Beth, for that introduction. And Dr. Haley.

DR. HALEY: Thank you very much. It's a real pleasure to be here. It's very exciting for me to be able to have this much time to talk about our CAT project, using this amount of time. So I'm glad there is interest, and we're very happy to be here.

I just want to let you know there are certainly many others at BU who are working on this project. Alan Jette, who many of you may know is Co-PI. We have Karen Bogusz, she is our project director. Our training director, Mary Slavin is

involved. We have a really great team of IT analysts or CAT programmers directed by Pengsheng Ni. Beth Barfield, our student is here.

We also consult with Ron Hambleton, who has been doing CAT and IRT work and education for probably over 30 years. He will look over our shoulder and make sure that what we do is sound and good.

Then we also are in our calibration work, which we will talk about in just a minute. We have subcontracted with Westat, and Bill Frey is leading the effort. They are a large survey center that SSA has used many times before. I think they are going to be very good at documenting all the data.

So, again, we want to make sure that you understand that this is exploratory work. It has not been endorsed. It has to be proven in the field, I think, before SSA is willing to adopt it. It is simply a way in which claimants and their health care providers can quickly and easily provide information about claimant's functions. So we are trying to do CATs both for claimants and providers

as well.

There are six parts of the presentation. We're going to talk about the functional domains, and some of the subdomains within interpersonal interactions, and physical domains.

Some of the formative work that goes into building items, calibration study and its plan. And then we will talk a little bit about computer adaptive testing and give you some of the details as to how it works, and how it doesn't work, and interpretation of individual scores. I know there is some sensitivity about scores, but that's what CAT does.

And the source can be used in many different ways. They don't have to be made from decisions. They can just be information. But we can show you how the scores might be possibly compared to the environmental job demands, et cetera. And then we have a small pilot study at the end of CAT development, just to make sure it can be used in the SSA system.

Then I have a question period after each

of these sessions. So if you will allow me to go forward, and then if you have questions they will come after each section.

We said that the two functional names we have chosen for our first feasibility test is physical demands and interpersonal interactions.

And our overall model recognize that there is physical areas, and cognitive, and mental. Here are the six areas that we will be tapping into eventually. The wording all comes from the ICF.

I'm going to take you on a little tour here. So if we look at one of the components of physical demands we will see that there is a change in basic body position component in one form of maintaining body positions. So these sub -- now these may not be in your handouts, because they are animations. So the full thing will be up on the screen.

Each of these areas was examined by ourselves, content experts, et cetera; and we built items around each of these content areas. So there are standing items. There are sitting items,

bending, squatting. In terms of whole body position there is walking, moving around. Moving around involves either the use of walking device or a wheelchair.

We debated for a long time if running is something necessary; but we did think of some professions that do require some running, policemen, fireman, et cetera. And then certainly carrying, moving, and handling objects with a lot of items for hand use since many people don't have to get around too much in the workplace, they just stay at a desk.

So all of these areas were identified as key to potential work jobs, work place environments. So we wanted to ask questions from a claimant as to how they did these types of items. So that's our physical demand model, content model.

We will take you through interpersonal interactions. This is much more complex. Now, this was not done just because we thought it was right. We looked at the literature. We have identified content experts. We have a lot of feedback into what components really made up our personal

interactions. So all of these components were considered essential to trying to build some content into this task. So questions on trustworthiness, and do you trust others, et cetera?

In terms of behavioral modulation, a whole series of areas that we thought were important in the workplace. These mainly had to do with control of workplace behaviors.

Now, we put in a -- some items regarding adaptability. You know, they may fit better in general tasks and demands. We're not sure, but we felt like we wanted to test them with this set first. So this is -- these are response to change and coping with stress.

Now, the good thing about this if they don't fit this -- these types of items within a background model, we can pull them out, save them, and see if they fit with others.

Then basic interactions includes a number of areas that we want to ask about. So this is a much more complex model, and -- now, we have -- we know that most of the people that are claimants have

not worked for a while. And so we did want to build some behavioral items that reflected back to either their very recent work or their current work. So this is a real short set of items, but it gives a sense if they are currently working now, you know, what are the issues?

One of the nice things about the computer format is if you ask them to work them out, and they say "no;" then these will be filtered out. They wouldn't show up.

Okay. So that's interpersonal interactions. Questions.

DR. GIBSON: I have two actually. Looking at what you referred to as your content model I'm just curious mostly about the ones on interpersonal. I found it interesting -- I understand why -- but I wanted to clarify why you included personality constructs as a part of function? Because personality is typically not perceived as something that is a limitation, because it's something that is stable over time. Although, it predicts the type of work you may like to do it, doesn't predict whether

you can or can't work.

DR. HALEY: I think it's more of an extreme personality trait that we're looking for. I don't have the items with me, but I would be happy to show you the items in that area.

DR. GIBSON: Thank you. I find that interesting, because big five items is what I was looking at. It looks like a big five factor model scale there. Then I also just had a question regarding what you referred to as your content model on the physical side. You talked about development of items on each of these. Can you share with us how many items you had on average for these, and the nature of those items? Because I think it would be very helpful for our work, to continue SSA's work on development of our content model further.

DR. HALEY: This will show up in the next section. I have it on a slide, so I thought I will give you the real numbers. We started with five or 600 items in each area.

MR. HARDY: I have a quick question on your interpersonal interactions. Later on do you go

into more detail on some of these things, such as agreeableness? Are they defined, and this is further up the road?

DR. HALEY: Not in this talk. But we can certainly show you the items that we have built in the area.

DR. RASCH: I am sorry. I will just mention that we have delivered -- the NIH deliverables to date, and the BU deliverables for Sylvia and to Mary for Panel use, and for use within OPDR. So all of the questions that are going to be tested that are in the current calibration test, all of the definitions, much more detail about the development of the content models and that process are included in those deliverables; and so the Panel will have access to all of that.

MS. LECHNER: My question has to do with -- a little bit with your process for developing all of your functional domains and items. I guess that's what you are about to explain.

DR. HALEY: Yes, I will try. If we don't, there will be a question section afterwards. Yes.

DR. HUNT: I am not intimately familiar with the ICF. How deeply into this structure are you still following the ICF -- I don't know -- pick one of them? Is it the third level, carrying, moving, and handling? Is it down to all of those? Are we still within the ICF structure?

DR. HALEY: Well, not fully. At some point if it made sense to alleviate, you would; but some of the major categories we tried to keep.

DR. RASCH: I will just add to that. The ICF was the framework that we started with. We choose it because it's the international standard for examining functioning. It gives us the common language to define functioning. What the ICF contributed were, in addition to defining and delineating the major domain of functioning, it went into great detail, as you can see, to define the subdomains within this domain. So that was a good starting point for us. But then, as Steve mentioned, accepted literature review and content expert's input drove the addition of additional subdomains.

DR. HALEY: Just one more point.

DR. PANTER: I am just wondering if we have access to kind of look at your methodologies or content model, because I think that will be useful.

DR. HALEY: Yes. Yes.

Just one more point. We will let the data drive us as to what domains are unidimensional, and what we can pull out. I hope we have multi-factor solutions. I hope we have even two factors, or even a large factor solution that we may interpret; but we will see.

Any other questions?

DR. BARROS-BAILEY: I was just going to say I think everybody is done with questions. Dr. Haley, you may move on. Okay.

DR. HALEY: Thank you. All right. The formative work in building these items around these content areas. I'm going to give you an overview, and this methodology has been used pretty uniformly now in major projects that are building CATs. Certainly, problems in your overall, and most other projects that are building CATs, are using a very

similar process. And there has been quite a bit of work and applications actually on methodologies that's available.

So we have catalogued all the items we could find in physical demands and interpersonal interactions from every instrument that we can find -- we could access. So that, hopefully, will give us the universe, at least in the past, of how people have tried to ask these questions.

We have done a series of focus groups where we have asked claimants and providers what they think are important areas to ask. And we have built new items, thinking that there are gaps. And we have had content experts help us build those new items. So these are people who have done instruments in the past in the areas that we're looking at. Then we develop initial item pool, which is evaluated by the experts and claimants, and particularly by cognitive testing, which we will talk about in just a minute.

Now, at this point we have a lot of volumes. So we have to narrow it down quite a bit.

Then we finally get to a point where we have an item pool that we're going to go out and test. We will call this a calibration phase. We send these items out and claimants and providers answer them. All of them. It's about a hundred or so. So that we get data to help us understand how these items fit together. Then we do our analytic hocus pocus and come out with models and scores.

So we have extensive literature review, and focus groups. We had four content experts actually in physical demands. We gave three for interpersonal interactions. And we did a series of cognitive interviews on all the others.

Cognitive interviews, if you are not familiar, it provides a sense of how this item sounds to the claimant or provider, and try to get -- to try to understand errors that we don't see either as professionals; or the people not understanding words the same way as we are. So it asks the participant a question; then it asks a series of pros. Like in your own words, what do you think this questions is asking? How competent are

you in your answer to this question? Can you think of a better way to ask this question? Well, very often they can. How do you like that answer? Was it hard or easy to answer that question.

So with that feedback, then we go back and we revise the items, and the content experts are involved in this process as well.

Now, the cognitive interviews did make a big difference. There were at least ten items in the physical demands that were completely rewritten; and seven in interpersonal. And it -- the whole process gave us a sense of, you know, the kinds of words that we could use and should used to avoid misunderstandings.

So some of the items, for instance, that are in the physical demands, are you able to lift a 20 pound object from table height to a high shelf? Now, this was done to represent one handed lift to a medium height, to a high height, which is one of the kinds of job demands that might be necessary in certain circumstances.

How far are you able to walk without

stopping.

How quickly are you able to walk?

And then, are you able to walk over -- to work overhead -- this is an item -- for 20 minutes, like organizing a high shelf in a closet. Some people have demands where they have to work on higher levels for some time. I don't know about you, but after 20 minutes if it wasn't organized I just quit. Sometimes you get pretty tired working overhead.

In terms of the interpersonal interaction items, these are just sample items. I feel good about myself. I am so tired when I wake up. It's hard to get going. I get back on track when I am distracted. That's a little bit more relevant. I can't stop myself from doing the same thing over and over. I have difficulty calming down. I get in conflict with others. That's how we're trying to ask the questions.

