NATIONAL QUALITY FORUM

+ + + + +

HEALTHCARE DISPARITIES AND
CULTURAL COMPETENCY STEERING COMMITTEE

+ + + + +

THURSDAY
FEBRUARY 23, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Dennis Andrulis and Denice Cora-Bramble, Co-Chairs, presiding.

PRESENT:
DENNIS ANDRULIS, PhD, MPH, Co-Chair
DENICE CORA-BRAMBLE, MD, MBA, Co-Chair
MARSHALL CHIN, MD, MPH, FACP, University of Chicago
LUTHER CLARK, MD, Merck & Co., Inc.
LOURDES CUELLAR, MS, RPh, FASHP, TIRR-Memorial Herrmann
COLETTE EDWARDS, MD, MBA, CIGNA HealthCare
LEONARD EPSTEIN, MSW, Health Resources and Services Administration
KEVIN FISCELLA, MD, MPH, University of Rochester School of Medicine
DAWN FITZGERALD, MBA, Qsource
ROMANA HASNAIN-WYNIA, PhD, Northwestern University Feinberg School of Medicine
ELIZABETH JACOBS, MD, MAPP, University of Wisconsin, Department of Medicine
JERRY JOHNSON, MD, University of Pennsylvania School of Medicine
FRANCIS LU, MD, University of California, Davis
MARY MARYLAND, PhD, MSN, BC, APN, Chicago State University

PRESENT(Cont'd):
ERNEST MOY, MD, MPH, Agency for Healthcare Research and Quality
SEAN O'BRIEN, PhD, Duke University Medical Center
NORMAN OTSUKA, MSc, MD, FRCSC, FAAP, FACS, New York University Hospital for Joint Diseases
GRACE TING, MHA, CHIE, WellPoint
DONNA WASHINGTON, MD, MPH, VA Greater Los Angeles Healthcare System
ELLEN WU, MPH, California Pan-Ethnic Health Network
MARA YOUDELMAN, JD, LLM, National Health Law Program

MEASURE DEVELOPERS:
CINDY BRACH, Agency for Healthcare Research and Quality
RON HAYS, Agency for Healthcare Research and Quality (by teleconference)
ANDREW JAGER, American Medical Association
MARSHA REGENSTEIN, George Washington University
BEV WEIDMER, Agency for Healthcare Research and Quality
CATHERINE WEST, George Washington University
MATTHEW WYNIA, American Medical Association (by teleconference)

NQF STAFF:
HELEN BURSTIN, MD, MPH, Senior Vice President, Performance Measures
HEIDI BOSSLEY, MSN, MBA, Vice President, Performance Measures
ROBYN NISHIMI, PhD, Consultant
ADEELA KHAN
NICOLE McELVEEN
ELISA MUNTHALI

C-O-N-T-E-N-T-S

C-O-N-T-E-N-T-S(Cont'd)

P-R-O-C-E-E-D-I-N-G-S

9:01 a.m.

CO-CHAIR CORA-BRAMBLE: I will have the pleasure of leading the group today. For those of you who don't know, I think I know most of you. I'm Dr. Cora-Bramble, I've had the pleasure of working with several of you around the table over my career, so it's a pleasure being here.

I will have the job of being the taskmaster and I hope that you don't say that I'm mean. But I will keep people on task today. There are many measures to discuss. So we are going to go ahead and get started.

My partner in crime, Dennis, is going to handle tomorrow's session, I will not be here. But I will be leading today's session. I'm going to pass it on to Nicole and then we will get started.

MS. MCELVEEN: Good morning. It's nice to see everyone again. I hope your travels were well, and we thank you again for

coming in to participate in the meeting. We are going to go through a few slides to introduce the meeting today. Before we get to those though we would like to briefly do introductions and go through any conflicts of interest as well.

I just want to remind the group that if you have, in particularly in light of the measures that we have submitted. If you've participated on any work groups, if you've been involved the development or testing in any way of any of the measures that were submitted we do need you to disclose that to the group.

And if you have an obvious conflict we will need you to refrain from the discussion of the measure and refrain from voting. You don't have to leave the room, but you cannot discuss or vote on the measure if you do have an obvious conflict.

So maybe start with Denice, just quickly.

CO-CHAIR CORA-BRAMBLE: Sure. I've been a consultant for Pfizer and for the American Academy of Pediatrics. But not anything regarding these measures.

DR. BURSTIN: I'll just say good morning, Helen Burstin. Welcome, everybody.

MEMBER CLARK: I'm Luther Clark, I'm at Merck Pharmaceuticals. I've not been a involved with the development of any of these measures.

MEMBER CUELLAR: I'm Lourdes Cuellar, from TIRR-Memorial Herrmann in Houston, Texas, I have nothing to disclose.

Member Epstein: I am Len Epstein at HRSA and I also have nothing to disclose.

MEMBER EDWARDS: Hi, Colette Edwards, Insight MD, nothing to disclose.

MEMBER FITZGERALD: Dawn Fitzgerald with Qsource in Memphis, and I have nothing to disclose as well.

MEMBER O'BRIEN: Good morning. Sean O'Brien from Duke University, nothing to

disclose.

MEMBER MARYLAND: Mary Maryland, Loyola Medical Center, nothing to disclose.

MEMBER FISCELLA: Kevin Fiscella, University of Rochester, nothing to disclose.

MEMBER MOY: Ernie Moy, AHRQ, I work with the CAHPS team, so probably can't participate in that discussion.

MEMBER TING: Grace Ting, from WellPoint, Inc., and I have nothing to disclose.

MEMBER YOUDELMAN: Mara Youdelman, National Health Law Program, and I was on the advisory committee to AMA's Ethical Force Program, so I can't do the CCAT measures.

MEMBER HASNAIN-WYNIA: Romana Hasnain-Wynia from Northwestern University in Chicago, and for the AMA measures, I'm married to Matt Wynia, who is the director of Institute for Ethics at the AMA, which is the group that submitted these measures, so I just need to disclose that.

MEMBER CHIN: Marshall Chin, from the University of Chicago. Matt Wynia has an affiliate relationship with the University of Chicago but that's the closest I would come to a conflict. Besides sitting next to Romana.

(Laughter.)

MEMBER WASHINGTON: Donna Washington, from VA Greater Los Angeles and UCLA, nothing to disclose.

MEMBER JOHNSON: Jerry Johnson from the University of Pennsylvania, nothing to disclose.

MEMBER JACOBS: Liz Jacobs from the University of Wisconsin School of Medicine Public Health. I was involved in the evaluation of the CAHPS measure, Cultural Competency measure. So I have a conflict.

MEMBER OTSUKA: Norman Otsuka from the NYU Hospital for Joint Diseases. No relevant disclosures, thank you.

MS. MUNTHALI: Elisa Munthali, NQF.

MS. KHAN: Adeela Khan, NQF.

DR. NISHIMI: Robyn Nishimi, I'm a consultant to NQF.

MEMBER LU: Francis Lu, UC Davis. Nothing to disclose.

MS. MCELVEEN: A few other logistics to remind the group, when you speak we do need you to use the mics because the meeting is being recorded and transcribed.

MEMBER HASNAIN-WYNIA: I just thought of something, for the measures that were submitted by George Washington University, some of the evidence that was cited was based on the Aligning Forces for Quality work that's being done that's funded by the Robert Wood Johnson Foundation, and I'm an evaluator of that program. I don't think it poses any conflict, but I just want to make sure that I disclose that.

MS. MCELVEEN: Okay. So please use the mics when you speak. Everyone should have at their station a little sort of tiny

remote-looking device. We will use that for the voting, just so you're aware.

Materials, I did email out a large PDF file of the main materials we'll be using today. If you also need access to any of the measure forms or any additional documents we do have thumb drives with all those materials uploaded, so if you'd like to view them on your computer as opposed to looking at hard copies we can provide those thumb drives for you. We do need them back at the end of the meeting. So does anybody need -- if you can hand those out.

And then finally restrooms, always important. Are outside by the elevators, if you go to the elevator and then make a right, you'll see the restrooms over there.

So if I could just draw everyone's attention to the screen, I'm just going to present a few slides before we get started.

To remind the group again, the main purpose, particularly of the second phase

of our project, is to identify and endorse standards that address health care disparities and cultural competency.

Our goals today are to evaluate the standards that we have submitted against the NQF evaluation criteria. And to determine if those are suitable for endorsement.

We will then review any related or competing measures if that's applicable for our project. Finally one of the things that we do with every consensus project is to identify any gaps within performance measures, again, specifically around addressing health care disparities and culture competency.

And the last exercise that we'll do on day two is we want to present to the group the results of our disparity sensitive measures assessment.

If you recall, we had a conference call in November -- I'm sorry, December to go through some of that information. So we've been continuing in that process and we want to

present those results to the group and discuss a few questions around that.

Our meeting format today will, as we go through each measure, the measure developer will provide a few brief comments to introduce the measure at the beginning.

They will remain available for questions from the committee if that's needed. The Steering Committee will then discuss the measure, vote on each of the major four criteria, as well as to vote whether you want to recommend the measure for endorsement.

And finally if we have any committee members or audience members who've called in then they will have an opportunity to comment.

Operator, this is Nicole, if you could let me know if we have any committee members who have called in on the phone? Okay. We'll come back to our members on the phone.

So continuing on, our evaluation

process will happen as such, as you all know you were each assigned a certain set of measures to review in depth as part of the preliminary evaluation process.

Also within that, we have assigned certain committee members to begin and lead the discussion when we get to a particular measure. So what we're asking is that that person will provide brief comments about the measure, particularly their own thoughts, their ratings around the criteria. And we will then open it up to the group for further discussion. After the group is done discussing the measure, we then will vote as I explained earlier.

Again, any measures that are related or competing will be addressed after each individual measure has been evaluated.

We have 16 measures that we have submitted. And here's a breakdown of the topics. Many of the measures are around communication. We have a few addressing

cultural competency and health literacy.

One measure that we were anticipating a submission from AHRQ, it was the Cultural Competency Implementation Measure. That's been submitted late so it will not be reviewed at the meeting. We will set aside a separate time to review that measure on our conference call.

And finally to just remind the group of the four major criteria that we use for our evaluation process. Again, starting with importance.

Under importance, you're going to have information around the evidence to support the measure. This a threshold criteria, the measure must pass importance to be continued to review it against the remaining criteria.

Scientific acceptability of the measure property is going to house the measure specifications as well as the testing around reliability and validity. Which is also

another very key component to our criteria. Usability, feasability and then finally any competing measures.

So for the voting, again you have a keypad that's been assigned to you. It's already on, you'll have 60 seconds to vote, it's very simple, you'll just simply press the number that corresponds to your voting response.

The results will appear on the two screens to the left and right of the large projector screen. So we're going to do a brief exercise to make sure that you all understand.

Mark, can you hear me?

OPERATOR: At this time, there are no participants on phone lines.

MS. MCELVEEN: Okay. Thank you. So do you have a slide ready?

MS. KHAN: Yes, so this is just a test vote, if you could just answer the question: isn't the weather in Washington,

D.C. great today? Press one for yes, and two for no.

Start voting, we gave you only 10 seconds for this; we just want to make sure it's working. Whatever button you pressed last is the one that gets registered, just so you know.

So why don't we try that again? We just want to make sure that we got everybody. I think it will work, we can move on.

MS. MCELVEEN: Dennis, good morning. Did you want to take a moment to say hello to the group? And also, if you have any conflicts to disclose.

CO-CHAIR ANDRULIS: Good morning all. I do have a disclosure related to the AMA's measures. Since I served on their advisory group. And I just also wanted to say that I think its been a really fascinating exercise to see what shows up and even more fascinating to see what we do with it. Thank

you.

MS. MCELVEEN: Okay. So are there any initial questions from the group before we get started? And our wonderful co-chairs will, one of their roles is to sort of keep the train moving as we go through the measures.

The first one we review we anticipate may take a little bit longer. But just so you're aware, to be sure we sort march through these efficiently, the group will have about 18 to 20 minutes to review each measure.

MEMBER LU: I do have a logistical question. Should we be putting our flags up to signal?

MS. MCELVEEN: Yes.

MEMBER LU: I'm sorry, I'm new to this process, but as I understand, we're going to be looking at each measure one by one. But for example, you showed at the beginning there were like four subcategories. And there are four for each?

At some point we may want to do comparisons if we have to choose between one the other. Is that part of this process as well? Or is that a secondary process?

DR. BURSTIN: That's a great question. So the way we do what we call related and competing measure is the first step is did they pass the evaluation criteria?

So we will evaluate each measure independently, we will then ask the committee to identify which measures are related or competing. And then walk you through an exercise.

And we'll put up the two sets of scores side by side for you to try to decide is there opportunity to select one that's best in class? Is there a reason to potentially select both?

Or even if they are slightly different and there's a reason for both, should they somehow be harmonized to make it work better in the field? So we'll get you to

that as we get through each of the measures individually.

DR. NISHIMI: Right now that's scheduled for the morning update too.

MS. MCELVEEN: Any other questions?

MEMBER JACOBS: I guess related to that, so when we talk about a particular measure if you know it's directly in competition with another measure, we should not talk about it today, that would really be held tomorrow.