Now, we read your stuff and some of the things were very helpful to us, because in your recommendations you indicated that there ought to be

an assessment of repetitive items, varying force requirements, duration of typical day, how many hours a day, balance items, reaching levels, which you just saw an example of; and unilateral, bilateral.

So we had to do -- what our content experts had to do is figure out what would be unilateral lift of something that was common to somebody, you know, from one height to another; and that's how we built many of the items.

Now, repetitive item. That was a little bit of a struggle; but an example would be if you drop cards on the floor, and they were spread out, and you had to pick them up. It wasn't just one bend; it was multiple bends. And to some people they told us at least that was really difficult. They could bend once and pick something up; but if they had to bend over and over again to pick things up, that was really a struggle.

So many of the items are couched in the content that we want to get at. They are not couched in the workplace, because most of these

people haven't worked for a while. So they have to be fairly common items that people, even if they are not in the workplace for a long time, would be able to answer.

And then the content areas addressed your recommendations. Was it an interpersonal function? Now, we didn't put in too much -- too many items on an issue in resistance. We're going to leave that for a general task, I believe. Then the neuro-cognitive items we're going to leave for learning from the client and knowledge.

So we initially started with 361 items, interpersonal interactions. We also took items, shamelessly, from NeuroQOL and PROMIS. The reason for that is we wanted to eventually link the two up. So if you have core items in one instrument, and the same core items in another one, you can do a linking function. So if you want to know how they score on PROMIS, and they have actually been administered this tool, you can make a link. You can say 40 on this instrument equals a 60 on PROMIS, whatever. Now, there may be some benefit for that in the

future. There is a lot of interest in how instruments link up. With our IRT analysis, it is very possible to do that.

So physical demands we started with 174. We also have PROMIS, NeuroQOL items. And we ended up -- as you can see, a larger number of interpersonal interactions, because we -- you know, we are not sure. We know we're going to have to throw a bunch of them out that don't work. Then we have provider items as well. Although, the provider items we cut down, because we are going to be lucky to get a provider to rate 90 items.

Physical demands we have 124, and 97 provider items. So that's what will go out for calibration work is that last set.

Any questions about that? Yes.

DR. GIBSON: I have two. Thank you for being patient with me. My first one is a philosophical one. One of the things that SSA will eventually struggle with is the linking of work analytic data to their measures of residual functional capacity. Do you see IRT methodologies

of having PROMIS for being able to do those linkages as well?

DR. HALEY: Well, I do see the possibility of linking these skills. These physical demands to job demands. And it's not necessarily an IRT process. It's really a consensus process by experts. It's called benchmarking; there are a number of methods; but it's a fairly structured agreement process where experts get together and they look at the skills on a continuum. And they compare it to job demands as people -- you know, then there can be some linkages.

DR. GIBSON: My second question is, looking at your slide number 34 where you describe some of your sample items, I'm curious what influences your choice of the different scales or measures that you utilize there. Some of your scales are very objective in nature, and some of them would be described as more subjective measures when you ask people to compare their capacity, for example, to someone else. So what entered into your thoughts? What references did you use? If they're

in the report, you can just tell me they are in the report, and I will look there. But I'm curious what influenced your scales.

DR. HALEY: I would say 95 percent asked about rate the level of difficulty. We just wanted to show you some other possible scales. The one scale that does refer to other people is to what extent do you walk? Have you walked against the speed of others? We found that item in other instruments to be very effective and clear. So that's why we included it.

DR. BARROS-BAILEY: Deb, and then Tom.

MS. LECHNER: You talked about the up and down process. I don't know what slide number, but it's on our page 15 where you start with your initial item total. You had like the 361 interpersonal items, and 174 physical demands. Can you talk to us a little bit about that dwindling down process. Because I know that SSA internally is working on a similar type of dwindling down of items. I'm just curious about how you all went about that dwindling down process, if you can speak

to that a bit.

DR. HALEY: Well, through a series of meetings we got the content experts together and ourselves; and we tried to organize the items in certain ways. We looked at it by content so we got things out that were too close in content. We did it by what coverage in the scale we anticipated that item to be in. So if we had too many easy items or too many hard items that influenced our decision. Because eventually, we want a hierarchy of all the items, and a good spread across the continuum.

I think those were the two major factors. And if we felt certain items just weren't worthy of the bite, you know, if we had other items around it that we felt were similar, we would eliminate it.

MS. LECHNER: Then kind of a follow-up question. Have you had any pilot studies where you have compared the self-report to actual performance on the functional testing?

DR. HALEY: Well, not within this. That should be done eventually. We have in other work that we have done in practice studies and others.

We, you know, found pretty reasonable correlations, point eight; point seven. They are not huge, but they are not -- you know, they are usually pretty good.

MS. LECHNER: And if you get a discrepancy between, let's say, a claimant report on these items and the provider report, how do you reconcile the differences between the two?

DR. HALEY: Well, that's not our decision. Certainly -- that's not our decision. It's certainly something we could advise on in terms of what SSA wants to do; but I think this is -- a real advantage of this approach is that you will be able to compare more easily, I would say, what a provider and a claimant says. And how people -- that could be just information that's brought in. It could lead to a decision of additional testing. It could, you know --

MS. LECHNER: And then as a follow-up, have any of the folks that you have used this protocol with been claimants or people who are applying for any kind of Workers' Compensation or

disability claims?

DR. HALEY: Not to date.

DR. RASCH: However -- I will just say, however, he is going to talk about it in a moment. We're just about ready to field a large calibration study where we will be testing the item pools on the applicants and their health care providers. And we're also trying to capture secondary health care providers, so that we get a better understanding of how well providers are able to answer questions about a claimant's functioning.

Because our premise is that therapist and other types of health care professionals might know more about claimant's functioning than, say, a primary care provider. And so we're trying to gather that data.

DR. BARROS-BAILEY: Tom, then Abigail, and then Shanan.

MR. HARDY: I will be very quick. Deb asked two of my questions.

I don't know anything about your report. I can't wait to read it. Sounds great. Back to

page 15, the items and instruments in that blue slide. I think I heard you say, quote, items from the universe of instruments. In your report you go into detail as to what instruments you have looked at and some commentary under that. Is that in there?

DR. HALEY: Yeah, the entire list of instruments that we examine is in there, and references.

MR. HARDY: And that would include anything dealing with the emotional, cognitive, all of those are as well.

DR. HALEY: Those are all there.

DR. PANTER: I just wanted to remind everyone that because it's such a high stake setting, this is now -- needs to be considered -- I think you are probably considering this all along -- but this would be different than any kind of research setting where you might use these measures, or even some of the clinical settings where you might use the measures. So I would like to just emphasize that being high stakes, the CAT

methodologies also needs to consider the issues that high stake settings required for CAT and security issues.

DR. RASCH: So I would just say that we're viewing this as a decision making aid for SSA. It's a tool that would help augment decision making that's already being performed by the individuals who are responsible for that within SSA, and simply that. The kinds of issues that you are talking about are really policy decisions within SSA, whether it would be implemented, how, how it's used, et cetera.

DR. PANTER: I agree. I'm just saying that it brings with it many different requirements of important implications. So it's just something down the line to be considering.

DR. HALEY: Yes. We appreciate that. Okay.

DR. GIBSON: My last question on this section, I promise. Following up on Deb's question about dwindling down of the item pool. So do I understand correctly that the decision was purely

rationale? There was so consideration of psychometric properties of the items at that point. And that that won't occur in terms of looking at the item's strengths and weaknesses, methodologically or psychometrically in the calibration phase when actual data is collected?

DR. HALEY: That's correct. We have no data right now. Once we get to calibration data that would allow us to decide what domains are really being covered here, and if the items really fit the scales.

DR. BARROS-BAILEY: Sylvia.

MS. KARMAN: Hi, thank you, Steve and Beth, for both coming.

I just have -- in this area I have one question about how we -- how you are going to determine the -- as I see, you did the cognitive interviews. I am sure that was intended to get at whether or not the people were understanding the questions correctly; but I think, you know, one of the items that you mentioned, you know, after 20 minutes of reaching -- you know, are you able to do

work reaching overhead after 20 minutes, like organizing closets. You kiddingly said, you know, after about 20 minutes I would give up. I am thinking yeah, I would give up too, but for a different reason.

So possibly what your question or one's question is trying to get at I know we will be faced with similar things when we look at items, or we develop items. You know, we're thinking that the item is getting at a particular question, but we're not sure. How is your study going to get that? So that's one question.

Another one I had was some of the elements that you had pulled from the Panel's recommendation and physical demand, some of these things we saw as measures, not as, I guess, content areas specifically. Like I see duration of a typical day or varying force requirements. So I don't know if that's just a definitional thing or what.

DR. HALEY: Well, it -- well, let me answer your last one first. We did look at your report. And we thought that was very helpful. We

knew that if we were to build repetitive items or different weights we had to do it in the context of the question. So it was the only way we could really approach it. We didn't want to do separate scales.

So there is a 50 pound -- series of 50 pound items, five pound items. So there is a whole series of those at different heights and lifts and things like that to incorporate different levels of functioning.

DR. BARROS-BAILEY: Tom.

MR. HARDY: Very quick question to you, Sylvia. I'm just trying to figure out where this fits in the world.

Should at some point SSA be interested in utilizing this, would this be kind of like an augmentation to the activities of daily living sheet or something like that? How would this kind of fit into the program? I don't quite get that. Maybe I shouldn't.

MS. KARMAN: I am still waiting -- I'm going to want an answer to my question too. But I'm

not sure. As I think both Art Spencer and myself when we introduced all of this presentation, it is really exploratory. The Agency knows that it needs to explore ways to obtain better functional information from claimants in a way that is not so onerous to the claimant, and also is not so onerous for the adjudicator to sift through.

So this is just one aspect of that, but it may very well be something that, perhaps, the Agency can use to augment, for example, getting information about activities of daily living; but I don't know that.

DR. HALEY: All right. Repeat your first question.

MS. KARMAN: All right. My first question went really to how -- ways in which you all are planning to determine whether a question is actually getting at the type of information that you are wanting. And I cited the question or the item that you had about, you know, are you able, you know, to work overhead, you know, after 20 minutes. You made a joke about well, I would quit if I had to clean my

closet out for 20 minutes; and I would too. But maybe you and I might be quitting for the same reason, but not the reason that you guys are trying to get at in this question.