CO-CHAIR CORA-BRAMBLE: Okay. Reminding everybody, I feel like we're about to get to the start line. We're about to begin the race. Eighteen to 20 minutes, the first one will take a little longer but I'll be ruthless, just so that you know.

All right, we are going to start off with Measure 1881, Data Collection Domain of Communication Climate Assessment Toolkit. The developers first will present sort of an

overview.

MR. JAGER: Okay. Thank you. Because we submitted nine measures which are all part of one toolkit. I'd like to make my comments a bit longer because they will cover all of the measures that I've submitted if possible?

MS. MCELVEEN: Yes.

MR. JAGER: Okay. Thank you for considering the measures, and I'll start with some very brief background and then discuss the measure development process and field testing. And then finally sum up with the importance of the measures.

So according to the Joint Commission, communication issues are the most frequent cause of sentinel events with issues often arising do to language barriers. Cultural differences and lower health literacy.

Certain patients especially those of limited English proficiency and those of

minority race/ethnicity face greater communications challenges.

LEP patients experience higher rates hospital readmission for chronic conditions. Longer hospital stays for common medical and surgical conditions and may have expensive tests ordered for conditions that could have been diagnosed through an oral history. And patients from minority racial/ethnic groups often face many of the same communication-based challenges, despite English language fluency.

To address these challenges the IOM recommended, in crossing the quality chasm, that organizations become more patient-centered and give patients more control over their care.

Likewise the IOM report on equal treatment recommended that health systems enhance patient-centered communications through steps including improved patient and community engagement.

Enhanced data collection and support for translation and interpretation services in communities where this need exists.

CO-CHAIR CORA-BRAMBLE: Let me just interrupt for a second. Some of this was included in the background information. Can I ask you to summarize it so we can get to the meat of the matter?

MR. JAGER: Sure, that was all I was going to say. The point of this is to say that physicians and other health care professionals' practice and organizations, and every organization in the health care system must communicate complex information to a wide variety of people, many of whom do not fully understand standard health information that they read or hear.

With these challenges and recommendations in mind, the American Medical Association developed the Communication Climate Assessment Toolkit, or CCAT.

The measures we've submitted comprise the nine domains of the CCAT, which is a 360 degree assessment toolkit designed to be used at hospitals and clinics to reliably evaluate the role of the organizational environment in either hindering or enhancing patient-centered communication.

The domains of the CCAT were developed by the Ethical Force Program, which is a multi-stakeholder consensus body formed to develop measures of the ethical environment in health care organizations.

The Ethical Force oversight body is composed of stakeholders from organizations throughout health care representing organized medicine, patient advocacy, health organization policymakers, government, insurers and pharmaceutical and other industry representatives.

This broad representation is important, as the Ethical Force Program uses formal consensus processes as part of the

validation for the climate assessment tools it develops.

Once a patient-centered communication has been selected as a topic for performance measure development, the oversight body appointed a national expert advisory panel.

The first charge of this panel was to review existing norms and performance standards for patient-centered communication.

Based on this review, nine domains were recommended to serve as a framework for the 360 degree comprehensive assessment. Each of the nine domains was carefully reviewed, revised, and approved by the oversight body using numerical one to ten rating scales.

And there are the low scores, in this case a mean of less than seven, reviewed and either revised or eliminated by the oversight body to ensure content validity.

In addition, each member had essentially a veto because of vote of three or

less would cause the domain to be revised or rejected.

Within the domain, our series of specific performance expectations, measured using both staff and patient surveys. For each of these, the expert panel and oversight body systematically reviewed each expectation for, one, its overall importance. Two, its feasability of implementation and three, its potential for measurement.

In this review process, each oversight body member gave each item numeric grades from one to ten for importance, feasability and measurability.

And those items receiving low scores in any of these three categories were reviewed, then either revised or eliminated.

The screening process was repeated three times over a year and a half, and revisions were made along the way to each consensus.

In addition, a report containing

the framework and expectations was circulated to a group of more than 100 external reviewers from across the health care system.

These reviewers received draft versions of the report via email and provided significant feedback about the value of the framework and the feasibility of meeting the expectations in each of the nine domains.

Over the last several years, these measures were further refined and validated through two rounds of field testing at 14 widely varying health care organizations, which included seven hospitals and seven clinics.

I can briefly discuss the field testing. In round one, the initial was for psychometric testing and to refine and simplify the tools. Reliability was assessed by testing the internal consistency or reliability of the domains, measured using Cronbach's alpha.

Standardized coefficients were

used to optimize the reliability of each domain. Specifically, items were systematically removed and alphas recalculated to determine when removing an item resulted in improved internal consistency. And the range of alphas for the patient surveys was .59 to .9 and for staff surveys .69 to .96.

CO-CHAIR CORA-BRAMBLE: Let me just stop you for one second. For those of you who have reviewed the AMA measures, the background information that was provided I thought was substantial.

Do we need to hear this level of detail? I'll just open it up to the group and let me know if you want to hear this level of detail, because a lot of it was included, or at least some.

MEMBER JACOBS: I would say no.

CO-CHAIR CORA-BRAMBLE: Okay, let me then ask to make a final comment so that we can go on to sort of discuss the measure.

Thank you for sort of the summary,

but I do think that having really drove down and looked at this in detail, I think that we sort of get the general picture.

MR. JAGER: Can I just summarize then?

CO-CHAIR CORA-BRAMBLE: Sure. So in sum, communication, we believe it's crucial to attempt to address in any attempt to improve health care disparities and improve cultural competency.

And the CCAT is designed to evaluate organizational performance in developing an environmental support effective communication.

The framework upon which these measures is based was developed using a robust consensus model that brought together a wide variety of experts from throughout health care.

Finally, I want to point out that while we submitted the CCAT domains as nine distinct measures there is in fact overlap

between these domains, both conceptually and in terms of specific patient and survey items that are included in more than one domain's measures.

In addition, the nine domains must be used together. The entire toolkit, not just one or two domains. As such, we were initially unsure as to whether we should submit the entire CCAT as a single composite measure with nine scoring components. So we discussed this with NQF staff, and based on three factors, it was recommended that we submit the measures as we have done.

The factors were: first, that the domains were each tested for reliability and validity. Second, each domain addresses an important issue and distinct aspect of patient-centered communication. And third, we did not calculate a single composite score that summarizes all domains.

Principally because such a broad composite would lose its utility in helping an

organization determine where to put limited QI resources.

CO-CHAIR CORA-BRAMBLE: Okay. Thanks so much. So the lead individual who is going to actually lead this discussion will be Marshall.

MEMBER CHIN: So I think probably a lot of us are new to this NQF process, so this is actually going to be an interesting learning test case for us. In many ways, discussion at this particular scale is purely similar to the next three. So the least force composite are the same.

I'll go into details in a moment but I think the issues that are raised by this case, are, the general topic, like in this case communication and climate, probably most of us around the room would think of this as important. The actual evidence in terms of -- that was supplied in terms of its impact is sketchy in the proposal.

Some of the validation material is

also marginal. As well as, many of the questions may not have been exactly the ones that if we were starting from scratch we would have had. And some are not up to date yet, is not out there yet in terms of validated or approved measures. And so, you know, is this good enough?

You know, so it's back to Helen's point earlier about, if there were competing measures, in getting something on the table.

So with this particular subset, this data collection one, and just to give you a flavor of what we're actually talking about. It's composed of three patient survey questions and then I think there's something like roughly nine survey questions of staff.

The patient ones have to do with" did a staff member ask your race/ethnicity? Did someone from the hospital clinic ask you what language you speak? Did someone from the clinic ask if you need an interpreter?

The staff survey ones have to do

with how frequently staff collected race/ethnicity language data. How often staff has access to information on language, that type of thing.

And these items then are actually combined into a single scale. As was mentioned, in terms of the rationale for why having this in terms of data collection, the documents basically refer back to the broader communication literature.

In fact, for all four sub-scales, it's the same literature that's cited each different time. There was not specific literature cited in terms of linkage of data collection to actual outcome.

So it's an issue where probably most of us around the table would agree it's a good thing, but in terms of the actual validated proof and citations, it doesn't exist here.

The development, I think, was well described. I would just add onto the prior

description. It was developed to and tested in a group of about five urban hospitals, four rural hospitals and clinics, then four FQHCs. So it was a fairly broad group that it was tested upon.

In terms of, then, I guess the reliability and validity. The reliability testing was Cronbach's alpha, as opposed to other types of reliability testing.

And then the alpha for this one was separated between data collection, which I think was a 0.65 and then for the staff survey, the alpha was 0.9. So reasonable.

And actually, if you guys have access to internet, there's a validation article that is available online. You can just access free. Do a PubMed on Matt Wynia, Wynia I think it is, and then it's the American Journal Medical Quality article from 2010.

The validity testing, and this consistent across the four different scales.

What they did was they took each sub-scale score, so in this case they had a collection and then they correlated with one each of three sort of global outcome perception questions they had on their survey.

One was: I received high quality medical care. The second was: my medical records were kept private. And a third is: if a mistake were made on my health care, the system would try to hide it from me. So these were asked in the patient survey.

So in the case of this sub-scale data collection there weren't correlations, odds ratios were basically at 1.0 for two of those outcome global measures. Then actually paradoxically, I received high-quality medical care, it was actually a slightly inverse relationship between data collection and receiving high-quality care.

I guess the other thing is just the face validity issues, as I mentioned, some of these questions really aren't up to date

now. So for example, you know, Romana was on the committee that recently updated the questions asked for language for example.

For example, in the recommendations, they actually asked: how good is your English, as opposed to just asking whether you need an interpreter, for example. And in some of these sub-scales, like in this particular one the provider question, when you look at them, they're not very parsimonious. Again, if we were doing this, it's not probably what we would do.

And when you look at other sub-scale questions it's the same issue that comes up. So I think the overall question is, well, for this particular one, you know, I think a lot of this is they probably think that collecting race, ethnicity, language data is important, even though they may not be showing linkage to outcome yet.

These questions of reliability and validity, at least in terms of validity the

data collection was not validated in their particular sample. The questions themselves aren't the best they could be, but you know, something is better than nothing.

So I think that's sort of the overall question at least for this data questions sub-scale, that it's better than what's out there, which I guess is nothing, I guess, in terms of an approved input measure, but there's really problems with it.

So probably I guess the question we need to ask NQF is in term of what's the bar in terms of, if you can give us guidance in terms of that before we jump in, maybe.

DR. BURSTIN: Those are all very great questions, Marshall. It is always very difficult for us when we enter into new areas of measurement. Of how high, for example, the evidence bar should be.

I think this is a tough line and this came up recently in our Palliative Care Project, for example. Some of the stuff is so

intuitive, I think, analogous to some to the cultural competency work. But very little evidence.

So there is an opportunity if you look at the NQF criteria on evidence to also allow the expert opinion of the group in the room to actually offer input when they feel like the benefits to patients significantly exceed any potential negatives of not in fact having sufficient evidence on some of that.

So that's where I think your expert input can also come to the table. In terms of how high the bar should be set, I think that that's something that you need to sort of decide as a group.

You have a lot of measures before you. Are there some that are better than others? It would probably give you at least an internal sense of what's good enough.

But, you know, importance is a must pass. And scientific acceptability is a must pass. These are very hierarchical, so

you need to get the importance first and then you've got to move to scientific acceptability.

Usability and feasability are harder, particularly for brand new measures like this. But I think you'll have a good sense of it once you get through the first measure.

And just to calm down, usually our first measure takes an hour and a half, Denice, just to warn you, that's typical. I've never seen a group do it in less than an hour. So I also think particularly this particular set of measures, because in some ways, if you've seen one -- there's so many similarities among them that I think you're going to get through most of the evidence issues and most of the scientific acceptability issues with the first one.

So I think, let's begin the process, let's see how the votes turn out.

CO-CHAIR CORA-BRAMBLE: There's

one comment here, and then we'll go around the table. Dennis.

CO-CHAIR ANDRULIS: Something that came up as I was reviewing the comments, and I won't comment on the specific one, I'll just comment generally, is whether there, the term "not quite ready for prime time" was mentioned a couple times.

And I think one of the questions that I wanted to ask NQF was: it seemed to me was there kind of a step down that you could kind of formulate or kind of get your hands around, or this group could kind of think about in the context of not a yea or nay.

Or is this, I know there's a need for yes or nay, but is this also, this group, an opportunity to think in the context of something that might bring it to a yea, being once step shy or two steps shy of that.

DR. BURSTIN: There's certainly an opportunity for the committee to make specific recommendations to the developers of things

that they might tweak. And the question is: can they tweak it and bring it back in a quick enough time?

But the point you made, you raised, Marshall, about the language question. And if now the evidence has changed probably since this was developed, that could be a potential recommendation you could make back in terms of minor tweaking.

But then you get into the issue of, but it's been tested in the way it existed. So those are complicated issues, I think that, in an area like this where there are so few measures out there. I think there would probably be more comfort with allowing perhaps some measures to flow out there to get used to learn more.