DR. HALEY: Well, we hope we orient people towards, you know, the physical part of this. Although, they could stop, because they just get bored, I guess. We did focus groups and particularly cognitive interviews. They said 20 people told us 20 minutes was about the right time. They wouldn't do it more than that. So we had tried originally to get information from the cognitive interviews for the examples, because that's what's so hard.

Now, whether that item is valid, you know, I'm not so sure we can tell unless -- you know, there are a couple ways. We will have empirical data. If it fits a model it should be -- you know, and it works in a continuum that makes sense, then, there is a certain amount of validity to that. You know, we will scale the items from easy to hard. And if it's an item that is put in place that makes

sense, it would be about that difficulty. Then we would accept it for that.

MS. KARMAN: Thank you very much. Another reason why I'm asking this, in addition to -- I know you all need to confront these kind of things, but also because the reason for the limitation -- the limitations, of course, the ideology for our purposes link back to a medical impairment. So that's -- you know, it's just -- I'm also cognizant of the fact that the questions -- your questionnaire is posing to people is to them. So individuals are filling it out about what they perceive their functioning is; and of course the kinds of items that we would be writing would be going towards what we would be trying to determine about work or evaluating about work.

So I understand that's a little different than having individuals completing a questionnaire; but that's one of the other reasons that I was doing it, because I know we are going to need to link the areas in which we are evaluating about work to the particular domains that we're interested in,

especially the things that are not so easily exertable, the cognitive ones. So thank you.

DR. HALEY: Okay.

DR. BARROS-BAILEY: I think I just turned you off.

DR. HALEY: I want to talk just briefly about the calibration study. This is a real challenge for us, because we have to get claimants and providers. We're looking for guidance. This is always the hardest part about building a CAT programming is getting people to respond to all the items. This is not a short form. This is a long form in order for us to calibrate them. So we're looking at SSA helping us with giving us a sample of claimants. Then those claimants will then help us with identifying their providers, geographical diversity, et cetera.

We want to sample 1,000 claimants for each of our three scales, and at least 500 providers, and some supplemental ones if we can. We know we're going to have a lot of denials with the providers. So we are hoping -- 500 would be a good target.

This will then give us information that will allow us to create calibrated banks for each domain, and will give us the information that will allow us to create CAT. So this is fundamental in moving forward with this work.

So we have been working closely over the past few months both with NIH and SSA to develop sampling procedures, and with Westat to help us collect the data. We will administer items by the web or telephone interview. Then we will implement analysis and build bank items.

So this just a little bit more information about our strategy. Westat is going to contact each claimant, the calibration survey administered by web or telephone. And anybody who doesn't have internet, et cetera, we can do it by telephone. So that's just briefly what we plan to do.

So let's talk a little bit about CAT. Now, CATs have been around for a long time; and, you know, they were built in the 60's and 70's in educational testing. So they are not new. And health care has grabbed them very quickly, and has

really found that it might be a great application for many purposes.

In most cases seven or ten items would allow you to get a very precise score of an item bank or an item pool that would have hundreds of items. And the reason is, is that we can select items that are better items to answer than just any item that is on an instrument. And I will show you how that works.

So there is a way it has to be scored. There is an item selection criteria, and there is a stop rule. So we can customize this. If people want to stop at ten items we can stop and get the score. If people want to do it on precision, we can do it on precision. We can do it by standard barrier, or we can do a combination. There is all kinds of strategies to stop the thing.

CAT administers a small sample of items because it relies on information, previous responses. The items that are administered are chosen based on how a person responds; and CAT reduces the number of assessment items needed for an

accurate assessment. So it's efficient.

Now, I'm going to show you how this is based. I know it's kind of late in the afternoon to do this to you. Think of an item that's administered. And this is from the calibration work. We can turn this item into a probability statement. If you look at a person who is scoring low level, around ten or so or whatever dimension you want, the probability of endorsing and able to do, as you can see, is very high. It's near 100 percent. With much difficulty and little difficulty, there are little curves. So as people increase, their probabilities increase in the categories that they will endorse. They will check that weighting scale.

Then as you move further into a person being better on this particular domain, there is a high percentage, high probability that people will say they don't have any difficulty here.

So these are item response curves. And every item have one of those curves associated with it that comes from the empirical data. And it's the

information that guides the score that we have. Is that okay?

So we're talking about probabilities, and each item will again have a very different signature as to what they bring forward to the assessment. Some items will be further to the right. Some will be further to the left, which means if they are further to the left, they are an easier item.

Now, what happens in a CAT is we start in the middle, and we know nothing; then, we have a score around 50. That's a normal curve. And all the things you see up above are items. Now, they are all the same shape, because we were lazy; we didn't build different ones. They will all be different -- slightly different.

If we wanted to administer an item in a CAT we would have an item that we would use typically; and let's say the person sat with some difficulty. There would be a mathematical function in the computer that would create this curve. This is for estimation. And one item we would score this person 44.2 with a standard error of 4.3. So we

could score somebody with one item, but our standard area is pretty big. We wouldn't want to feel very confident about this particular score.

Now, we have the one item, and we have the first prior -- or the norm curve; and we then say, okay, let's have another item. It could take -- any of these items up here -- there is only three, because we only had room -- but think of it as we having 20, 40, 80 items up there to choose from. And each of these items has an information function associated with it, which tells you where that item is most valuable along the continuum.

So the item that is going to be chosen is the one that's closest to the score estimation. So if that item is chosen, that response scale is chosen, we have about three bits of information, and then we have a new score. So we went from 42.2 to 45. Now -- and the score comes from the peak of that curve. That's what the estimation is.

The width of the curve has to do with standard error. The other thing that's happening here is the standard error is decreased, which means

that the estimate is getting more precise.

So let's say we want another item. Again, we're getting a slightly different estimation, but we're getting a standard error that's decreased to 2.5. So this will continue like that. All of these items will be part of your score. Now, after four items, we get a 46.9 and standard error of 2.1. And I won't go any further, but let's say our standard precision level is less than two, we would probably do one or two more items and achieve that.

So those are the mathematics. Those are the things that are happening under the black box of CAT. People are choosing a rating scale point, getting a new score estimation, and then a new item is selected based on that new score estimation.

Now, what you notice there is that all the items are coming from sort of the central part of the item banks. We didn't get an item from the far left or far right, because we didn't need to. It wouldn't have given us any information. Remember how some of the probability things were -- you know, if they are either going to say no difficulty or

very difficult, can't do, it's not going to provide us any information. You have to go after those items that really can provide us with information in the area of the continuum. So the items are focused right where we think the ability levels are. That's what makes it efficient.

And so again going back to this thing you saw, the calibrated items then will give us the ability to do the CAT work.

Any questions about that?

DR. BARROS-BAILEY: Shanan.

DR. GIBSON: Sorry, I am the question person seems like.

If you bear with me I want to kind of walk through and make sure I'm following you in my own IRT for dummies kind of CAT model here, because I have an information curve with my Master's thesis 16 years ago or something.

But I'm going to use an old example with the SATs. I think everybody on the Panel can relate to that. Stop me and correct me if I am getting too simplistic. Someone sits down and take the SAT now,

they are sitting at a computer screen, and they give them a question that's middle difficulty. If they get it right, it then chooses a harder question. If they get that wrong, then it goes to a question in the middle. And it keeps asking questions going higher and lower based on the amount of information each question possesses until it narrows in very precisely on what the score is, or what the person's score should be.

It could be within a standard error, we become certain; or in their case maybe that standard error goes up to "X" number of questions. That's kind of the goal there, to move it up and down, right --

DR. HALEY: Right.

DR. GIBSON: -- until you get there, at least in an academic setting like that?

DR. HALEY: That's correct, exactly.

DR. GIBSON: So my follow on question has to do with methodology that you are utilizing for calibration, coming up to where I was going. Again, if I'm wrong just tell me I'm wrong, because I'm

trying to work this through in my brain.

From my perspective, the development of the calibration for an instrument is based on the assumption of a normal data distribution, because you are finding the information in a normal situation. If your calibration that you are going to be utilizing are claimants in Social Security we can't assume that's a normal data distribution, and how is that going to impact the calibration?

DR. HALEY: Well, in most of our previous work, even though we used disability samples, we still have normal distribution. It -- if it deviates from that slightly, or even a little more than slightly, it's not a real problem. As a matter of fact, a distribution -- I don't know if you are talking about scores here or items. But if you have a distribution that is a little flatter than normal, sometimes it's the actual data. So there is no real assumption of normalcy.

DR. BARROS-BAILEY: Tom.

MR. HARDY: I'm completely lost, but that's okay. I guarantee you I'm going to read

everything, and I will try to catch up and understand this.

So this may be -- I know this is a stupid question, because I don't know what I'm talking about; but I'm trying to sort of figure this out in my head. I get the theory, you ask a question. You are telling us the probability then is to kind of figure out as to what the next question should be.

For the purposes of some of the information we are trying to gather, sometimes we want to know that a person can't do something; and sometimes we want to know the full range of updating. Am I losing something or is that information also sort of captured there or am I lost?

DR. HALEY: Well, think of a ruler or a continuum of some kind, and it is physical functioning; talk about walking around a person's house versus walking in the community. If a person can walk in the community they most always can walk in the house. Almost always. Not always, but almost always. So that's what the IRT model will

suggest. That we don't have to ask all those questions below a particular point, are they having trouble.

Let's say they are having a little trouble going up and down stairs. Let's say two steps, which for most people it is easier to walk around the community. May not be, it may be. We have to look. If they can go up and down stairs, these two steps, it's not worthwhile usually to ask about all those other questions, because they're functioning -- they're functioning at a particular level that would indicate there is a high probability that that doesn't happen.

Now, if people do that -- if a person can't walk indoors, but can walk outdoors, what the program will do is it will identify that as a very unexpected response. And if people are doing that constantly, then, there is a flag that comes up that says this is a really strange pattern of responses; and that we have got to be very careful about how we interpret them.

So it could be legitimate for some reason

if that's the case. But often when you get these kind of patterns that are not logical, it is usually based on people just randomly putting in answers. So we can certainly flag those circumstances.

DR. BARROS-BAILEY: I had a question. I'm familiar with CAT because that's one of the methods that my credentialing organization was considering in terms of going from paper based testing to computer based testing in terms of the different models; and it was incredibly impressive, the psychometrics and the statistics for those models. They decided for other reasons to go with another model. And I'm wondering what kind of models did you examine when you decided to go with the CAT model? Why the CAT seemed to stand out to be superior for your project?