I mean, there's sort of this debate as well, if they're not out there, they're not getting used, we won't learn more about actually how they perform in practice.

But that's, you need to decide.

An NQF-endorsed measure can be used for any accountability purpose or quality improvement. So that needs to be in the back of your mind.

If you think this measure is sufficiently ready that, if a health plan picked it up or if somebody else decided it was an appropriate measure, would it be a reasonable one to compare providers?

CO-CHAIR CORA-BRAMBLE: Okay. Around the table, Liz.

MEMBER JACOBS: Marshall, I'm glad you have the first one, not me. Thank you. I had a question about feasability, and I don't know how this fits in the context of feasability of the rest of NQF measures, and maybe we're not supposed to be thinking about that, but if we were asking people to do this whole CCAT thing, because basically we've been given the whole thing to evaluate. I mean, that's a lot of items, it's a lot of questions. And it doesn't seem that easy to do.

But I didn't know if people are routinely asked to do these sort of things as part of NQF, and other NQF measures and how this fits into the context of what other measures look like.

DR. BURSTIN: I think it really comes down to the fact that there is a hierarchy. So we thank importance in evidence is premier, followed by the scientific acceptability of the measure, followed by usability, followed by feasability.

So there is a reason feasability is last. And it's because, if its really that important and it's really that reliable and valid and you think it would provide really important useable results to end users then you would then consider feasability as part of that hierarchy.

One can make the argument it's really hard to do clinician group CAHPS, and that's endorsed because people thought it was important enough to get through those first.

So that's how I would frame it, really think about it as a hierarchy, walk through it as you get to -- feasibility will it be a concern. I don't know that there have been measures that have been, very few measures go down, I think, on feasability because by that point many of the major issues have been brought forward.

If it's not important enough, you probably wouldn't expend the effort. If it's not valid enough you probably wouldn't expend the effort.

If it's a really good measure and it's the only way to collect it, it's something you need to weigh in your minds as you do those votes.

MEMBER JACOBS: Can I ask a question about that? So how long did it take for people to complete the entire CCAT? Do you know?

MR. JAGER: Sure, so for patients, on average, we think it's about ten minutes,

actually.

MEMBER JACOBS: For all the items?

MR. JAGER: Yes. There are about 33 items and then an additional ten items for patients who speak a language other than English.

And then the staff survey is more on the range of 15 to 20 minutes.

MEMBER JACOBS: Okay, thank you.

CO-CHAIR CORA-BRAMBLE: Other questions. Can I ask the group just to turn your name tag just a little bit so I can see, and call out who it is? Okay, Romana, and then over to you, Kevin.

MEMBER HASNAIN-WYNIA: So I have two questions, one for Helen for NQF staff but it relates to something you said a few minutes ago. How in terms of using kind of the expertise around the table to make a decision about kind of the importance.

But then what struck me as I was reviewing these measures is that across the

board. Much of the evidence that's been cited or importance has been through expert reports.

Through IOM reports, through joint commission, NCQA and others. So in terms of kind of the face validity, and I'm not just speaking about this measure, I'm talking about a number of the measures.

If we use that as a criteria for importance then what we get down to is the level of evidence. And that's where we end up struggling.

So I guess, you know, if we can't cross that evidence bar, then what happens?

DR. BURSTIN: If you feel like you can't cross the evidence bar then the measure will go down. But I do think it's important to note there's not a requirement, if you look at our evidence requirements.

There's not a requirement that there be an RCT or that there be a Cochrane review. We know in many of these areas there won't be. So I think you need to decide based

on the evidence that's available.

And that's why actually our evidence task force did this work about a year ago. Specifically saying that we recognize it's really quantity, quality and consistency.

So I think of you take all three of those together it may be there is an area of research where there's only one really good paper. But it's a really good paper and you don't need six in an area like this.

So I think that's what you're going to weigh. But there's no expectation that there needs to be RCT level kind of evidence. Particularly in some of these kinds of measures where you are not necessarily, for example, changing the clinical course.

Or ordering something or not doing a procedure. This is the same issue we're having in care coordination for example. A lot of the evidence is actually very similar. More experiential, not the classic sort of heavy duty evidence we would rely on in

clinical measures.

CO-CHAIR CORA-BRAMBLE: Okay, Kevin and then Colette.

MEMBER FISCELLA: Two questions, the first is, did I understand you to say that we could recommend a measure for say, just internal quality improvement as opposed to accountability, or not?

DR. BURSTIN: No, so there's an expectation that any measure we put forward could be used for any purpose. The QI, any of the accountability applications.

MEMBER FISCELLA: So that means that it really would need to really meet that threshold that we felt comfortable for external reporting.

And the second question has to do with the actual specifications in terms of how CCAT is administered. It looks like, when I went to the AMA website it looks like there's a number of consultants that can help out. But I didn't actually see specifications.

CO-CHAIR CORA-BRAMBLE: Could you respond to that?

MR. JAGER: So the instruments are available for a download on the website, so anyone could use them. But in order to have access to the expertise and the algorithms to calculate the scores as well as our national averages.

We recommend that sites using the CCAT working with especially trained consultants. And they can assist with preparation, because sometimes there's IRB's to be dealt with. Things like nurses unions.

And they assist with the data collection and bring the data and then we perform the analysis and provide the scores. And a feedback report, which also enables them to interpret the results and thereby focus their QI's for example.

Well, we have licensed consultants that we bring in on a yearly basis to make sure that they're trained in proper

methodology.

And we provide recommended methodology for data collection for example. But there's no sort of standard thing that you must, you know, get a 50 percent response rate or something like that.

CO-CHAIR CORA-BRAMBLE: Okay. Thank you. Colette.

MEMBER EDWARDS: I had a question for Marshal, since I didn't review this one. Was the major goal the actual data collection and the importance of getting the data? Or data collection plus the potential impact that it had?

CO-CHAIR CORA-BRAMBLE: Marshall can you turn on your mic? Thanks.

MEMBER CHIN: I think this just sub-scale was mostly data collection, per se. I think the scale itself wasn't necessarily designed to capture the downstream effects. Although there is a three validation questions, one of them was like the patient

saying overall how high was the quality of my care.

MEMBER EDWARDS: So it's really just getting people accustom to the importance of gathering the data. So at some point in the task additional data can be collected to see if it makes a difference?

MEMBER CHIN: True.

MEMBER EDWARDS: And then just a comment, are we in the comment stage?

CO-CHAIR CORA-BRAMBLE: Yes, one more question and then we will hear from the other committee members. Luther.

MEMBER CLARK: This may actually be more of a general question, but one aspect of these measures, throwing in the baseline information was how they correlate it with indicators of health quality.

And I guess my question is how critical is that there be a correlation, and if there is not a correlation which there was not in a least a couple of these. What is the

AMA's plan, how are they planning to approach that?

MR. JAGER: Well seven of the nine measures did in fact correlate with quality, trust, and I'm forgetting the other measure that was quoted. But the two that were not were language services and data collection and we believe that there were other variables that are influencing that.

But from a quality perspective as well as an ethical perspective we believe that improved language services and data collection are sort of important on their own.

And as we continue to collect data we are always analyzing and trying to improve the instrument.

MEMBER CHIN: And those three questions, again, they weren't calling against like chart review measures of quality. But they were patient perception. The three questions the from the patient survey were, I received high quality medical care. Which

needs to be closest to a question that we might be looking at.

The other two to me I think are a little bit more marginal. My medical records are kept private. If a mistake were made in my health care the system would try to hide it from me.

MEMBER CLARK: So I guess that was my issue, because that is an important measure, at least critical for the measure as we have it in front of us, or not. Even given what we might know generically the interim is important of these language programs.

CO-CHAIR CORA-BRAMBLE: Okay, Donna, and then Liz, and then we're going to ask the other committee members to voice their opinion. Oh, so sorry, I missed you, Ernest. So Donna first, then Liz, then Ernest.

MEMBER WASHINGTON: Yes, it's a question for the developer, looking at the sample sizes that were included in the validation study. It looks like the numbers

are sufficient to stratify by race/ethnicity even by broad categories. I wonder of those sort of analysis were done but not published?

MR. JAGER: So because the scores are based on staff and patient components, and we're unable to determine, for example which provider saw which patient. It's sort of hard to stratify the score by race/ethnicity.

We could stratify certain components or individual items, which we in fact do when we report back to the site that uses it.

But there's not a real good way to say a certain subcategory or demography group scored a certain way because of the 360-degree comprehensive assessment nature of the tools.

CO-CHAIR CORA-BRAMBLE: Okay. Dr. Jacobs.

MEMBER JACOBS: I'm going bring up an issue that's just coming to me as we have this discussion and based on what Helen said.

In addition to looking at all

these measures and some of the work that I do, sometimes I feel like we hold what we should do up to some standard of evidence when actually we can't be totally confident.

Or you can't actually address disparities unless you know someone has a language barrier. They need an interpreter for example. Or that they are actually asking people if they need help.

So this is just a bigger issue that I face in the work that I do and I think that we're talking about here that I'd like us to keep in mind is some of these things I think we need to be asking them. Even if there isn't a ton of great evidence.

And actually this is better evidence for other things that we asked people to do in health care. So I just want to throw that out there.

I mean, as a scientists, I'm like, oh my god, the science is not very good. But then I'm like, do we really need that great of

science to decide that we should do this.

So that's more a comment than it is a question.

DR. BURSTIN: That's a great question. And I wanted to read you the section of our evaluation criteria specifically on potential exceptions to evidence because I think that's important. And we probably should get it into this light for folks.

So we recognize there are areas like this where some of the stuff is kind of intuitively obvious. And are you really shouldn't study that someone shouldn't have pain.

I mean issues like that as we encounter in Palliative care. So the specific language says potential exceptions to the empirical body of evidence.

If there is no empirical evidence, expert opinion is systematically assessed with agreement that the benefit to patients greatly

outweigh potential harms. And it would pass the criteria.

And if we say if you guys agree that it's judged by its potential benefits for patients clearly outweigh potential harms. So there is, we've already built this in explicitly for those areas where we know the evidence base is just growing or there's some places where you're just not going to get that kind of evidence.

CO-CHAIR CORA-BRAMBLE: Good point, Ernest and then Dawn.

MEMBER MOY: I guess my question is mostly a question for the developer. It seems like the dimensions that are captured in the patient survey and the provider survey are in some ways hitting at the same thing. And somewhat duplicative.

And I was wondering just why to add that perhaps unnecessary complexity to the issue. And number two because you have these two different components, what kind of

guidance are you going to give to potential users if they disagree.

So say the patents say they're not collecting data and the providers say, yes, we're collecting lots of data. What do you do with that? Do you average it out and say okay, it looks about average.

It seems potentially unnecessary complex and I'm not sure what you do with that data that don't necessary correlate.

MR. JAGER: So by design the patient and survey items asked about similar things, because we're looking to get the different perspectives.

So in the example that you've given, if the patient says no, no one asked me my ethnicity and 90 percent of the staff says yes, we always ask. That's useful data.

Regarding the scoring component they are equally weighted so this is to counter if you have a great number more staff respondents than patents or vise versa. They

are equally weighted so we would get an even score.

But we do report both components so that you can see that we got a 30 on data collection and a 70 from on patients and 70 for staff that's important information, there's a disconnect there.

And we also report key items and compare not only what staff and patients say but also what executive leadership says and whether or not there's a policy regarding that issue.

MEMBER MOY: Can I ask a followup question? And then are we being asked to endorse this as a composite as it were? Or as an individual component? It seems like it actually is two separate things.

DR. BURSTIN: I think that's a discussion for you to have. And it's not exactly clear to me. It seems like they are components in the larger tool and the question is is it a composite?

And if it is a composite is it submitted in a way that allows you to have an overall score. Which wasn't clear to me.

CO-CHAIR CORA-BRAMBLE: Very good point, in reading these that was issue. Dawn.

MEMBER FITZGERALD: Yes, and I'd like to go back because listening to Elizabeth's comments about this desire to have this information available. And to your comments about the lack of evidence can still lead to an opportunity for a measure when it's important enough and significant enough.

And I admit to being a little bit conflicted because I think on the one hand that those are both very valid points. But then I go back to Kevin's very specific question about the purpose of the data.

And while I'm willing to kind of go to the cliff and terms of saying that I think it's important and the measures dictate the desire to have this kind of information available.

I'm not sure I'm willing to take the leap of faith to say that they are appropriate for public reporting or quality improvement purposes. Because I personally have some concerns with the lack of information on the consistency of administration of the data.

And the extent to which without that level of consistency and how to administer it making comparisons across plans or providers would be troubling to me.

CO-CHAIR CORA-BRAMBLE: Okay. Marshall do you have one more, or are you done?

MEMBER CHIN: Yes. I was going to follow up on Dawn and Liz's points. That I'm clear in my mind anyway. I think the distinction between importance, I think Helen's was importance and validation.

I think what Liz was talking about was more importance that there may not be existing data showing that we can collect

race/ethnicity language data that at least is better outcomes.

I have a couple of experts opinions saying that that's good enough, lets go ahead and try to develop a measure.