DR. HALEY: Well, based on our experience in doing CATs for other applications, it seemed like exactly the way to go.

Now, some people will argue that, you know, the CAT requires computer power, and it's too much burden. And so people develop short forms

based on IRT that are fairly good. But short forms, you know, run you into the same place where everybody has to take the same item, and CATs avoid that.

So, you know, based on what people are doing in the quality of life and the functional fields, and what people have done for education, this just seemed like a really -- this methodology seemed to match the problem.

MS. LECHNER: Going back a little bit, not on the topic of calibration, but just on the topic of your responses in looking at the one, for example, of, you know, how quickly you are able to walk or your ability to work overhead for 20 minutes. You know, with SSA's process they will be matching or trying to match or compare claimant abilities to occupational demands.

And so have you given some thought about how your responses might have to be modified to use the same -- you know, same rating scale, in other words, that have or will be used for the job analysis side of it. So that the claimant's self-

reportabilities could be compared to the way that the jobs are rated.

DR. RASCH: So we understand very well -- it's part of why we wanted to be here today -- that the job demand side has to line up with the person capability side. Those are the two pieces of the equation to yield an understanding of one's ability to work, right. Those are the two pieces of the equation that yields participation. That was what I was trying to say earlier. So we understand that these two pieces are necessary, and that they need to line up.

Our work has been on the side of measuring capabilities of individuals in a comprehensive uniform, fast, efficient way, and getting information from both claimants and providers.

We look to this group for the job demand side, and we hope that your work would line up with ours, because both pieces are critical to informed decision making; but how that is going to work out is, I think, as yet undetermined, because they are separate efforts.

DR. BARROS-BAILEY: Dave.

DR. SCHRETLEN: Thank you. This is a wonderful presentation. I really appreciate it. I thought for a long time that this is potentially -- it's a very exciting approach to assessment, and enormous potential utility to SSA.

My understanding -- expert understanding of CAT terminology is not particularly deep nor particularly current. But you earlier used an example that was certainly very consistent with my understanding, and that is that CAT owes itself best to constructs that are very homogenous in which you can have items that are very clearly arranged in a hierarchical order of difficulty. So that really no one who is unable to lift 20 pounds is likely to be able to lift 50 pounds or 100 pounds. That makes a lot of sense to me. If you can't run 100 yard dash; certainly, you are not going to be able to run a marathon.

But how does this translated into some of the interpersonal characteristics? That's not as clear to me. Is the idea that you are going to

define interpersonal sort of emotional characteristics in a clearly hierarchical fashion? And that's just not -- I don't quite understand how that -- how CAT technology addresses emotional personality and interpersonal characteristics in the same way that it addresses relative strengths or persistence ratings.

DR. HALEY: Well, I agree with you. It's a big challenge. And people that worked in this area -- the PROMIS example is a good one -- is they have really struggled to get a good hierarchy in some of their social banks. So that's why we have a lot of items to, obviously, test. So we could cast off many if they don't fit a continuum.

But you are absolutely right, if we don't have an unilateral -- an unidimensional scale by which we could predict where items are going to go along the continuum, we won't be able to build a CAT.

DR. SCHRETLEN: Okay. So thank you. That's very helpful.

So it is the case -- it is not just my

understanding is out of date or something, but you really do need sort of unidimensional constructs that are hierarchically arrayed with items that are assessed at different points along a continuum of ability, because it is really ability measured more than --

DR. HALEY: Behavior as well, I think.

Now, the promising thing is that there are newer models of IRT that allow a little less unidimensionality. It's fairly unidimensional. There are multi-factor models that are starting to be used now, because the field has struggled with this. You know, they can't find full unidimensional constructs; but they can find enough of a unidimensional scale if you use multi-dimensions and the statistics are good enough to put into a CAT.

Now, some of the standard errors are a little larger, and people accept that as well. So it's much more difficult, I agree; and we may fail. But we hope we can pull out enough items that make enough sense to people that it's a construct to start with.

DR. SCHRETLEN: Are you saying, then, that what you do is you have a large pool of items, sort of principal component analysis, factor analysis; and then those individual factors, if you will, you think are -- you can assess using a CAT approach?

DR. HALEY: Correct. It may very well be that in interpersonal interactions we had two -- let's say we have two major factors; they are quite distinct. We would have two CATS for that area.

MS. LECHNER: Based on the number of items you currently have, do you have a sense of how long it takes the claimant or the health care professional to complete it? And do you have a, it's this long now; we want to eventually get it to "X" length?

DR. HALEY: Well, remember the calibration study is in the developmental phase. It's not something that claimants ever would see on a routine basis. This is just to get us to a point where we can build a CAT. No more than 45 minutes currently; and probably between 30 and 45 is the burden that I see right now on the claimants.

DR. PANTER: Should I?

DR. HALEY: Sorry. A CAT would take less than two minutes.

DR. PANTER: I am just wondering if you were confused on the point that this is one stage where there will be more items than would typically be asked; and then later will just be a few items would be asked in just a few minutes per construct.

MS. LECHNER: A few minutes per construct. In other words --

DR. PANTER: That in the end a successful CAT would require only three items, four items per major area, or it depends on how well they can estimate. But it would not take too many items generally to get at the estimate for each of the major areas.

DR. BARROS-BAILEY: Go ahead, Beth.

DR. RASCH: I want to add to that. Precisely. Thank you.

So if we developed -- we don't know how many CATs we're going to end up with; but if we developed the proposed six CATs, it might take 18

minutes for the claimant to complete -- cover all the domains of functioning in the six CATs, because it would take two to three minutes per CAT. So it's very -- that's the benefit of it. It's very efficient, and it really offloads the respondent in terms of the number of items they have to answer. And same with providers.

DR. SCHRETLEN: I wonder if we could return to a question that Shanan asked a little while back about the pilot study -- the calibration study, rather, that you are going to be doing. It's using claimants and providers. And Shanan asked about maybe -- I thought I understood her to be asking about what -- about people who are not claimants or providers, but rather ordinary, healthy people.

And you had said that you don't -- you don't necessarily need a normal distribution for the individual items; and I understand that. But don't you -- but you don't want to attenuate the range of potential responses either, do you?

Wouldn't it be desirable to have a sample

of individuals who actually represent the very broad range of functioning in each of these domains, including people who were healthy and nonclaimants or not disabled at all?

DR. HALEY: Well, you know -- it's certainly a matter of resources. Having a normative sample for these items would be great. If that's what you said or meant, I apologize for misunderstanding.

But given our time frame and our resources, we felt that the majority of resources ought to go to claimants. I mean, these things are built best if they are built for the people. That the calibration work is for the people that its intended for.

If we did a normative sample, and let's say a large one, it could mask out all our disability people. And then our information would be so skewed -- you know, because most people wouldn't be disabled in these items.

DR. SCHRETLEN: Yes, I really do appreciate that. I certainly wasn't implying that.

Just asking why, just based on the normative sample, just that -- I suppose down the road -- I mean, this is a calibration phase you are talking about; but down the road at some point you will want to know how claimants self-ratings and their provider ratings compare to the self-ratings and informative ratings for people who are not claimants, I assume.

DR. HALEY: It would be interesting information. You could indicate at what percentile they would be. You know, you have to have age -- would have to be based on age as well as other factors. So it would be complicated, but it could be done.

DR. BARROS-BAILEY: Sylvia, I'm going to have you ask a question, then we're going to move on. I am kind of the time keeper here so you have the opportunity to finish your presentation. So Sylvia.

MS. KARMAN: Thank you. Very quickly then. Given the stage that we're in at this point in terms of developing a content model as one of the -- you know, initially at least areas that we're

interested in your methodology, certainly, among the things that we will be needing to deal with is refinement of our content model. I'm wondering if you could speak to that. Regardless of what we end up with instruments, you know -- I mean, I'm not speaking to whether or not this method or that for the actual questionnaire. Just for the content model in terms of refinements, what are your plans with regard to that as you get information back from your pilot?

DR. HALEY: Well, once we get data back from the calibration work that is really going to guide us as to what types of items sort of hang together, and what's going to give us the best shot at getting some kind of unidimensional scale, and we hope it looks like personal interactions. It should. But we won't do much more with the content model at this point. What we will do, though, is when we look at the items that we throw out, we will make sure we don't throw all of them out from certain content areas. We will be aware of that.

And so we will get it down to, I don't

know, maybe 70 items or so that will go into the CAT. We will try to make sure that we have a broad spread of content. But we will throw items out if they don't fit, if they don't work. Because if they don't work in the scale, it's going to make the CAT less precise. So we have to be careful to get items that don't fit well with the scale.

DR. BARROS-BAILEY: We're almost at 3:00, but we have, in terms of looking at our agenda, maybe some flexibility here. I think I would like to maybe have the Panel and the team present engage in some decisions. After the break we have maybe about a 15 minute opportunity to have further discussions. It seems to me I don't want to shut down the discussions that are going on, because I think they are really important.

Do we want to go ahead and break now, and come back and have further discussions in terms of the presentation? Finish the presentation and have further questions? Or do we want to go ahead and go over the next 15 minutes, then take a break and come back for public comment?

DR. RASCH: Steve says that he can get through the remainder of the slides pretty quickly, if you want to do that and then come back for discussion after. That's just a proposal.

DR. BARROS-BAILEY: I think that's a great idea. I saw a lot of heads nodding. So Steve, if you would finish up the formal slide presentation, we will take a break, and we will come back for 15 minutes before public comment. Thank you.

DR. HALEY: Okay. Interpreting scores. What you see in your slide is what we call item measure. It describes the item and what people are scoring on the item. And it's the kind of thing that will allow us to look at different regions of the scale, and to look at the items that are either unable, which is a blue -- at least on my screen. Purple is a lot of difficulty. There is the clear is a little difficulty. Then the final blue is no difficulty.

So those are expected responses at different levels of the continuum. So if a person, let's say, scored a 50, you just draw a line up

through those items, you would see what would be expected, what that person's profile would be.

So around a 50 if experts looked at those sets of items they could decide well maybe that person is able to remain on their feet. They could stand a while. And there would be a series of items that you could then used to describe sort of that region and ability. And that's been done a lot in areas -- we have done it some, and there have been many others who have created levels of that continuum. So let's say a cut point would be 34. Between 34 and 52, this is how I would describe claimant's abilities.

So that gives some level of interpretation. And that would be done with a bookmark method where you would get consensus of people looking at your items, and then describe what those scale's are, and how to think about them.

Now, this is the place where we have, I think, one of the chances to integrate with your job description. So let's say a person is between 34 and 52. We have described what they can do. Let's

say they can walk 100 feet; they can do steps, et cetera, et cetera. What would -- how would that relate to other -- to job demands? What job would need a person to be able to do these functions at that level? So I think this might be the thing that would allow us to communicate between abilities and job -- job demands.

Now, this has been done -- I won't talk about this much, but this has been done in reading and math. So many areas where they have a reading ability -- let's say they can do a particular type of phrase. They understand certain phrases that are complex. And people then have -- well, if they can do that, then here are the books that they should be reading.

So they translate from the ability to what it is that is available out in the community in terms of reading that a person then would be able to do. The analogy would be work abilities, and then, you know, job demands; and we would be happy to talk with you much more about this in the future if you are interested.

So we will be doing, as part of our initial work, a small feasibility study. This will be with 120 claimants; and this will be with the actual CAT log. This will not be the calibration items once a CAT is built. We will compare it with measures that are used in the field. Probably a PROMIS type measure or some others, just so that gives the CAT some validity with existing measures that are out there.

So this is what we have covered. Our next steps are develop and calibrate clearly, to work on other dimensions. Then there will be at the end of this a large demonstration project comparing the use of CAT with typical procedures; but that's way down the road once all the CATs are done. And again, just to make sure you understand this is exploratory work and examines how we might approach getting functional information from claimants and health care providers.

And this is where we work at BU on the medical campus. One of the nicer buildings on the medical campus. Please feel free to join us, time

permitting. All right. Thank you very much.

DR. BARROS-BAILEY: Thank you so much Steve, Beth, and Beth for presenting. I think -- let me kind of test the Panel. Are there more questions that the Panel -- yes. So let's go ahead and take a 15 minutes break and come back and continue on for another 15 minutes or so until 3:30, because we do have people signed up for public comment. So thank you.

(Whereupon, a recess was taken.)

DR. BARROS-BAILEY: Okay. We're coming up against some time constraints. So I'm going to limit this to about ten minutes. And so I would like to open it up to the Panel to see if there are additional questions -- follow-up questions. Abigail.

DR. PANTER: It's a quick question about the claimants. I know that you are trying to maximize your sample size for your claimants, but I wonder if there were anyway to identify certain job types that might be within the large groups of job types within your claimant samples so that you could

evaluate what kind of scores you are getting across groups and information like this?

DR. HALEY: There will be a way for us to group conditions that people say they are applying for; but we won't be able to sample those groups as we go through unless we -- unless we are getting all -- monitor the data as it comes in. But as we're getting tapes, information that's likely not going to have all the information we need to make good decisions about which ones.

I will say that we are trying very, very hard to make sure we have enough people that have upper extremity deficits so that we can scale the upper extremity items; and that we see as a particular challenge.

The other group of people, you know, who really want this calibration study are people who use walking devices or individuals who use wheelchairs. And so we're going to do our very best to get enough people within our sample in order to scale those items, because we have wheelchair and device items in the pool; but they don't click in --

they indicate the uses.

DR. BARROS-BAILEY: Beth.

DR. RASCH: I think your question was whether we're going to include information on occupation, and be able to understand how scores relate to particular occupations.

DR. PANTER: Yes, I would like that, just because the more we know about the job, about the work situation, the better off we will be in understanding the scores.

DR. RASCH: So we haven't planned that, but we certainly can.

DR. HALEY: Remember, though, a lot of people haven't worked for a while, so it may not be very useful.

DR. BARROS-BAILEY: David.

DR. SCHRETLEN: One of the slides you showed is the example of a math, verbal ability scale. And I was struck by that because I'm thinking that what's going to be difficult is to capture certain interpersonal and emotional things. This is about mobility, physical things. Yet, if

you look at this, it looks like the items are in the hierarchy; but it is not clear to me what makes them hang together as a scale so that at the higher level of difficulty there is, are you able to walk a mile without resting? A little further down, are you able to pull a cord on a lawn mower? And a little further, are you able to get in and out of a squatting position? Then, are you able to reach behind you to get a seat belt?

So there are very dispirit kinds of things. Some are strength, some are ability, some are flexibility. And although there is a name, "mobility scale," the implication is that there is somehow -- there is something unitary underlying that. Or am I misunderstanding? Is this like multiple CATs in one slide?

DR. HALEY: Well, it is -- it's just the example. And what you don't see are all the items that are part of the bank. We just didn't have the room. It's small print anyway. So those are just representative of the entire bank, but they are not the entire bank. And they are as a schematic

anyway, because we don't have the data yet. So they are our best guess, perhaps, as to where certain items are going to fit.

DR. SCHRETLEN: I appreciate that. But I guess what I'm asking is a slightly different question. That is, even if you have more items, then they would be more heterogenetic, not less. So what I'm wondering is were you just using this for purpose of illustration in illustrating a hierarchical arrangement of items? Or is this actually sort of a speculative CAT for mobility? Because if it's the latter, it seems like it combines very, very dispirit kinds of functioning into a single functional scale.

It may be -- at the beginning of the talk Beth was talking about the emphasis on functional orientation; but is it functional to the exclusion of trying to identify homogenous subsets of items that have to do with pushing, pulling, squeezing, whatever, and so forth?

DR. HALEY: Your point is well taken. There are a number of upper extremity items clearly

in this. And we -- in some respects we didn't have people; but that will fall into a different dimension so that it may very well be a different construct than some of the physical demands. And if it is, then we will have two CATs, or we will have a multi-dimensional CAT with two dimensions, and scores for overall mobility, and scores for upper extremity.

So it is true that we have to have a meaningful continuum of items to make sense.

DR. SCHRETLEN: And it seems to me that both of those are important. Meaningful and continuum. Here we have a continuum, but the meaningful part I'm not sure I get it, unless from what you just said it's that the constructs will emerge subsequently. That you don't want to go in an a priori with some notion of what constructs are. They are going to emerge from the data. But I'm not sure that you can guarantee that happening just from using principal component analysis.

DR. HALEY: Well, we do have the anticipation that there will be an overall mobility

and an overall upper extremity dimension. That's our hypothesis. We will see.

For adults sometimes they are merged. We have seem in PROMIS -- I think they are doing it in NeuroQOL -- is there they are having one dimension, even though it's upper and lower extremity. I haven't seen the data to support that, but I think there will be two at least. It might be more, but hopefully two. And those items that don't fit or don't make sense within a continuum we throw out or get rid of.

DR. BARROS-BAILEY: Steve, I know that you and Beth need to run to the airport. So if there is maybe one more question from the Panel that anybody would like to ask; but I want to respect the fact that you may need to be leaving.

Do we have anymore questions?

MS. LECHNER: I just have -- not a question, but a comment. And that is just that it's refreshing to hear discussion and talk of SSA moving somewhat away from the -- totally impairment based model to talking and considering functional

applications for the disability determination process.

DR. BARROS-BAILEY: Okay.

DR. HALEY: It's been a real pleasure for us to be here.

DR. BARROS-BAILEY: Thank you. I would like to thank Beth, Steve, Beth for your presentation. It was very stimulating. I'm sure we will be hearing more about your project as it continues. Thank you for your time, and we will be in touch. We look forward to receiving those materials.

DR. HALEY: Thanks.

DR. BARROS-BAILEY: We are at the point now in the meeting that we have everytime we meet, the place where we welcome public comment. We have had two organizations sign up for public comment.

When we have organizations or representatives from organizations present to us under public comment, that is allowed for ten minutes, after which time we have question and answer.

So the first organization is the American Physical Therapy Association Occupational Health special interest group with Rick Wickstrom and Karen Jost, if you would like to provide public comment.

There is a button on the console.

MR. WICKSTROM: Thank you for the opportunity to make some brief comments. I'm here as a representative of the Occupational Health Special Interest Group, which is under the orthopedic section of the American Physical Therapy Association. I'm a physical therapist in private practice from Cincinnati, Ohio. I am also certified ergonomist, as well as a certified disability management specialist.

And I found this process to be absolutely fascinating. And it's -- I have listened actually on a lot of the calls from afar, but it's a completely different experience to be here in person, to actually have handouts and see what the slides are referencing.

And the Dictionary of Occupational Titles from a physical capacity perspective has been quite

elegant in its simplicity since it was implemented -- I didn't realize -- 1939; but I completely concur that that work has changed a lot since 1939. I really think this new taxonomy will be an opportunity to make sure that the scaling and the content are more functionally relevant and consistent with science. I really get the sense, for example, looking at the strength measurement that a bunch of content experts sat around the room and constructed a scale that wasn't necessarily based on the work of Stover, Smoke and others that were actually doing research at that time.

And I think this is an opportunity also to introduce other factors that will help us better identify a combination potential for individuals applying for disability benefits. And as I looked around the room -- and I'm assuming all of us that are here that are attending this meeting was maybe an essential function in some respects to our job.

So I look at the universe design concept, which is chairs all the way around, and this work being done entirely sitting. But I also am mindful

that this particular function, which takes all day of our time, could be done, for the most part, by a one armed bandit. It could be done, for the most part, by a person that had limited sitting or standing tolerance, because they would have the flexibility to get up and down. I mean, the work is actually done and designed as far as the room set up for a seated conference, but it doesn't have to be done that way.

It could also be done with technology advances with somebody with low vision, because as I was at a blind work shop that creates products for the Federal government, the technology with the IPhones and taking pictures of devices and having them immediately converted to larger pictures just makes me aware of the potential that we're not tapping into and being aware of because of the shortcomings of our taxonomy.

I even look at the transcriptionist that is sitting there and I am mindful of the last deposition that I did on the expert case on a personal injury case; and the person that was a

transcriptionist for me actually had like a dust mask style of microphone over their face. The whole time I was talking they were talking into a voice recognition software.

I just think it's important that we recognize, yes, we need to look at the way the work is maybe currently being done as we sample the jobs, but we need to have a taxonomy that also is periodically kind of updated, has factors that show how the work is changing as our technology is changing, we're doing things differently.