It's the validation point which I think is trickier. Especially as Dawn and Kevin said, because this could be used for accountability purposes. Across like the four different measures for this particular instrument, the developers present very nice data showing spread across respondents.

It's generally like a 20 to 25 point spread across respondents. So there were high scores, there were low scores. The challenge is though is that we really don't know what the meaning of that is.

For example, you know, the three questions they're using as their validation ones, again, I think they're questionable if these are the right questions to use.

The perceptions in and of

themselves aren't the problem for me. But I think it's a high bar if we had to say, well this tool had to correlate the traditional clinical quality measures.

So I think it's okay in terms some of these perceptions. But I'm not convinced that these are sort of the right ones. And if it has to be for accountability purposes I think we need to have a pretty high bar there in terms of the validation.

CO-CHAIR CORA-BRAMBLE: Norman.

MEMBER OTSUKA: I just wanted to keep the perspective of the clinician and the American Academy of Orthopedic Surgeons sends out a needs assessment to their members, 30,000. And culturally competent care is always important.

But whether they are willing to do something about it or not is not a high priority. So my plea to you is, I agree, there's got to be some evidence, and I agree we're practicing sort of like best medical

evidence here.

And doing level five stuff. But keep the clinician in mind, and they all agree, they're all on board it's important.

But let's give them something that is important with some level of evidence. And it's tough to do all these measures, you know, if you're a busy clinician in a hospital or ask your staff to do it. Thank you.

CO-CHAIR CORA-BRAMBLE: Thank you. Comments from any of the committee members at this point? Ernest, did you have something else to say? Okay.

DR. BURSTIN: Just one response to Marshall, I think the issue that was raised by Dawn about consistency of data collection is under scientific acceptability. I just want to keep those separate.

I mean I think there are some validity concerns about a measure that might fit into the evidence piece. But I think that piece in particular I would argue is the

second criterion.

So for the importance vote it's really about evidence and that's display of results is really the first one. Because that shows you there are three parts to importance.

The first one is evidence, the second is it a high impact area, and obviously we wouldn't be sitting here if in same ways it wasn't.

And the third is, is there a gap in care or is there a known variation. And they've clearly have provided some data on the variation side that I think, again, fits under importance.

CO-CHAIR CORA-BRAMBLE: Okay. I'll invite the other committee members to make any other comments.

MEMBER EDWARDS: I just wanted to ask when, I can't remember who asked, it was Dennis, saying can you make a recommendation that something be tweaked. How do we handle that when we go to vote?

CO-CHAIR CORA-BRAMBLE: I think we can state it, and then it will be included in the transcript.

DR. BURSTIN: And after the meeting the developers will be asked to respond to a series of, again, this is really early in the consensus process. You guys will have your deliberations today.

You may have a series of questions and, you know, it may be that may be he needed to answer some of these harder questions perhaps. You'll then have a chance to have those questions come back to you, perhaps even re-vote on the measure if you think the additional information is so compelling.

It then will go out for commenting. There's a whole long series of steps here that you're really at the very first step at this point.

MR. JAGER: I do want to say that Dr. Wynia is going to try to call in about 10:30.

CO-CHAIR CORA-BRAMBLE: Any final comments from, yes, Kevin.

MEMBER FISCELLA: I was just going to say that the timing of the administration of patients reports of what they're experience was with their provider makes a huge difference.

So that if you query somebody right after the visit they can answer fairly reliably about what actually happened. If you query somebody say a month later their affective heuristics really take over.

And you've just got a sort of a global sense of, you know, was my experience positive or negative? And people tend to rely on those heuristics in order to answer.

And the further out you go the more those sort of affective global ratings sort of bias the individual responses.

CO-CHAIR CORA-BRAMBLE: So you're advocating for immediate survey, or late survey? I couldn't understand by your

comments.

MEMBER FISCELLA: In general if you want to get that specificity you want it done immediately. But the other issue it brings up is that if people are administrating them at different times you're going to get huge bias in terms of responses.

CO-CHAIR CORA-BRAMBLE: Okay. Are we prepared to vote? Any final comments?

MEMBER HASNAIN-WYNIA: I have a question, and it relates to the measure developer. Based on kind of where you started and I think kind of on Liz's comment about feasability. I'm still not clear, because these measures were submitted separately, individually. Even though they are part of a larger organizational assessment tool.

I'm having a hard time connecting the dots in terms of the implementation. So let's say we vote on these measures and one passes. What happens if they're suppose to be part of a whole tool to gauge the

organizational climate in terms of communication?

I'm very confused about that.

MR. JAGER: So I guess I don't really know the answer if that would happen. The tool kit is developed to be taken as a whole. There are four components right, but the scoring component is based on the patient and the staff survey.

There is overlap of the items though. An item in data collection could also be an item in work force development, for example.

CO-CHAIR CORA-BRAMBLE: I think we could also, depending on, once we go through each of the measures. Depending on the outcome we then may be able to step back and say, well, you know, we did our sort of due diligence but this is what we find in looking at it in its totality. And I think that may be the way to go.

MEMBER HASNAIN-WYNIA: I imagine

that we can also in the kind of the request for tweaking or those comments later on, we can also raise that question back to the developers, right?

CO-CHAIR CORA-BRAMBLE: Correct.

MEMBER LU: There was a mention about Dr. Wynia perhaps calling in and I'm just wondering of some of these questions that have come up that I think are quite important. Would it be worthwhile to bring him in at this point?

CO-CHAIR CORA-BRAMBLE: To bring who in, I'm sorry.

MEMBER LU: Dr. Matt Wynia.

CO-CHAIR CORA-BRAMBLE: From the AMA?

MEMBER LU: Yes.

MR. JAGER: So he's on service at the University of Chicago and he was going to try to call in by 10:30 today. But he doesn't have control of his schedule because he's attending.

CO-CHAIR CORA-BRAMBLE: No, I think we have to go ahead and vote and then we'll circle back if we have to. I wanted to check if there were any comments on the phone before we take a vote.

OPERATOR: We have no phone comments.

MEMBER JACOBS: Can I suggest something a little bit different? If we feel like it's important to talk to Matt. Which is, you know the other measures, all these other measures that are based on the same tool, maybe we could move on to a discussion of the next one? Before doing the vote. I don't know, maybe that's not NQF's process.

CO-CHAIR CORA-BRAMBLE: The concern is that we're going to get them mixed up and when's it's time to vote I'm not sure that we're going to be able to figure out.

MEMBER JACOBS: The next four are AMA measures.

CO-CHAIR CORA-BRAMBLE: I don't

think so, I think we need to vote.

MEMBER O'BRIEN: I just would make the same suggestion, so I'll just weigh in on the pro side.

CO-CHAIR CORA-BRAMBLE: It's the pleasure of the group. My concern is that when it's time to vote these are, there's overlap and you know I'm not sure it's going to be as easy to keep our vote specific to a measure.

But as a group you feel it can be done I certainly will defer to all of you. So Mary and then Kevin.

MEMBER MARYLAND: So just in terms of practice, and I understand the need to have a vote. And if we have to follow that process or access it by the finds for this one.

But perhaps after we've discussed the second which may not be as murky as the first, if we need to revisit the first vote I would suggest we do it sooner rather than later.

CO-CHAIR CORA-BRAMBLE: There's a period of calibration among us as group members, as there is when we do grant reviews. So I think the first one there will be come internal calibration, that's my sense.

MEMBER FISCELLA: I see a lot of the core issues as really common to the measure, so I would support doing it all at once and giving Matt a chance to weigh in.

CO-CHAIR CORA-BRAMBLE: Okay. So at least three members are interested in doing it. I'm happy to do it that way. I'll just defer to the NQF staff in terms of the logistics.

DR. BURSTIN: I think this issue should vote on the first one. And think about it but they need to get into the process. So I would agree with the calibration.

MEMBER FISCELLA: Can the vote be revisited?

CO-CHAIR CORA-BRAMBLE: Why don't we do that? Why don't we vote and if we need

to let's revisit it. I'd feel more comfortable because then we'll end up with a vote as opposed just amorphous material.

MEMBER EDWARDS: Before we vote I just need to get some clarity about how to vote. If your vote has qualifications. With the tweakings, I mean, it's not, it's a qualified, yes, so how do we do that?

DR. BURSTIN: I think you should vote on the measure that you have before you. Before you can have assurances that anybody can tweak or change anything.

MEMBER EDWARDS: Meaning if you're not comfortable vote no?

MEMBER YOUDELMAN: Or do you vote yes, because you want to tweak it?

DR. BURSTIN: No. It can always go back to the developers. So you can certainly re-vote, it's not a big deal so if you just want to do a quick kind of get one under belt. Knowing you may get more information.

I think the issue is I don't know that anybody but Matt could really answer some of the tweaking kind of questions.

CO-CHAIR CORA-BRAMBLE: And the other thing that's the issue for me is there anything that the developers going to say to us that is going persuade us to change our vote.

I think they can clarify but is it really going to change substantively how we would vote?

MEMBER TING: I'm sorry, one last clarifying question, so from my own personal stakeholder, i.e., the health kind of perspective, is if I don't think it would work do I vote from my stakeholder perspective or should I look at the general global industry perspective?

DR. BURSTIN: You're each asked to serve as individuals not as stakeholders. We try to get the mix of stakeholders at the table. But you're here because of your

expertise. So you should vote based on what you think is the quality of the measure.

MEMBER MARYLAND: So my question is going to be maybe on middle ground. We definitely have to vote, yes or no. And right after that vote can we then give you a brief here are our antidotes if it's possible to address?

CO-CHAIR CORA-BRAMBLE: I have no problem with that. That would be appropriate in my opinion.

We're voting. We're going to vote. So I would invite you to look at the other members of this sub group, how they voted. Marshall if you have final remarks, in terms of recommendations or, this would be the time to say it.

MEMBER YOUDELMAN: If we're not allowed to vote, do we just not vote or do we press an abstain button?

CO-CHAIR CORA-BRAMBLE: No, just don't vote, and there are a few people I think

who were conflicted in terms of this.

MS. KHAN: Can I just get a show of hands who's not going to be voting so my numbers aren't off?

CO-CHAIR CORA-BRAMBLE: Two individuals are not voting.

MS. KHAN: Okay. So we're going to be voting on importance to measure, we're looking at high impact, was it moderate or high. Performance gap moderate or high and the evidence if it's a health outcome with rational or the consistency of the evidence is moderate or high. And the quality and the quantity are moderate or high or low with special circumstances.

So was the criterion important to measure reported and met? Press one for yes, and two for no and you have 60 seconds to answer the question. Has everyone voted?

CO-CHAIR CORA-BRAMBLE: Okay. Rocking and rolling, next.

MS. KHAN: Okay. So your final

response was 19 yeses and zero noes. We've only done one criterion. Sorry, three more to go.

We're going to be voting on reliability now, so reliability testing was conducted with appropriate methods, scope and adequate demonstration of reliability.

To what extend was the criteria and reliability met? Press one for high, two for moderate, three for low and four for insufficient information.

So you can start now. Has everyone put in their vote? So we have two high, 12 moderate, three low and one insufficient information.

So then looking at validity to what extent was the sub criterion validity met? It's one for high, two for moderate, three low, four, insufficient information.

Sorry about the music guys. Has everyone put in their vote? So our final is one high, seven moderate, nine low and two

insufficient information.

So now, voting on scientific acceptability of the measure properties, our votes for reliability and validity are rated moderate or high.

DR. BURSTIN: The reason we are having a little consternation up here, is we actually have an algorithm for scientific acceptability and basically low validity on anyone means it doesn't go forward.

So if you've really just rated that as low validity then the measure stops.

MEMBER HASNAIN-WYNIA: Does that mean that mean if any one vote of low validity it stops, is that what you're saying?

DR. BURSTIN: It's that the majority of you voted low.

MEMBER HASNAIN-WYNIA: Oh, a majority.

DR. BURSTIN: Yes. That's a good point insufficient information is not clear and this may be an example if you had more

information you might in fact, those two votes would flip that. So I think that's the question here.

CO-CHAIR CORA-BRAMBLE: Okay. So I'm told this concludes the voting for this particular measure.

DR. BURSTIN: Although going back to the initial point the question would be are there additional, you know, was the reason it was voted down, low on validity anything you would like to prepare a set of questions for Dr. Wynia when he is available.

CO-CHAIR CORA-BRAMBLE: That would be, that could flip the vote. That will change the outcome in terms of the algorithm.

So when you ask your questions at least try and focus on this validity issue because that can make a difference as whether this measure is accepted or not.

MEMBER TING: I'm sorry I don't have the information in front of me but could you someone give me the validity correlation

for that one where the patient answer is linked to a high quality of care?

MEMBER CHIN: For the high quality medical care question it was actually negatively correlated at .95. And the other two questions it was 1.0 odds ratio, so no validation with their data.

MEMBER TING: I wasn't too fond of the other two questions to be honest.

MEMBER CHIN: Just for the record it is 10:19, we started at 9:20. Thank you, very much committee members. Just for the record. All right. Are we ready for the next one, 1888, Lourdes.

MEMBER CUELLAR: Yes. So I'll introduce the next measure. And the title of this measure was Work Force Development of Communication Climate Assessment Tools.