And another key point is, from a risk -- from my perspective as an ergonomist, the risk management modeling that goes on. What is considered heavy lifting is considerably different than what was considered heavy lifting back in 1939 where current risk models are saying, you know, anything over 70 pounds is quite a significant lift. But we have an opportunity not only to identify what the demands are of the job, but also to identify where those jobs are maybe outliers within an industry. Do a good job of adjusting your scaling

to reflect more reasonable levels of capacity to reflect the way work is currently being done.

I'm real excited about the CAT presentation. It's, I think, really an opportunity to apply some of those methodologies to each of the individual content factors that come out of the content model. I think that would make more sense to me than putting several different areas of dexterity, and materials handling, and crouching all in one mobility category. I think that type of methodology certainly has some exciting promise.

And I really -- and from a standpoint of our Occupational Health Special Interest Group we look forward to the opportunity to look at how the September 2009 report with the initial list of content factors has, perhaps, change to incorporate, or include or exclude, or adjust that factor. I think that would be a good opportunity to actually survey our members to see whether those new content factors, or how they are now being written or portrayed, would make sense from our perspective as a provider or from the perspective of the patients

that we serve.

So nothing but positive comments for what you all are doing with this effort. Really happy to see the increased research focus, and just appreciate the opportunity to -- as our Occupational Health Special Interest Group to provide additional support, resources, or collaborate with you to make this effort a success and not drag it out over too many years. I really think we need that interim content model, factor model sooner than later, so we can begin to at least start talking in an common language to see what is going to help make that evolve during the process. Thank you.

DR. BARROS-BAILEY: Thank you, Rick. I would like to maybe open it up to the Panel to see if anybody has questions or comments. Sylvia.

MS. KARMAN: Hi, Rick. Thank you very much for coming to talk with us today. I really appreciate your interest in our work. One of the things that I just want to, you know, put out there continually, and as much as I can remind people that when we do put out into public our initial list or

initial content model those elements will be for testing. And so, yes, you are absolutely right that would then be open for discussion, you know, more exploration of other people's ideas; but, certainly, some of those things may come off the list in the long run as we're testing or other things will get modified. I just thought I would put that out there. There is no expectation that when one sees this that it's finished.

MR. WICKSTROM: That's very encouraging, and I just would like to -- I very much appreciate that approach, and really we look forward to just -- I guess looking at the public commentary report the hardest thing that I struggled with is that I didn't have any sense of priority, as far as of the things presented from the public. It was just a long, long laundry list. So we had no idea how that has impacted your thought process for decision making about the original content elements.

So it will just be very interesting to see that next step so that we can provide feedback. And so thank you very much.

DR. BARROS-BAILEY: And Rick, I just want to maybe add another comment. I know that when you started your public comment, you were talking about the issue of accommodation. And I know that in some systems that is the way that they could be applied, and just that it's not one of the considerations for the systems that we're looking for in terms of SSA's needs. So that's not a consideration in terms of accommodations, so.

MR. WICKSTROM: I understand it's statutory, and it's not a consideration. But in the way that you present the factor it presents also how a person with a specific disability can have something like one handed work potential or work option, and you are capturing that's the way the job existed. It makes it very clear where the jobs exist that people are able to operate predominantly with one arm. I'm not implying that you are going to accommodate those individuals, but it does provide a way and better insight in terms of matching or identifying opportunities for individuals that have specific kinds of

disabilities.

DR. BARROS-BAILEY: Okay. Maybe I misunderstood. I understood the discussion to be including extra taxonomic data elements to consider accommodations. And I just wanted to clarify that that wouldn't be the application use for SSA.

So thank you, Rick. I appreciate it.

We have another organization, the International Association of Rehabilitation Professionals who will be providing public comment. We have Angie Heitzman and Ann Newlicht, who will presenting for ten minutes. Thank you.

MS. HEITZMAN: Good afternoon. I'm Angie Heitzman, vocational rehabilitation consultant with IARP. This is Dr. Ann Newlicht. She is with me today. We are here representing IARP. As we have before, we wanted to take this time to comment on today's proceedings.

We appreciated hearing from the Commissioner, and particularly regarding SSA's continued commitment to this process and the development of the OIS. We're pleased also with the

development of the Office of Vocational Resources, and its mission and charge. And we're also anxiously awaiting the OIS development plan, which I'm sure everybody is.

One thing we wanted to discuss was the focus groups that are being run now by ICF International. Three of our vocational experts participated in a group the day before Thanksgiving. And the process and questions that were gone through there they found were very thoughtful, and they appreciated the time taken with them.

There were a couple of concerns regarding how the focus group was developed. One person was identified -- we're not sure how -- but she was contacted by ICF; and she put out a notice on our list serve saying somebody is doing these focus groups. Does anybody want to join me? So three people quickly jumped on it, and were involved in it, but it was like the day before, maybe two days before; and it was very short notice.

There was concern that ICF didn't seem to understand what the VEs would do; what the role is,

and what IARP is about.

In the future when you are running focus groups, we would recommend contacting the professional associations who are able to identify and retain good people for the focus groups. And one of our concerns with that is the fact that that may be the only contact ICF has with VEs, just with those three people on that one suboccasion.

DR. NEWLICHT: Regarding the job analysis demonstration project I want to provide some comments. Over and over we hear the Panel bringing up the need for scientific rigor, and we concur. That's very, very important, being that we served for 20 years in the Social Security Administration. I think that we need to know that when we're providing answers to hypotheticals we're doing so in a defensible manner.

We're glad to see that although analyzing the same occupation, such as cashier, that there were differences Panel members experienced in different circumstances; and that this variable condition is real life. This supports the need for

multiple measures, and multiple layers, and multiple sources of information.

Regarding measurement, we share Sylvia's concern about reliance on measure tendency, and the possibility that they can cover up needed information. We would recommended decent wages as more realistic and appropriate. We're not sure that it's necessary to be absolute in every variable. We do need to make sure that's it's correct in terms of allowing us to make decisions, but having one answer for every job just may not be possible.

We acknowledge the importance to capture duties that are less common to fully represent a job; and certainly the importance of operational definitions can't be overemphasized. We all need to be speaking the same language.

In analyzing various job analysis formats we would hope that a method can be developed to take advantage of the best parts of each, realizing that a sample of convenience can provide some data so that we can compare the different methods as a point of discussion was the right place to start. It was

clear that there are good parts to both; and I think that that would come -- become more clear as more methods are looked at.

The process of job analysis does vary depending on job complexity, especially as we talk about interpersonal demands and cognitive demands that are different from a job that's very, very simple, maintain independence, as opposed to something that's very highly skilled, or just a different kind of job.

We support adequate training and retraining analysts on the standardized method. We think that's critical, and needs to be thorough and ongoing.

I guess in thinking through just the job analysis process sample, we think it is a process. And as this unfolds we fully support the need to be very complete and to try to answer all of the questions that need to be answered and be scientifically rigorous. We also support the need to finish it in someone's lifetime, so that it can be used. That this isn't an ongoing process that

lasts forever.

I think as an VE -- and I know that this Panel has discussed this in depth -- but Social Security decisions are ultimately individual. And that any one comes up all the time. So a constellation of issues that an individual brings to a claim, and -- just can't be ignored. And so at some level I think we need to think about what is -- what we can do, and in some cases then rely on vocational expertise and vocational experts to provide clinical judgment when there are some unanswered questions.

MS. HEITZMAN: Do I have another minute? Good.

We wanted to just have a couple brief comments on this afternoon's information and presentation. Dr. Rasch had discussed the needs for better methods to collect claimant data earlier on in the process, and that that would be most useful in decision making. We agree with that. We agree with the need to holistically measure function.

From our vantage point this would be

helpful; however, it's important that decision makers at all levels have the training and tools to make appropriate and defensible decisions, not just at the hearing level.

Regarding Dr. Haley's presentation, we were concerned that some of the items recommended, whether or not they are actually measurable. And also along those measures some of them appeared to be very subjective in nature. Again, going back to the defensibility of how we do our work, we're always concerned about the use of self report as a reliable measure.

In regards to the physician's actual knowledge of the claimant's ability to do a task, we have seen it before when we have been in these positions. Sometimes the doctor will just say, can you do that, to the claimant. The claimant will say no. The doctor will say, okay, and check that off. And we're just concerned again in defensibility with that type of an issue.

Terms needing to be defined, such as "some difficulty." What does "some difficulty" mean? How

do we operationalize that?

And in the end result, what is the purpose for the data? How is it going to be used? Is it that it actually terminates a job matching kind of thing where you prepare the clients' response to -- a doctor's response to the job? Or is that not what the end result is?

We're concerned also about the CAT system, although it is sufficient, it does not always give all the data points that SSA may want to access; and the ability to see the full range of a person's abilities.

DR. BARROS-BAILEY: Okay. Thank you. Panel, any questions or comments?

All right. Thank you for providing public comment.

We are at the point in our agenda where we have opportunity for deliberation. I want to right before lunch kind of plant in the panel's mind a couple things that we need to, I think maybe keep in mind or discuss. And the first question that I like to put out there was a question that the

Commissioner asked us to address. And the question was -- we might not be able to answer it today. We might at least start at that level considering it -- is provide anything in our work, anything in our advice and recommendations to SSA that might assist them in terms of seeing some benefit from this project earlier on. So I will put the question out there. Shanan.

DR. GIBSON: Sometimes it seems like we are setting this up, and we probably are. Funny how that works out.

I know that I am considered by some to be relentless for pushing in doing this scientifically and methodically. That doesn't mean that I don't see the need for trying to find ways we can do things efficiently and create some momentum that will move this project forward in a positive way.

Several of us have talked about is that viable? What happens if it -- if we do something that's not quite ready? And for whatever reason I had this idea this morning during the Commissioner's talk about one area where I thought maybe there is

something we can do to help to get it out there sooner than waiting until the end.

So my question to the Panel is could we consider while the content model development is ongoing to maybe focus some energy on scale development measures? We know that there are certain scales and measure that are going to have to be created. There is a good research literature out there on many of them. And for example, if we could decide on things. We have identified some that are necessary. We know they need a better frequency scale. We know they need measures of repetition.

But maybe we could help them develop good scales. That is something they could integrate into their existing system while doing it consistently with their existing scales to run the test to see what works. And it would be a small piece, but something that could be rolled out in advance.

DR. SCHRETLEN: It seems reasonable to me for us to think about and consider it. I think in a sense we had already begun that process with the report. I know that we have had many discussions.

In our roundtable that very issue came up. We were trying to, you know, distinguish between the constructs and how we would go about measuring them. But it seems to me like it would be very reasonable for the Panel to begin thinking about.

DR. GIBSON: And from my perspective the worse case scenario is they get some better measures of existing constructs they already have to start with. For example -- I was making an example, a list here -- different types of frequency scales to be considered. I believe currently it is continuously, frequently, rare type of thing; probably consistently. That's not the best; but what are some options we might look into?

Well, there is a number of times per day type of scale. There is hourly, daily, weekly kind of scale. There is how many minutes or hours per day, and how much cumulative time type of scale. So right there we can identify four types of frequency scales, look and see which might be better, and perhaps come up with a frequency scale that would give them better data.

DR. FRASER: I just had a thought about ICF activities. Maybe through Sylvia can get some kind of updating of kind of what's going on, and maybe some challenge points in their activities so we can, you know, have some -- provide some feedback for that process.

DR. BARROS-BAILEY: Those are two different points. I'm going to park that one right now until we can maybe get some more discussion on scales and measures. We will come back to that. I want to make sure that we fully develop that idea in terms of the scales and measures.

Any other input into the idea that Shanan proposed, and that Dave commented on? Sylvia.

MS. KARMAN: Yes, what I'm -- I'm interested in this. I think that would be -- I certainly see that as a possibility. I'm wondering what some of you may have had in mind in terms of activities, like roundtables, perhaps? I am sensing that some areas of functioning are going to be lending themselves more readily than other areas.

So just wondering is there -- or is this

something that we want to take up in subcommittees, and then report out? Certainly, we want to talk more about it; but I'm very interested. I think that would be very helpful.

DR. BARROS-BAILEY: I think one of the things that immediately comes to mind is that we might be able to -- we have a couple subcommittee meetings tomorrow. We might be able to have some preliminary discussions about it tomorrow during those, and see if it's something that could be either handled by an existing subcommittee, or if we want to do another ad hoc -- the ad hoc seems to work well for us -- in terms of addressing this.

So Tom.

MR. HARDY: I'm thrilled with the idea of doing this, and we have talked about this individually in different places. I think the message we have got both from the Commissioner and from the public comments, let's kind of start moving things forward.

I like Shanan's idea, and I was going to suggest why don't we do an ad hoc committee to just

get this started, focus on something very narrowly, establish a process, figure out what's the best way of addressing -- setting this up. Do it on just one scale. Get the darn thing done. Figure out what's the most efficient way to do it; and then we can start rolling it out in pieces. That might be a wonderful way to approach the problem.

My question of roundtables, I don't know if a roundtable is necessary. I would ask why you think we would want to do that? I am just curious.

MS. KARMAN: Then, I guess maybe what I need to find out then is what level of involvement you all are thinking of having? I guess I was anticipating -- or what I was assuming -- perhaps incorrectly -- you were looking to provide us some information to get us moving in a certain direction as opposed to developing the scale.

So when I suggested roundtables it was really more on the line of bringing individuals who may have a particular area of expertise and focus on a certain area of measurement, for example, the mental cognitive elements. A measurement for how we

look at residual functional capacity is not the same thing as what you would be looking to measure in the world of work. So there may need to be some discussion about that with individuals, and that might help inform us about what to do next.

If you are talking about actually doing something, that's different. Then we need to talk about, all right, what's your study design? Where are you going with that? And which is fine.

DR. GIBSON: I was thinking if it would be an ad hoc group, it would be an ad hoc group that was charged to work very closely with your research group within your OVRD; that we will also follow the business protocol that you are developing, because one, this would be a good try out for that within the research methodology, and utilizing the plan as it's being developed.

I did not see the Panel actually developing your scales, but being an integral, collaborative part of helping you find research on it, directing towards that. I was also very much focused on scales that measure work activity and

wanting to measure work activities for those items, which are -- I hate to say your easy wins, but most likely to be the ones that are most concrete, and therefore, quickly useful.

We know we have to measure sitting, standing, walking. We know we have to measure in terms of frequency. Let's start with the animals we know best. I think that, personally, would be where you would see the quickest movement toward establishing integrating something new in a roll out type of capacity; but that's my thinking. That's why I'm throwing it out there.

MS. KARMAN: Okay. So that sounds like what I was saying, what I was understanding was, in other words, helping us identify the sources, approaches that we should investigate; and providing some direction along those lines as opposed to doing it. So I mean, providing information about, you know -- so in other words, not conducting a literature survey, for example, advising about what literature -- who has been writing in that particular area, that kind of thing. And that's

what I was suggesting.

DR. BARROS-BAILEY: I think that's a great distinction to make since we're advisory, and we're not developmental in developing the OIS. I think that's an important discussion to have. I think that's clear that it would be advisory. Shanan.

DR. GIBSON: Last comment on this. The reason I used frequency as my exemplar, as I just said, this is about measuring work activity. However, those scales with things like frequency of standing, sitting, walking translate directly to measuring the activities and then the capabilities of individuals. So these are scales that should be able to maintain double duty. How many hours a day do you stand? How many hours a day can you stand? How many hours a day does the job require you to stand? Same scale always.

DR. BARROS-BAILEY: Deb.

MS. LECHNER: You know, when developing a content model, there is so many -- well, there is several different components of it. There is the physical demands. There is the psychological,

cognitive behavioral, whatever you want to call those. There is the environmental piece. And so -- I think we have all had discussions from time to time that certainly the physical components there are going to be probably fewer changes, less dramatic changes than perhaps in the mental cognitive.

So you wonder if -- you know, sometimes you wonder if there are pieces of it -- rather than just try to develop the whole content model at once, are there -- can you break it down? Could you introduce the physical components first, and start moving that ahead while more work is being done on the mental cognitive? Can you start with the environmental piece and move that ahead?

So that's just another way of structuring the output of the content model that might move it ahead faster. There might be several -- you know, there would be development going on simultaneously. Just a thought.

DR. BARROS-BAILEY: I think immediately that comes to me -- and I don't know if I'm

understanding what you are saying; but if -- the way I'm interpreting it is maybe go out and maybe capture data maybe involved with physical. Then go out mental cog. So your -- I don't know how more efficient that would be, because it sounds like you are going out -- it would be more extensive to go out and capture data three times, instead of once. And spending the time to make sure the instrument is complete and capturing the data once. So I don't know if I'm misunderstanding what you mean by that.

MS. LECHNER: Yeah. And I'm not sure that I was thinking along the lines of capturing data solely on one aspect of the content model, as much as I think there are several steps of development within the content model. So there is the development of the items, and then there is the scales that kind of fit, similar to what Shanan was saying. Scales of those items.

And then, you know, so could work sort of be going on simultaneously in parallel? And that maybe how the internal SSA group is working on it anyway. But it is just -- you know, thinking about

how would -- how would you fast tract this project?

Could there be some strategic ways of working on it? Some divisions of labor so that things are occurring simultaneously and using smaller groups and moving that forward and having -- so that, perhaps, an internal group within SSA is working on the physical demands and interfacing with the physical demands subcommittee while work is going on simultaneously with the mental cognitive; and they are interfacing with the mental cog subcommittee. So you have got things happening in parallel. And I'm not talking about the data collection phase.

DR. BARROS-BAILEY: Thank you. Shanan.

DR. GIBSON: I could see where Deb sees deficiencies there, because I do theoretically; but I think practically, at least from what I understand, they don't have enough personnel to have some working with mental cog, and some working with physical. I think we have got the same small group of people trying to handle it all, and that's probably a limitation. We might break us up; I

don't think they can break their staff up.

DR. BARROS-BAILEY: Allan.

DR. HUNT: I think this is a very simulating discussion. I am very stimulated by the Commissioner's question or challenge. However, I think it's something that bears a little more thought. I guess I would like to see some pondering about this before we ventilate in public.

DR. BARROS-BAILEY: Yes. I wanted to bring it out there just as a general topic of discussion. What it sounds like to me is that we may need to maybe go ahead and take it to subcommittee, talk about a little bit more and come back with some ideas in terms of that. As I indicated when I introduced the question, we may not get to an answer; but I think it just bears on the fact that we are aware that SSA needs it today than yesterday; and that there is a sense of urgency. It needs to be done scientifically. It needs to be done rigorously.

And we can't, you know, assume that it's going to be research ongoing for 20 years on it. So

where is that balance to make sure it gets done, and gets done right and efficiently? Okay, Deb.

MS. LECHNER: Another thought would be that we might want to wait until we see the R & D Plan. Because it might -- you know, if we see the plan it might be easier to say, okay, how can this particular plan -- given that we have identified and agreed upon what a quality project looks like, now let's talk about fast tracking. So that we don't undercut any of the quality issues.

MS. KARMAN: When Shanan mentioned the initial recommendation -- not recommendation, but suggestion or idea for us to consider, I was anticipating that that was -- you know, giving us some sign posts in terms of here is some things to consider. Here is some areas of literature to follow-up on or look through. To actually get into developing one segment of content model or doing one piece, I just -- I'm thinking that it's certainly a good idea for us to wait until we got the plan to look at to see if there is other things that might rise to the top.

As you know, it is easy to pick fruit, but I'm concerned about the sequencing or staging of things. Even regardless even if we had, you know, twice the staff, I just think the sequencing may be a problem. So anyway.

DR. BARROS-BAILEY: So that's for consideration for us all tomorrow when we are going through our meetings.

Any other discussion on this particular question? Okay. Comment.

DR. SCHRETLEN: Yes. Just I recall from the first day you were outlining some of the -- what's called business process document, and its four components and the classification in each of those components, that I would think that if we made some significant progress on those over the next year, that would be quite an impressive accomplishment. And I would think that the Commissioner would -- I would assume that's what he is looking for. I think that's a concrete manifestation of the work we are doing.

DR. BARROS-BAILEY: Tom.

MR. HARDY: I don't want to beat a dead horse, but I'm going to anyway.