And it's really looking at communication, and it's looking at work force development. And this is another AMA, is the story for this particular measure.

The numerator statement on work force development is centered on patient-centered communication. And indicates that an organization should ensure that the structure and capability of its work force meets the communication needs of the population it serves.

Including employing and training a work force that reflects and appreciates the diversity of their population.

The measure scored on two items from the CCAT survey. That are patient surveys, and those two items are, did doctors explain things in a way that you could understand? And do hospital or clinic staff come from your community?

For me that was an interesting question, especially when you come from a big urban city the definition of what is community, that was the first question I had there.

The secondly there were 21 items

related to the staff survey most had to do with communication and training. And their were only two leadership questions, and I had a question related to that as well. Because so much of the rest comes from the leadership of the organization itself.

The other indication, or the other question I had on this, ultimately the board of trustees was involved. There was no questions related to the board enrollment in the process.

And most successful organizations also have some sort of community advisory board and there was no questions related to actually population based input.

You had to have a minimum of a hundred patient responses and 50 staff responses.

The denominator statements were two components. One were the patient response and the second one was a staff response. And the measure type of course is a patient

engagement in experience.

For evidence of high impact they're actually looking at, they're correlating communication to patient, or poor quality or quality patient care.

In the summary, what they're saying is effective communication is critical to providing high quality care. And can be effected by a number of modifiable factors. Validation of the measure of the study comes from these questions regarding patient-centered communication itself.

So briefly the benefit that they're outlining in this measurement is understanding and improving communication may be the key to addressing a disparity which obviously is an important health care goal.

Some of the questions I had here, some of the citations as in Marshall's were dated going back to the early 2000's. There's been a lot of research and data has been submitted since then.

In addition some of the validation factors I think were in question. For example in my organization which is the largest organization in Texas in the health system.

We must do 12 surveys a year and then validation, you know, people will provide answers to surveys but how can you validate that as truly accurate or answering a survey just to answer. Especially when you work in organizations that have a lot of surveys.

And again, going with education, you can educate your staff but that doesn't necessarily validate that they're going to utilize the information that they're given to actually put into practice.

So again, this whole measure is based on, the other factor that I had here too. On a lot of the citations a lot is working with the Spanish speaking patients, which of course is our largest population of low English proficiency.

But really we're a very diverse population and there really needs to be more studies in other minority population as well. So those were some of the questions that I had.

Looking at my committee members, everyone voted that this would be a high impact. Some of their comments that I have here. Research has demonstrated the language barriers were either real or perceived.

And I think perceived is a major factor. Because there's a lot of studies have shown perception weighs heavily on how patients respond to surveys. Can directly impact inherence and therefore apply to just over total quality of care.

The rational, some of the comments here, was not well presented by some of the authors. And I think this is a very important point here. That one could extrapolate using face validity that well trained work force should improve communication.

But there was not much cited in the literature to provide evidence for this. So I use that as a background to open up the discussion, and I think, Dennis, here is the question that I can't remember where it is. Someone used that phrase that you said several times in some of the comments.

The other things, there were some citations for the medical record again, the use of an electronic medical record is not universal. So that again provides a weakness in the study as well. So I'll open up for discussion.

CO-CHAIR CORA-BRAMBLE: Thank you, Lourdes. I'll invite the other workgroup for members to chime in at this time. You all have to really calibrate it quickly.

Any other comments? Okay. The group at large, any comments for Lourdes? Dr. Johnson.

MEMBER JOHNSON: Jerry's fine. One of my big problems with a lot of these

measures is the extent to which there is a lot of discussion about overall communication. Or overall competence in contrast to what it seems these specific domains are suppose to be addressing.

So we're looking at five or six or seven or eight domains and spending all this time thinking about the domains but a lot of discussion and evidence is, and even some of the survey questions seen to be more about overall communication.

So this one was suppose to be about structure, to me I think it's about structure and training. And those two staff issues, maybe one of them relates to that but I'm not sure about the other one.

I mean, there are two for patients and then I guess the rest of the questions are for staff. So I'm continually struggling with exactly what are we evaluating here. Should we be just trying to just focus on the domain.

In this case work force

development. And if so, I wonder about if the survey questions are really the best ones.

CO-CHAIR CORA-BRAMBLE: I would also add that I'm increasingly concerned that we should be looking at this in its totality as a tool. As opposed to each one of these individual matters. I just don't think it gives us the full picture.

MEMBER CUELLAR: And Jerry, your point is well taken because while there are a lot of questions related to the staff, did you receive training in this, did you receive training in that. Only two patient questions, so there's really no validation that the staff training really in any way enhanced their care, or their quality of care.

CO-CHAIR CORA-BRAMBLE: Other comments, Mary.

MEMBER MARYLAND: And mine is just anecdotal, even when you presented, Lourdes, she used the term doctor versus provider. And in primary care nationwide it is frequently

someone other than a physician. I just think we need to be conscious of that.

CO-CHAIR CORA-BRAMBLE: Point well taken. Any other comments? Are we prepared to vote? Record time, all right, Ms. Elisa.

MS. KHAN: So we're going to be voting on importance to measure importance. Was the threshold criterion, importance to measure and report met? Press one for yes, and two for no. Let's try that again.

There we go, so you can start voting. Is everyone done? So we have 17 for yes and two for no.

And again looking at reliability, to what extent was the sub criterion in reliability met? Press one for high, two for moderate, three for low and four for insufficient.

You can start voting now. Did everyone vote? We are going to move forward, so it's 13 moderate and five low.

And again moving on to validity,

to what extent was the sub criterion for validity met? One for high, two for moderate, three low, four insufficient information. So you can start voting now. Did everyone put their vote in? We have ten moderate, eight low and one insufficient information. So it passes.

Moving on to usability, we're looking at meaningful, understandable and useful for public reporting and accountability. And meaningful, understandable and useable for quality improvement.

So to what extend was the criterion for usability met? One for high, two moderate, three for low, and four for insufficient information. Okay, everyone voted? So we have two for high, nine moderate, four low and one insufficient.

Okay, moving on to feasability, so looking at 4A, data generated during care for via electronic sources, foresee to

susceptibilities, unintended consequences are identified and 4B, data collection can be implemented.

So to what extend was the criteria and feasability met? Press one for high, two moderate, three low, four for insufficient information. Okay, I think everyone completed their vote. So we have one for high, 11 for moderate, five for low, and one insufficient.

And we're voting on overall suitability for endorsement. Does the measure meet all the NQF criteria for endorsement? Press one for yes, and two for no. You can start voting now. Did everyone vote? We have 11 for yes and six for no, so the measure will pass.

CO-CHAIR CORA-BRAMBLE: Okay. Rocking and rolling. 1901, Dr. Lu.

MEMBER LU: Okay. So if you have Attachment B with you you might want to turn to Page 17, because there we have the summary from our workgroup in terms of looking at this

particular measure.

In terms of importance to measure and report our overall group in terms of impact rated it a five as high and I think that was very strong. In terms of performance gap again, three rated it as high and two as low.

And then in terms of the overall evidence, three was a yes, and one was a no. I think overall in my assessment here, I think that consistent with the other two parts of this AMA tool. I think the from my perspective the importance aspect has been met.

I think that where the rubber meets the road is the second area of the scientific acceptability. In terms of the reliability and validity, that again where the main evidence comes back to the survey that was done of the 13 health organization and only nine of them continued on to the second phase of the study that led to the published

article, the peer review article.

And I guess the question really that kind of ties in with the other two scales that we looked at. Or subsections, really is this study sufficient to really move this forward.

Now in terms of this particular, I think for credit it's a peer reviewed article but is it sufficient, I think that's the question.

But in terms of the performance evaluation section here that I'm looking at the Cronbach alpha was 0.84 for the patient survey. Reliability of the patient survey was not assessed due to the low number of items.

And in terms of the validity testing I just focused on assessing the domain specific scores and the patient reported measures of quality and trust.

So this is what they're using for their main validity argument and I welcome other peoples comments on all of this. It's I

think our group overall, as you can see up front there, kind of put it in the moderate range for both reliability and validity.

So those are my comments on the key sections there.

CO-CHAIR CORA-BRAMBLE: Thank you. Okay, comments from this work group? Comments from the committee at large? Colette.

MEMBER EDWARDS: Can you just give a few examples of some of the questions?

MEMBER CHIN: I have here, Colette. So from the patients survey, did know whom to call if you want to complain? From the staff survey, senior leaders have rewarded staff and departments that worked to improve communication.

My direct supervisors have intervened if staff were not respectful towards patients. My direct supervisors have monitored wether I communicate effectively with patients.

My direct supervisors have asked for my suggestions on how to improve communications with the hospital or clinic. My direct supervisors have used my feedback to improve communications within the hospital or clinic.

Staff members have spoken openly with supervisors about any miscommunication. Staff members have known whom to call if they have a problem or suggestion.

CO-CHAIR CORA-BRAMBLE: Other questions or comments?

DR. NISHIMI: I just want to let the committee know that Matt Wynia is on the phone right now. Operator have you moved Matt from the audience line to the speaker line?

OPERATOR: This line is open.

DR. NISHIMI: Thank you.

DR. WYNIA: Hello everyone. Can you hear me now?

CO-CHAIR CORA-BRAMBLE: Yes.

DR. WYNIA: All right, first I

want to apologize, I wish I could have been there in person today, I think Dr. Chin knows I'm on service right now. And there's a very strong desire for people not to take time off when they're on the in-patient service, to travel. So my apologizes, but I'm happy to answer any questions that might have might have arisen this morning.

CO-CHAIR CORA-BRAMBLE: Okay, so what, I think we, the best way to proceed is that we'll vote on this measure, and then we'll go back to the questions for you Matt, from the first measure, all right? Okay Alisa.

Oh, questions? I'm sorry. Lourdes.

MEMBER CUELLAR: I just have a quick comment. And I think there was some recent studies and I didn't come to the top of my head right now.

But one of the things we need to keep in mind that patient satisfaction, or how

they perceive their satisfaction in quality of care do not equate. And a lot of times we're seeing now where the goal for many organizations is to get those high numbers of quality of care, of patient satisfaction.

And it doesn't correlate necessarily to the outcome. And so I think we need to keep that in mind with all these measures as well.

CO-CHAIR CORA-BRAMBLE: Thank you. Other comments? Yes.

MEMBER JACOBS: I've just addressed that comment and I actually think even if it doesn't. I'll go back to my earlier comment, even if it doesn't impact outcome, it's still important that patients feel happy and comfortable with the care that they are getting which satisfaction can measure. So I just think that's so important to know if people are actually striving to do.

I met with a patient yesterday and I actually called her, she had some abnormal

test results. She told me, she's 40 years old, first time the doctor ever called her with test results. I was like, "That's really sad."

So I actually think that's really important, and I mean this is someone that who will come back to me regarding this test results because of making that call. So I think it's really important.

CO-CHAIR CORA-BRAMBLE: Okay, thank you. Jerry?

MEMBER JOHNSON: Yes, on the same topic of kind of less important, in validity and what's not, because I think it's going to keep coming up.

Where's the satisfaction measures okay, I'm not overwhelmed with that. I am quite comfortable with the quality measure that is used in a lot of these. As they're kind of a validity standard, it's the patients' perception of quality.

So that may not be as hard in

outcome measure we would like, from morbidity, mortality, disability, but it correlates with a lot of intermediate measures.

So I'm just making a plea for, and I think I'm in agreeing with Elizabeth here, at least when it comes to patients' perception of the quality that they perceive of care that they receive, that that's a reasonable validity standard in studies like this.

CO-CHAIR CORA-BRAMBLE: Okay. Thank you. Yes.

MEMBER O'BRIEN: I guess I'd be interested in hearing Dr. Wynia's response to this question, I'm just curious about the rational for combining some of the patient items and the staff items into a single composite instead of reporting them separately.

In my mind I think of the patient's items as being outcome measures in the sense that they are result of all these structures and processes that are maybe in

place. If that's good communication.

When I look at the single patient measure in this item, and I look at the label attached to it that has to do with performance evaluation. One comment is that if my gut sense of does that match up with what the label of the measure is in terms of the performance evaluation. I'm not sure that a patient response is able to really get at what's in place to measure performance.

You know they can get at the outcome but they can't really answer the question, you know, is the organization taking serious steps to evaluate performance and act on it.

And then this measure itself, 50 percent of the weight for how you'd assess the organization's efforts to measure performance, becomes from the percent of patient responded that they knew who to call if they wanted to complain.

So if you are trying to get at

efforts measure, performance is 50 percent of that and it comes from patients, knowing where they complained.

To me it doesn't, I would think about reporting this separately, and I might think of, you know, domains along the line that are here, but then a separate patient outcome domain in that.

And I would just like to go on to another comment. Is that for me my hangup with it, with any of these is really mainly about the public reporting component. And I don't know this is partly an issue with NQF, and partly an issue with the measure.

It's just that it seems clear to me that this is a, oh, my gut sense is not a content expert. This is like incredibly useful tool for an internal organizational assessment.

You know, no matter what weaknesses it might have if an organization takes this on, I feel they would be likely to

learn something.

So for me it's the public reporting where I start to raise more issues. And if I'm thinking about reporting something and what would consumers out there, you know, who might go to a web site, and look at something they want to know about. I would think it would be more the outcome type patient responses that matter.

CO-CHAIR CORA-BRAMBLE: Okay. Marshall, did we want to get, invite the feedback or wait until we take the vote? I'm sorry? Yes. So, I don't know the name. Matt there's a direct question in terms of the choice of questions that were used for that particular measure. And we invite you to respond to the Dr. O'Brien.

DR. WYNIA: Yes. Thank you. I think he's bringing up a really important point and in some ways it's kind of validating to hear this conversation because it very much reflects the conversation at the oversight

body.

Which is our sort of expert panel that has been working with us on developing and testing these measures for the last five years.

One of the real conundrums is trying to develop an organization wide measure that is reflective of both the patient and the staffs experience. And retaining some degree of simplicity in terms of reporting to the organization and potentially to the public.

So the oversight body has been very concerned that at the end of the day, we are able to give organizations a numeric score. From zero to 100, where 100 is the best, and zero is bad.

And that entails developing this, you know, composite scoring system where we bring together both patient and staff feedback. And their experiences are weighted.

We've gone back and forth with the idea of differential weightings, and in some

domains giving the staff a greater weight, than what we give to the patients' scores.

An the sense of the oversight body was that, that might increase the complexity of understanding the measure to a point where people would start to just get confused. Which we already feel is a risk with some of these measures. Because they are, you know, multi-factorial already.

So that's been the conversation at the oversight body. And that's why we weight and the same. Even though there are a couple of domains I think you probably already looked at one of the other domains where we only really have one or two items, from the patient survey that are directly relevant to that domain.

There are just some domains where the patients' perceptions, patients are not able to see, you know, what's going on sort of behind the scenes. And yet it's a very important issue.

So I'm not sure if that's a full and complete explanation, but I couldn't affirm with you that you're having the same questions as our oversight body has been grappling with. In trying to balance the need for a relevantly simple score where you can say you got an 85, and a 85 is not as good as a 95.

CO-CHAIR CORA-BRAMBLE: Thank you. I think it addresses sort of a global issue. I'm not sure that on the very specific ones, perhaps you Dr. O'Brien felt it addressed this. I didn't think it did, but it's your question.

MEMBER O'BRIEN: Well, I mean I heard their thoughts.

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER O'BRIEN: So, I mean, and I anticipated that these probably were the similar issues that had been discussed and ultimately, you know, when you're developing you have to make a decision and go with it.

And there's many different approaches, I can't make a judgement myself.

CO-CHAIR CORA-BRAMBLE: Okay. Dawn, did you have a question?

MEMBER FITZGERALD: Yes. I, just to go back again, to the issue of validity and the conversation.

I don't disagree with regard to the fact that the patient's perception is an important variable. But I'm approaching validity from a methodologic standpoint, which again validity is yes, does the questions appear to be relevant. But it also goes back to the issue of, you know, when you're considering the administration of that, can you consistently identify the population to whom the survey will be administered.

And I don't think its been made clear that there is any systematic approach to administering the survey that would allow for you to be able to suggest that there would be no bias associated with that selection.

As indicated by the fact that, you know, if you respond to this right after the clinic visit you could potentially get a bias relative to another administrator who surveys a patient a month or a week or sometime in the future.

So, you know, it's two forms of validity, one is the relevance of the questions, but the other is the validity of the way in which the survey's administered to identify population. Which again goes back to where my concern lies.

CO-CHAIR CORA-BRAMBLE: Thank you. Marshall, did you have a question?

MEMBER CHIN: Hi, Marshall here.

I've got a question about would a group try to reduce the number of items in the overall survey as well as each sub-scale.

Especially in the staff components of a lot of these scales. There's an awfully lot of questions which seem to have a fair amount of conceptual overlap, both within

sub-scale as well as we're starting to look across three or four different sub-scales.

So to what degree did your group try to reduce items, because if this goes out, especially, you know, like if these are approved, then they could be pushed back in terms of, usability, feasability issues.

And so I'm wondering to what degree you guys have already explored trying to reduce items?

DR. WYNIA: Yes. May I reply to that right now, or?

CO-CHAIR CORA-BRAMBLE: Oh yes. No, we're inviting your response now.

DR. WYNIA: Okay, thanks. Yes, so that's an important concern, and we have tried to look at whether there are items that are conceptually overlapping or even frankly redundant.

And one of the balancing acts that we're trying to pull off here is that these items are often directly reflective of the

voted expectations that were weighed out by the oversight body. And were voted on by the oversight body and so they, you know, they have a list of things that they're trying to address.

And so we have items that are often specifically addressing those consensus expectations. So part of the validation of the entire tool set, was the voting process. To design, you know, what is it we're going to try to measure? And all those important issues that are relevant across multiple organizations, and feasible to measure, and so on.

And so in developing the instrument, we were trying to be attentive both to not having an instrument that's so long that it's not feasible to carry it out. And also touching on everything that is laid out in those consensus expectations.

CO-CHAIR CORA-BRAMBLE: Okay. Two other comments? Oh, one other comment.

Elizabeth, yes?

MEMBER JACOBS: Hi Matt, it's Liz, this is a question for you, or is it Andy?

MR. JAGER: Andrew.

MEMBER JACOBS: Andrew. I was wondering, to get to this issue around the sampling, and how the sampling was, in this particular paper. There is this issue around like if you ran it -- how many people did you, what was your response rate? Of people who actually participated, both staff and patients? Like did 50 percent refuse, did 80 percent refuse?

DR. WYNIA: Well, Andrew do you want to get that or?

MR. JAGER: Sure I can try and if you want to add something.

So we aren't able to calculate really reliable response rates a lot of times. Because we don't know always who refuses the survey, when it's given out on paper to the hospital or clinic.

In phase one we had nearly 6,000 patient respondents, and almost 2,000 in phase two. And over the last year we've had about 1200 patient respondents.

And for staff, phase one there were about 1200 respondents, 650 in phase two. And over the last year we've had 4,500 staff respondents.

DR. WYNIA: We do know for some of the sites. So some sites, were less stringent in terms of, you know, reporting back to us how many of these they, because there were people who ended up photocopying some off. So I'm just being very blunt.

There were a few sites that photocopied additional ones off. And it wasn't always clear how many of those got reported back. So we do have response rate data from each of the sites.

But I think what Andrew is reflecting is we're not 100 percent confident in those response rate data on the patient

surveys from some of the sites.

The response rates in general ranged in the 20 to 40 percent range for the patient surveys.

For the staff surveys it's more like 50 percent. But again that's quite variable from site to site.

And one of the things we learned over the different waves of field testing, were some ways to improve staff response rate by insuring that the survey was sent out with the appropriate cover letter signed by the right person and so on.

MEMBER JACOBS: I have another question which is related. I am wondering, I can imagine the, I'm not imagining, these questions are very sensitive for employees to respond to, related to their own organization.

And I want to know, did you get any sense of, you know, like some supervisors might say, you must fill out this form and give it back to me. And then they'll will

fill it out. And then they'll say, this is a threat to validity right. They'll say, oh, you do a great job of communicating with me.

I wanted to get a sense of how you monitor that, if you know if any of that sort of social or response bias went on or?

DR. WYNIA: Yes, again a very important, very important issue. And the best we can generally do, just in any survey where you're dealing with a sensitive issue, something we do a lot of, in our other survey work, in ethical issues at the AMA.

Often the best you can do is to give people a clear cover letter that says, this information is not going back to your boss. So the cover letters that go out with these surveys are designed to provide some reassurance that your name is not on this survey.

We're not asking for your name or any other identifying information. And the information will be only reported in aggregate

and not on an individual level. And your individual boss won't see the results of your survey's. It's being sent directly back to the AMA for analysis. Not to the hospital for analysis.

CO-CHAIR ANDRULIS: Romana?

MEMBER HASNAIN-WYNIA: Matt, this is Romana. And I should tell people that this is not a discussion that Matt and I have, you know, at home, or dinner. Actually, it's forbidden because I might lose my mind if this is what we talk about at home. So I'm asking this question following up on Marshall's.

So you describe the process of kind of an oversight body, you know, coming up with these consensus expectations, and then kind of the response items that are on the assessment tool.

I guess I am also concerned about the potential push back from the field, if the burden of collecting data on so many items is so pervasive.

And your response in terms of the kind of, you know, these came because of this expert body that put forth the list of items that should be in this assessment tool, happened prior to testing.

So after you tested the assessment tool in your various hospitals and clinics. Did you see an opportunity to reduce the number of items, based on those test results?

DR. WYNIA: Yes, and the answer to that is yes. We did end up reducing some items. And I guess there's one other aspect to this that I didn't mention in responding to Marshall, which is some of these items remain in because the sites want to know the answer to that particular item. Even though there is some cross-over with other items.

So believe it or not, we more often get responses back from sites that say, Well, could we add some questions? We want to know more about this or that.

And so anytime we try to remove

one of these items it's possible that a site comes back and says, well actually I was really interested in that particular item.

And that happens quite a lot where, you know, the sites still, they do want their overall score, and that's what we are looking at in terms of validation of these as domains.

But they also want to see the results of individual questions, because that's important for quality improvement. That's where they can say, well you know what, the reason our score is low here is because were not doing well on this particular issue.

CO-CHAIR CORA-BRAMBLE: Jerry?

MEMBER JOHNSON: Yes, I wonder if you could help me understand how we convey to the public how they think about the significance of a particular score, of 75 versus an 85, or 50 versus a 70.

I think from a public reporting standpoint, that's going to be crucial. So

higher is better, and there's this continuum, but how much better? Because the answer to that, I would think, would influence what kind of quality improvement efforts an organization should make.

And I can't get a sense from reading any of these of what, I mean, kind of what are the anchors of significance of any of the scores is it.

And I know lower versus higher, but what's the significance, how should the public even think about that?

DR. WYNIA: Yes, that's a nice point. What we were aiming for, and I think we have mainly achieved, is a scale which is something like a traditional grading scale, where, you know, a 70 is probably a C.

Now that's not to say that there are not a few domains where most of the hospitals that have viewed this so far, are getting a little less than a C. And there are other domains where most the hospitals that

have used this so far, are getting a B+. And it's a rare hospital that gets a whole lot of A's right now.

But I think that that's a fair way to think about it. That the average score in the average domain is going be around, you know, somewhere between 60 and 80.

And above an 80, puts you in pretty good company in terms of your performance on any particular domain.

There's a table in the paper, or a figure in the validation paper, that kind of shows the range of scores, on the nine domains, at each of the hospitals.

And one of the things that, that demonstrates is that there were none of these test hospitals, despite the fact that they are, you know, they're very interested in these issues and many of them are, you know, you would expect to be pretty high performers.

But there were no sites that scored uniformly high across all nine domains.

And similarly there were no domains in which every hospital scored either high or low.

So there was a pretty good spread of scores in each of the domains, and there's a generally pretty good spread of scores within a hospital.

Which was what we were aiming for, because the idea of doing a nine domain assessment, is that your hospital finds out that you're doing better than you expected in terms of addressing the language needs of patients. But you're not doing as well addressing literacy issues. Or community engagement could be improved.

So being able to target your interventions to those areas that might need the most improvement, was what we were shooting for. Does that help?

MEMBER JOHNSON: Yes, thank you.

CO-CHAIR CORA-BRAMBLE: Kevin, you had a comment?

MEMBER FISCELLA: Yes, three

questions. First I should say is, I really want to applaud you for this initiative. I think you're getting at really important concepts that often aren't captured in other ways. And I hope this project continues.

So my three questions are, first is do you have plans to issue guidelines to standardize that data collection in the future? Is that sort of in the works, so that organizations do it in a standardized way?

The second question, I realize your n is small, non-organizations for the phase two, but were there correlations between the staff and patient sub-scales?

And the third question, relates to whether there was any, or whether you assessed differential item functioning for different suburbs in terms of responses?

DR. WYNIA: I'll handle the first. Andrew might be able to give you actual data. I'm standing in the hallway so I don't have the data on correlations between patient and

staff surveys. But we do have those data.

And I may need more explanation on the third question. But let me say in terms of standardized data collection, the short answer is yes, we're constantly trying to improve the standardization of the data collection process.

And we're constantly balancing that against the need to do assessments that are reasonable and that hospitals and large clinics and so on, are willing to undertake.

So we entered into negotiations essentially with hospitals when they decide they want to do this. With one of the consultants that we're working with or whomever. And we try to talk them into the most standardized effective data collection method that they are willing to carry out.

And usually it's pretty good. We do have a sort of rank order set of possibilities for how to distribute the surveys for example. And there's a set of

documents that we share with sites that are using this about exactly this issue. How to ensure that you're getting a reliable sample in order to get valid data.

On correlations between patient and staff survey items entered, do you have that available to you?

MR. JAGER: I don't have it at my fingertips, I could certainly forward that when I get back to the office this evening to the committee.

DR. WYNIA: And I can give you a general sense, which is, they are correlated, but not great.