I like the idea that we are talking about here. I like the idea of taking something from the scale, as Shanan suggested. I wonder if we couldn't use this test and just run it through the whole system, this one small piece, run it through the new plan that you are working on and see how that piece would fit through. Like maybe using physical, it's concrete that would maybe show us where the business plan needs to be tweaked or changed, or how we're going to approach different problems. I would like to see people talking about this. It makes a lot of sense.

DR. BARROS-BAILEY: Thank you. I understand. Okay.

Any other discussion on this particular question?

Now, I want to bring from the table -- you had asked, Bob, Sylvia some question about ICF International. And so Sylvia.

MS. KARMAN: So let me see if I recall the

question was about, to what extent there may be some opportunity to interface with staff about --

DR. FRASER: I -- you know, we obviously, can't all be interfacing with them. I thought maybe through you, you can facilitate progress. Whether that be locating personnel for focus groups, whether that be pointing them to literature, et cetera. If they reach certain challenge points that slow them down, is there someone we can, you know, sort of keep things moving? Again, under the manual of efficiency, you know, if you can update us every six weeks or whatever you feel is appropriate.

MS. KARMAN: Thank you for the offer. I think we can try to coordinate that, so that, you know -- just what you said. If they need to identify experts we will already have people in mind, we could past that information along, or whatever may be needed. So I think that would be good. We can work that out.

DR. FRASER: Where do they work out of?

MS. KARMAN: As I understand it, there is an office in Northern Virginia; and I think there is

another one in Texas.

DR. FRASER: It can be helpful, you know, just people in those areas if they need them, that kind of thing.

DR. BARROS-BAILEY: Okay. Other areas for deliberation that anybody want to bring up at this point?

Okay. We have some administrative business to handle in your packets. Let's see, it's under the last red tab -- or red section that's tab two. We have the Minutes from our last teleconference. This is the teleconference where we voted on the general recommendation number eight on OIS planning.

I want -- I would like to entertain a motion to accept the Minutes.

MR. HARDY: I would like to make a motion that we accept the Minutes.

DR. BARROS-BAILEY: And Shanan seconded that motion.

DR. GIBSON: Yes.

DR. BARROS-BAILEY: Okay. Any discussion?

All those in favor?

PANELISTS: Aye.

DR. BARROS-BAILEY: Opposed?

Motion carries. Okay.

Also, under that tab, it's the very last sheet of paper there, we have dates. Debra Tidwell-Peters did a scan of the Panel for dates for next year in terms of the next three meetings. We have identified three dates, March 15th through the 17th in San Francisco. June 21st through the 23rd in Seattle; and September 20th through the 22nd in Denver. If you would put those dates on your calendar, that would be wonderful. And as we know sometimes because of logistics, those locations change. So those are proposed locations for the meetings.

Okay. So March is looking like it's going to be a very full meeting, the impression that I'm getting. At this point in the meeting I usually open it up to the Panel for recommendations, suggestions for agenda items for March.

One of the considerations that we have had

in the past that we would like to address is maybe some of the presentations from the earlier statistics from Census, other sources for sampling. So one of the thoughts is having a day of sampling presentations. Other thoughts out there that you would like to see.

MS. LECHNER: Will it be premature for ICF to give a presentation to us at that time?

MS. KARMAN: It might be. I think it depends on the topic, but I think it might be. We may really want to wait until June for that.

DR. BARROS-BAILEY: Other areas for presentations?

Are there going to be any current projects that will be at a point maybe of completion before March that we might have presentations from staff on? A general question.

MS. KARMAN: We may be in a position to provide the results for the Ochmann Volt (phonetic) study, for example. So that's one thing. You know, as we get through January I will be in a better position to know what we can predict to be able to

present on.

DR. BARROS-BAILEY: Other -- Allan.

DR. HUNT: What about the International IRS research?

MS. KARMAN: Okay. One of the things we have done is under the -- our research design, one of the first things we're doing is pulling together all the information that the Agency has already collected, investigated with regard to Occupational Information systems or classification systems. And a lot of them were national. Some of them were private sector. Some of this has been done over the years. We had a contractor do this in the late '90's.

We also did point out some investigation for international systems. How are they used? Just so that we are able to with confidence say that we really have examined the full range of what's available.

In addition to that, we are going to augment that investigation with an investigation of other classification systems, particularly those in

the Federal government. Mainly because -- it's not so much that we're interested in the resulting data; but that the methodologies that may have been used to develop their content model -- we would call the content model their instrumentation. You know, they had to make certain design decisions. So based upon what their purpose was, some of their design decisions might be things that will be very important to us.

So what I'm getting at is we are combining that work with -- the work on the international review with what's available nationally, and what we learned from that. So it was really to try to ramp that. When we looked at the international aspects of it, we thought well, it almost seems like you want to be able to say, okay, well, that's nice, but what's available in your own backyard, you know?

So we thought well, we do -- the Agency has -- has that information, and we also have some additional -- some things that are more recent. So it's really a matter of pulling those things together.

DR. BARROS-BAILEY: Okay. Other items for the agenda? Dave.

DR. SCHRETLEN: Yes, because the job analysis is just so far outside of my expertise, I have found it enormously helpful to do those exercises where we -- the Panel has rated a grocery clerk; and then to hear the presentation of your experience of going out -- Bob and Shanan and Deb's experience of going out and observing and speaking with the supervisor, and so forth. It's just very helpful to me, as a panelist, to get a clearer idea of what -- what -- concretely, what this involves. Because, ultimately, this is developed -- you know, we're advising Social Security about the development of a new Occupational Information System.

So at minimum, I would appreciate some more opportunity to just talk with you guys about what you found. Like, for instance, one of the things that you talked about was the fact that the instrument that Deb used do not include any assessment of cognitive or mental job demands or work characteristics. Whereas, the King County

instrument did. But I never heard about the King County instrument, whether it assesses complexity of information processing, or frequency, or just the presence or absence of it? It's just not -- again, it's still quite abstract to me how this is done.

So it would be interesting -- you know, I would actually really appreciate seeing other instruments used, and maybe getting a little bit more into the detail of how do you go about assessing these things? Because that kind of level of detail might actually be very helpful to me in terms of -- as SSA moves forward, being aware of what the weaknesses are of these instruments that exist. What the contours are of how they go about assessing things just in terms of -- although, the job analysis is outside of my expertise, there are some areas that I might be able to bring to the table in terms of thinking about it as I learn more about that.

DR. GIBSON: I joked at lunch that -- you weren't sitting close enough to hear me -- that they won't let you do a job analysis. I still haven't

given up the idea -- one of the things that was very helpful for me was to go to a DDS, because I didn't know anything about the disability adjudication process. There are members of the Panel who have done work analyses for various reasons. There are people who have never done a work analysis. The little pseudo one we used for informational purposes was vastly different than actually going.

It may be around the time people have a lot -- they are able to allocate to something like this. But taking something like this compiled instrument I created from the King County ones, or utilizing Deb's, and somebody actually giving this a try I think can be very formative for folks who want to see more contours of the process. So I think you are right, it is very important to try to do this.

DR. BARROS-BAILEY: One of the things that I think we have done over the last year to try to get the Panel on the same page on a lot of these issues has been through the professional development. So I have seen a lot of discussion today, what you are bringing up of interest, and

other areas as well.

Let's -- let me take this back and work with Sylvia and Debra on it in terms of our agenda, and see what might be the best way to address this. It might be through professional development, or like you say, actually, applied, such as the DDS versus the year of experience that we have there.

Does that sound like maybe a good way to approach that? Okay.

Go ahead, Sylvia. Then I have an idea. Go ahead.

MS. KARMAN: As long as we were on the topic of the job analysis exercise that Shanan, Bob, and Deb did, I think we may want to have more discussion about the process. Because I think that that was really informative. It certainly raised some concerns that Abigail, you know, also was able to reflect on with regard to, you know, how to reconcile differences in ratings, and that kind of thing.

So I think it is just a normative process -- what I call a process, as opposed to

instrumentation and content of the job analysis -- that may be very helpful to us, especially as the year progresses in our work with ICF International on the business -- as the process for job analysis moves along.

DR. BARROS-BAILEY: I didn't plan that, but you segued into what I had on my list, which was the discussion we had earlier on resolving differences. I know that Abigail and David, both of you have some expertise in that area. It might be helpful to have a presentation, maybe consideration on the agenda for March in terms of resolving differences in the literature out there, and what the methodology is that might be helpful to this process. Deb.

MS. LECHNER: Yeah, I think one of the things I would like to see is that -- and I'm not sure quite what to call it, but the session that we had on Tuesday morning -- I think it's labeled as a fact finding discussion. I think we all found that very useful for you and Sylvia to update us on some ongoing issues, and just to give us time to reflect,

and ask questions, and think about things and discuss it. So those were -- that was a very helpful session. And I would love to see that repeated -- that format repeated. I'm sure it would be lots of different topics; but I found that very, very helpful.

DR. BARROS-BAILEY: I have gotten feedback from, I think, almost every Panel member who was there who felt the same. So it may become a routine part of every meeting for us. I agree that it was very valuable to do.

Any other thoughts in terms of things for the agenda that we need to consider for March?

As I said, it is going to be pretty full sounds like to me.

Well, anything else anybody needs to bring up in terms of deliberation?

So at this point, I think we are concluding our second quarterly meeting for OIDAP for this fiscal year. I would like to turn the mike over to the Debra Tidwell-Peters to formally adjourn the meeting as our Designated Federal Officer.

So I would entertain a motion to formally adjourn our meeting. Abigail.

DR. PANTER: So moved.

DR. BARROS-BAILEY: Second?

MS. HOLLOMAN: I will second.

DR. BARROS-BAILEY: Janine.

All those in favor?

PANELISTS: Aye.

DR. BARROS-BAILEY: Nobody wants to stay here longer. Debra.

MS. TIDWELL-PETERS: Thank you very much. That means that the first meeting of fiscal year 2011 is now adjourned. See you in March.

(Whereupon, at 4:30 p.m., the proceedings were adjourned.)

CERTIFICATE OF REPORTER

I, Stella R. Christian, A Certified Shorthand Reporter, do hereby certify that I was authorized to and did report in stenotype notes the foregoing proceedings, and that thereafter my stenotype notes were reduced to typewriting under my supervision.

I further certify that the transcript of proceedings contains a true and correct transcript of my stenotype notes taken therein to the best of my ability and knowledge.

SIGNED this 27th day of December, 2010.

_____
STELLA R. CHRISTIAN