If memory serves, we're talking about correlation co-efficient in the point four range, point three range. Which again, points to the fact that patient experiences, and staff experiences, and perceptions are not the same. Which is why it's important to look at both.

Incidently, I think earlier I

heard someone talking about reporting the patient and the staff measures separately. And just as an FYI we didn't submit these for NQF endorsement, but we do report those data separately.

So the hospitals get back both the staff and the patients scores, separable. As well as the scores for every individual item on the surveys, obviously.

I'm sorry, there was a third question that I --

MEMBER FISCELLA: Yes, differential item functioning for the suburbs?

DR. WYNIA: So are you asking, are there some items that are more important, within a particular domain?

MEMBER FISCELLA: No, if they function differently for example, by patient education for example. Whether those items are formed differently.

DR. WYNIA: Yes, the answer that is that they do, and we often in fact, are now

looking at a sort of stratified analysis.

So we can show people their data according to language. According to literacy level, education level, and so on.

We haven't incorporated that into the scoring, once again because we're trying to keep the scoring as understandable as possible. And if we start giving differential weights to items, and then using, you know, multi variable models to determine the relative weights of each individual's responses.

We felt like that would become a tool that hospitals might not want to use because it would just be to complicated for them to understand what was going on.

But we do report, we are able now to report those kind of data. So that people can see whether folks with, you know, lower education level are reporting similar experiences of care.

CO-CHAIR CORA-BRAMBLE: Okay, two

more comments, and then I am going to ask that we go ahead and vote. Ernest?

MEMBER MOY: This is Ernie Moy, and I was glad to hear about your responses that the provider and the patient components were typically reported separately, because I do think that they are probably capturing something very different and I'm concerned about putting them together into a similar composite?

But along the same lines, I was also concerned about the patient responses in that. For any given sub-domain the number of questions seems to be fairly sparse.

And the other issue that potentially is a confounder is that it maybe, you know, I'm curious about the correlation about the patient responses across the different domains, and to see whether or not they are actually capturing something different or just some kind of generic patient satisfaction, or satisfaction communication

element. And so I was wondering if you could comment on that. Correlation across the different domains.

DR. WYNIA: Yes, I think what you're reflecting on, is also reflected in some of the earlier comments, about tying to reduce the number of items on these surveys.

And we've paid particular attention to item reduction within the patient survey, in part that's because patients are much less likely to respond to a very long survey.

So we've done a lot of item reduction to get this survey to a point where we can get a lot of patient responses, including from patients who may have lower literacy levels, or who don't speak English.

And the trade off there is that we have a number of domains where there are quite a few items in the staff survey that address that domain, and there are a few items relatively, from the patient survey in that

domain.

CO-CHAIR CORA-BRAMBLE: Okay, Luther? Then Mary, yes?

MEMBER CLARK: Sure. My question actually is in the same lines as the one Jerry ask. And it really has to do with the expected significance of a change in score. So if one gets a, does the survey has a base line score to identify issues that need to be addressed.

So that would be certainly important in terms of knowing that they exist, and will give you some perhaps measure of how compare to others. But once you introduce some corrective actions or measures, what level of change in score would you say would represent importance or significance or targets for improvement?

DR. WYNIA: We've considered a five point change in score, to be what I would think of as clinically significant.

Given the numbers of surveys and

so on, that we've had back, we had statical significance at lower levels than that, so a change of one point would be statistically significant.

But I think a change of five points is clinically significant. And I say that because we did analysis looking at the relative change in patient reported quality and trust, with a five point difference.

And for most of these domains there were really quite substantial changes in the odds that patients report. Quality care and trust in the organization. When there's a five point difference.

And I think Andrew probably has the table, or the chart in front of him. But for many of these domains we would see for example, a 30 or 40 percent increase in the odds that patients would report quality care and trust in the organization with a five point change in the domain score.

CO-CHAIR CORA-BRAMBLE: Okay.

Mary, and then Dawn.

MEMBER MARYLAND: So my question is you indicate that the number of measures were decreased, the number of questions, so what impact was on the reliability and validity once you either combined or decreased questions, was that looked at?

DR. WYNIA: Yes, so that was done. That whole process took place during the first phase of the validation. So it's incorporated into the process of checking those alpha scores.

So we would run the alpha score and then we would remove a few items and see if we were still getting the same alpha score or good enough. And that was how we did item reduction through that first round.

CO-CHAIR CORA-BRAMBLE: Okay, Dawn.

MEMBER FITZGERALD: Matt, perhaps you can help me a little bit in terms of responding to a question that continues to

concern me.

And it has to do with earlier on you were talking about we asked you about the consistency or the desire to put some more parameters around the administration of the survey in order to be able to consistently validate the results using a standard protocol.

And your response was a very valid one in saying that, you know, putting too many restrictions on it makes it difficult to administer and it could be more complicating for the providers.

But on the other hand this measure is now up for NQF endorsement. And as such means that, you know, that all of a sudden now we are talking about providers having to be accountable for the measure in terms of potentially expectations from payers, providers, et cetera.

Given that, that's what this endorsement means, are you still as

comfortable with the level of flexibility you have in terms of the administration of that survey? Or given that level of importance, would you reconsider your response?

DR. WYNIA: I think again, I hope I'm not speaking out of turn, or just being to blunt here, but frankly NQF endorsement kind of changes the calculations, I hope.

So my hope is that with NQF endorsement we are able to implement more stringency in what we can require, and people will be willing to go along with it. Because they want to do a measure that they think is going to be helpful to them.

We haven't had that in the past, and I think we've responded as best we can, to try and maintain a level of integrity in the data collection process. While being responsive to the needs of hospitals who are doing a lot of other stuff and have, I'm sensitive to the demands that hospitals are facing for performance measurement.

And so this has been a completely voluntary activity that these sites have done because they're particularly interested in insuring that they're providing high quality care to every patient who walks through the door. And I applaud them for that and want to help them.

That said, if we have NQF endorsement behind us, I think we gain leverage, in insuring that the data collection process steadily improves.

CO-CHAIR CORA-BRAMBLE: Okay, thanks so much. Last question or comment, Luther?

MEMBER CLARK: No, I don't think so.

CO-CHAIR CORA-BRAMBLE: Oh, I'm sorry. One more, Luther, no? Okay Marshall.

MEMBER CHIN: Okay, so this may be a question for Helen. About process, you mentioned that after we vote we can go back to AMA and that and Andrew in terms of

conversation and suggestions.

And this may be new terrain for you, because the aspirant for an MI is different than a complex survey like this in terms of the issues. Okay, okay. The question is, as for doing this voting, how substantial can these recommendations be?

For example, this is really helpful, the information that Matt and Andrew are supplying and my guess is that if they knew what the answer was going to be used for accountability purposes, for NQF endorsement. They would have done a different survey in terms of the way they did this and all.

And so that, I can always think of a pretty substantial recommendations we might have based upon some of these questions. Which are entirely doable, but would be a substantial amount of work.

But which is do-able with the existing data base, and would probably lead to scales which are pretty different than the

ones that are there right now.

Which again, in hearing Matt's latest answer, you know, I think his philosophy would probably be similar to ours in terms of the different purpose than what the scales were originally designed for.

So that's on the table, so if we vote yes, for example on these sub-scales in their current form, that still leaves the open potential for some fairly substantial revision before they actually really get NQF approved.

DR. BURSTIN: It's a really good question, I mean, this is complex stuff these are tested surveys, so, you know, I think when there are minor tweaks, and I don't think we are talking minor tweaks, it's perfectly reasonable the developer might be willing to say I can adapt to some of those minor tweaks.

I think what you're talking about is a pretty different survey. And that's not what we would be doing in the terms of this process.

MEMBER CHIN: Well, let me see if I've got an example. I think with the existing questions they could basically re-do some analysis to come up with, for example, a more parsimonious data set.

So for example, Matt's answer was basically a committee wrote this, that's the way it comes across right now.

But if NQF said you had to basically do a know what, you know, a really parsimonious data set. They would do it.

That's one example, and there are other ones in terms of the staff versus patient question issue. The administration issue that Dawn brings up, and there's a variety of things that I think are really important, but and they are eminently do-able. But it will take some work.

But it would probably be a much better instrument than as currently here. And once it goes out NQF endorsement and all. As Romana said there could be some major push

back if, and Norman's point about the usability by commissions.

That, you know, if a bad instrument gets out there in terms of feasability, you know, this is the one chance in terms of when it first comes out.

DR. BURSTIN: So it's actually two answers, the first is, you can make any recommendations you so choose, it's certainly up to Matt. Hi Matt, its Helen, for the developer to take them under advisement, see if they think are the things they want to change.

In the terms of this project though, the issue would be how quickly could they actually potentially do any of these changes, re-analyze the data, re-analyze your liability.

I mean this is where it gets to be to the point where it just might be difficult enough that it all so often times the developers will say, really helpful input,

I'll go back to the drawing board, make those changes.

Now I will also point out that is also very common and appropriate that as the measure gets put out into the fields for wider use and often an NQF endorsement does lead to that wider use.

There's often experience or implementation that leads to significant improvement in measures that we always happy to take those improvements.

We can do an ad hoc review at any time they can be submitted through an annual update process, or as part of the three year maintenance. So there are opportunities to continue to iterate and improve the measure moving forward.

And I think you have to decide if basically what you have at hand, is it, does it meet the threshold. And I think the question for the developers, how much is doable within the time frame of this work.

MEMBER WASHINGTON: Hi Matt, it's Donna. I'm sorry, were you commenting?

DR. WYNIA: No, I was asking if I should comment.

CO-CHAIR CORA-BRAMBLE: Donna's going to raise an issue or ask a question, and then we'll invite you to comment, Matt.

DR. WYNIA: Thank you.

CO-CHAIR CORA-BRAMBLE: Hi Matt, it's Donna Washington. I'm apologizing if I'm asking a question that was addressed a couple minutes ago when I was out of the room but one of the criteria had to do with usability of these measures and improving performance.

And I wonder, when I looked at the web site and it looks as if the AMA suggests use of paid consultants for organizations to help interpret their results and target them toward performance improvement.

I wonder if you have any data either from these paid consultants or from any other related studies on, Number 1, how useful

these scales have been in quality improvement. And Number 2, if these measurements tools are responsive then to these interventions?

DR. WYNIA: Donna, thank you for giving me the chance to talk about that actually, it was not raised earlier.

We do have some data on this because we have a few sites that have used the tool several times now. So in the way we intended them originally to be used.

Which is to say you check performance, you do some interventions, you re-check. And we have some really interesting information on that.

We actually presented this at SGIM last year, on one of the domains where several of the sites have tried to address their relatively low score in health literacy.

And we learned, Number 1, that just because you measure something and try to improve it, doesn't mean it will improve.

So we had three sites that did

specific interventions to try to improve their score in the health literacy domain.

One of them saw very substantial improvement, another saw basically no change, and a third actually got a lower score on their next assessment.

And so we went back to those sites to try and figure out why did your quality improvement effort work, or not work.

And there are a number of things, none of which will be surprising to any of you, having to do with leadership commitments and support for the interventions and so on that were probably at play in terms of why some organizations are capable of taking performance improvement information and using it, or assessment information and using it for performance improvement. And others have a more difficult time.

CO-CHAIR CORA-BRAMBLE: Okay. Are we ready to vote?

MEMBER CHIN: Matt, I think you

were willing to answer, respond to my question and comments?

DR. WYNIA: Oh yes. I'm sorry Marshall, yes, I would like to say something about that. Because we are always looking for ways to improve these instruments.

So even over the last year, once the Joint Commission Roadmap Document came out, for example, we went back to the instrument to see whether there were things that we could add or tweak, to be sure we were attending to all of the issues raised in the road map document.

As you know, the CLASS standards are about to come out with an enhanced version. We are going through these instruments to make sure that we are addressing each of the issues in the enhanced CLASS Standards.

So the idea of continuing to improve the performance of these instruments, over time, is absolutely on the table for us.

And it's often just contingent on having faith that are using the tools at any particular time. So the more sites we have using them the more opportunity we have to continue to test and re-test and make improvements over time.

CO-CHAIR CORA-BRAMBLE: Okay, we're going to go ahead and vote. And Adeela, I'll pass it on to you.

MS. KHAN: So looking at in points to measure importance, was the threshold criteria and importance to measure and report met? Press one for yes, and two for no. You can start voting now.

We have two people missing. One more. Okay, we're all set. We have 19 yeses, and zero noes.

And looking at reliability, to what extend was the sub criteria on liability met? Press one for high, two moderate, three low, four insufficient information. You can start voting now.

And we have one person missing, so if you could all just enter it one more time. All right.

We have zero for high, 15 moderate, four low, and zero insufficient information.

And looking at validity, to what extend was the sub criterion validity met? One high, two moderate, three low, four insufficient information. You can begin your vote. There's two more people. There we go.

Zero high, 13 moderate, six low and zero insufficient information.

So we're going to move on to usability. To what extent was the criteria usability met? One high, two moderate, three low, four insufficient information.

We have one high, 13 moderate, three low and two insufficient.

Going back to scientific acceptability of measure properties. So was the criterion scientific acceptability of

measures properties met? You can press one for yes, two for no. There we go. We have 13 yes and six no.

And going on to feasibility. To what extend was the criteria in feasibility met? One high, two moderate, three low, four insufficient information.

DR. WYNIA: I don't know if I'm still on the open line, but I actually need to go. I can come back in about 20 minutes, if that's okay?

MS. MCELVEEN: Yes, that's fine. Thank you, Matt.

DR. WYNIA: Okay, I'll call back in a little bit.

MS. KHAN: For feasibility we have one high, 14 moderate, three low and one insufficient.

And overall suitability for endorsement does this measure meet the NQF criteria for endorsement. Press one for yes, and two for no.

And we have 14 for yes, and five for no. So the measure will pass.

CO-CHAIR CORA-BRAMBLE: Okay, what we're going to do, is we're going to stop now, take about a 10 minute break and by the time we get back Matt then can re-join us on the line in case there are any further questions.

(Whereupon, the above-entitled matter went off the record at 11:28 a.m. and resumed at 11:41 a.m.)

CO-CHAIR CORA-BRAMBLE: All right, we are going to get started. We're going to do a few things, we're going first address Measure 1905.

Then we're going to go back to the first measure that we did, 1881. And we will invite feedback from the author. And then we're going to deal with public and member comment.

So Measure Number 1905, our lead person there would be Kevin.

MEMBER FISCELLA: I'm not sure how

these assignments were made but --

CO-CHAIR CORA-BRAMBLE: You know I sat in a grant review committee with you, you never said that.

MEMBER FISCELLA: But actually this is something near and dear to my heart. This concept of leadership commitment in the domain of communication deployment assessment.

One of the challenges in looking at this, was I thought that the, unfortunately the evidence that was supported was fairly generic to the item. To the issues surrounding, you know, the importance of communications and disparities of quality.

When in fact I think there is a fairly compelling body of literature showing that leadership commitment does matter, in terms of what organizations do.

And I think anybody in organizations knows that intuitively. And there's no whole organizational management literature on that, that I think would have

been helpful to cite.

And that certainly affected my scoring. I think the committee gave it a four highs and two lows.

In terms of performance gaps, actually this had the highest delta of any of the sub-domains of 9.4, between the highest and lowest performing organization.

Let's see, in terms of reliability the Cronbach's alpha's were quite high. Probably given the number of items here, a .87 for the patient and .91 for the staff survey.

The issues in regarding usability and feasibility really are no different than the previous ones discussed.

Just to give people an idea of what we're talking about, some of the questions for the patient ones, sort of had to do with, a sort of climate.

It wasn't easy to ask questions at the hospital with information in the waiting areas helpful? Was it easy to reach someone

on phone? Do you feel welcome?

Are you happy with the care you got? Does hospital clinic communicate well with patients? Would you bring a family member to the hospital or clinic?

And then for staff items, really some of them were directly focused on senior leadership. Has senior leadership that taken steps to create a more welcoming environment for patients.

They've taken steps to promote a more patient-centered environment. Have make affective communication with the diverse populations a priority. They've rewarded staff and departments that work to improve communication.

So I, you know, I think a lot of these items have, at least in my view, pretty strong, at least face validity.

So I think, certainly relevant to the evidence sub scales. I think that this one certainly is quite important and performs

well relative to the other ones.

CO-CHAIR CORA-BRAMBLE: Okay. Thank you, Kevin. Comments from this particular work group? All right, comments from the group at large. Liz.

MEMBER JACOBS: I mean, how much of an overlap is there with this measure versus the other measure? I'm somewhat confused by how distinctive measures are.

When I was reviewing them I felt, I don't know if you or anyone else in the workgroup have a sense of that.

MEMBER FISCELLA: You know, I think, certainly from the staff survey, I think they are, they do get out a fairly unique domain, in terms of how staff perceive leadership.

I think for the patient ones, I suspect that there's quite a bit of overlap. And I would bet that the correlations are going to be quite high with other sub scales.

MR. JAGER: So as was pointed out

these, the patients items there are a good amount of overlap. Because they tried to keep the survey quite short and at a level that was not too complex.

Because we're trying to make sure this is accessible to people with lower literacy and people who may not speak English well.

So there is a good amount of overlap. But we do think we're measuring this great domain. And as you can see, there's not uniform performance in any one domain at all sites.

Or, you know, ones, I think sites that are uniformly well, or uniformly poor. And I don't have the coefficients here, but I can send them. Regarding the correlation of the domains.

DR. WYNIA: Folks I just came back on the line.

CO-CHAIR CORA-BRAMBLE: Perfect timing, thank you.

DR. WYNIA: Sorry I got paged away.

MEMBER JACOBS: Do you use one set of items for patients and then, like are there similar, are there over, like do you use one like patient item in multiple measures? Is that what you're saying? Because you kept it short?

DR. WYNIA: Yes, that's correct.

CO-CHAIR CORA-BRAMBLE: Your colleagues from the AMA was taking a stab at it. But it's okay, however you want to do it.

DR. WYNIA: Sorry.

CO-CHAIR CORA-BRAMBLE: Okay, Matt, you go ahead.

DR. WYNIA: No, I think you got it exactly right. There, some of the items are in, you know, they contribute to multiple domains.

CO-CHAIR CORA-BRAMBLE: Okay. Any other comments from the group, before we vote? Yes, sir?

MEMBER JOHNSON: Yes, it's a question about the clear distinction between the staff questions, which are focused on leadership. And maybe even use the word leadership.

And the patient questions which are oblique, they don't really focus on leadership.

Was that because you did not think it was appropriate to ask patients directly about leadership, which is what this is suppose to be about? Or did not work and you cut them out?

Or are they just, I'm impressed by the fact that the patient questions are not focused on leadership but the others are.

CO-CHAIR CORA-BRAMBLE: Matt, your response.

DR. WYNIA: Yes, the answer there is that we didn't ask questions directly about the leadership of the organization to the patients.

On the assumption that this was an area where, you know, patients probably wouldn't know whether the senior management was supportive of something or not.

They would have experience with the people that they interact with. The caregivers and the other staff.

So the best we could do was ask them about the things that we expected them to have some experience with.

CO-CHAIR CORA-BRAMBLE: Okay, any other questions?

MEMBER LU: Yes.

CO-CHAIR CORA-BRAMBLE: Yes, I'm sorry. Francis.

DR. WYNIA: Well it just dawned on me as a side note. This same issue arises in a couple of these domains. Where the patient items are more oblique than the staff items.

And the other ones where this comes up are the performance improvement domain. And the training domain where you can

expect staff, you can expect the patient to know whether performance along the important dimensions of training is occurring, but you wouldn't expect them to know whether training per se, had occurred.

CO-CHAIR CORA-BRAMBLE: Okay. Francis?

MEMBER LU: This is just more of a comment, in that this whole area of leadership commitment, assessment, I would say is one of the prominent parts of the class enhancement initiative.

And so in that one of the additional standards that's being put forward specifically addresses this. And others also strengthen this whole are of leadership and organizational commitment as part of that effort.

So I just wanted to add that additional information, in the sense that that's another body that's looking at these kinds of topics. And to have some kind of

cross walk to help assess this aspect of things I think would be quite important.

CO-CHAIR CORA-BRAMBLE: Excellent comment, Francis. Okay, Donna.

MEMBER WASHINGTON: Hi, this is more a comment rather than a question. In just looking across the domains, this domain as well as the others. The results of the validation study, I was struck by how closely clustered the scores were for several of the domains.

And now looking specifically at the items for this domain and thinking about the fact that the patient questions really may not be giving that leadership but may be measuring more generic satisfaction, communication type thing that are addressed with questions in other domains as well.

Then it just sort of suggests that perhaps some of the domains such as this that may have been better assessed by staff alone. Have results that are more attenuated by

including the patient items.

So I guess it is more of a comment as well as a question. I just wonder if you could respond to that, Matt?

DR. WYNIA: Sure, I think you're correct that the scores are somewhat attenuated as a result of combining the staff and the patients and we get some degree more variability in the staff scores than we do in some of these patient scores.

Partly because the patient survey is shorter and therefore there are fewer items to be incorporated. And partly because there are some of these domains where we're really only able to get patient feedback in kind of an oblique way.

In terms of looking at the outcome of an organization that is committed. Rather than asking directly about is this organization committed? So that's not really an answer to your comment but more an amplification.

I think you're right, the trade off here is that we felt like it was important to include both patient and staff data in each of these domains.

Because the idea of the entire instrument as a whole, is that we're doing a 360 evaluation, that incorporates input from staff, from leaders, from patients.

And that all of them count. All of their experiences count in these domain scores.

CO-CHAIR CORA-BRAMBLE: Okay. Thank you. I'm going to ask that that we then get prepared to vote.

(Off microphone comments)

MS. KHAN: So looking at importance to measure and report was the threshold criteria in importance to measure and report met? So you can start voting now. So we're waiting on one more person if you want to click again. We have 19 yeses and zero nos.

Moving on to reliability, to what extend was the sub-criterion reliability met? You can start your vote. One high, 16 moderate, two low, and zero insufficient information.

And looking at validity, to what extent was the sub criterion in validity met? You can start voting. So we have zero for high, 13 moderate, six low and zero insufficient information.

And measuring scientific acceptability of the measure properties, was the criterion scientific acceptability of measure properties met? Yes or no\. You can start voting. Fourteen yes, and five no.

Moving on to usability, to what extent was the criterion usability met? You can begin your vote. We have three high, 12 moderate, 3 low and one insufficient.

And feasibility, to what extend was the criterion feasibility met? You can start voting. Three high, 13 moderate, two

low, and one insufficient information.

And overall suitability for endorsement, does this measure meet all the NQF criteria for endorsement? Yes or no. We have one person missing if you want to try that again. There we go. So 14 yes and 5 no. So the measure passes.

CO-CHAIR CORA-BRAMBLE: Okay. Thank you. So we are at public and member comment. I don't know if there is anyone? We need to go back to the first measure.

DR. NISHIMI: Operator, can you open the participant line and inquire if there's any public comment?

OPERATOR: For public comment from the phone line hit star one on your telephone keypad. We have no responses.

CO-CHAIR CORA-BRAMBLE: Okay. Thank you. We're going to go back to the first measure that we considered and find out, now that Matt is on the phone, see if there any additional questions. So that would have

been Measure 1881. Any questions, I think some of them came up as we discussed the other measures. Liz, and then Mara.

MEMBER JACOBS: I was surprised that the other three measures passed and this one didn't. Because it seems like the conversation was similar on the issues and was similar for all of them. So I wonder if people think we should re-vote on that one?

CO-CHAIR CORA-BRAMBLE: I have no problem with that, you know, I'll defer to the group. So we'll go around the table, Donna, actually you first, then Donna, then Marshall.

MEMBER YOUDELMAN: Thanks, so I'm just going to reiterate that I technically have a conflict because I was on the advisory panel that drew up the consensus report which then gave rise to the CCAT.

But I ask because I also have done a ton of work on data work generally. If I could speak generally and not specifically to this standard. So I was given that answer but

I do technically have a conflict.

I also was surprised and I've had to be silent all morning. But exactly what Liz said, that the others passed and this didn't because to me if --

CO-CHAIR CORA-BRAMBLE: Wait a minute, if you're at conflict I'm not sure that you can comment.

MEMBER YOUDELMAN: I'm allowed to comment generally about data collection, correct?

CO-CHAIR CORA-BRAMBLE: I know but you're saying you're saying you're surprised the measure passed.

MEMBER YOUDELMAN: Right, strike that. Sorry. Realigned okay.

CO-CHAIR CORA-BRAMBLE: Comment in general terms, not specific to these measures.

MEMBER YOUDELMAN: Thank you. I have something to say. To me the data collection is really critical so that we have that.

Because in looking at any other measures, AMA's measures, anyone's measures, so any NQF measure. That we need the baseline data to identify if there are disparities.

So to me it really was surprising that while I think there's general evidence in the field. I'm trying to be careful here. Of the importance of data collection as we've seen from the IOM report. Unequal treatment from the IOM development of standards on data collection from the office of minority health adopting data collection standards.

So I think there has been a lot of work, and the Joint Commission requiring data collection from hospitals. So I think overall my sense is there is lots of evidence and support for this type of data collection generally. And I just wanted make that statement, I'll shut up now.

CO-CHAIR CORA-BRAMBLE: Okay. Thank you. Donna, and then Marshall.

MEMBER CHIN: Yes, there's no

question that data collection is an important item. The issue is that this was not a good measure. So this was the only one of the four that none of the three validation criteria measures correlated in a positive manner.

And this issue of accountability, we don't want to get a measure up there that could be used for accountability purposes that isn't validated.

So my suggestion for Matt is that it would be great to have a re-do. Such that you're doing ongoing data collection and ongoing surveying.

And I want to recommend that for this particular sub scale you eliminate all the provider staff questions. Just pick three or four patient questions that ask directly at these issues. Perhaps updating with new IOM chronic conditions. But that face validity alone probably, I think would be strong.

MEMBER WASHINGTON: I was going to advocate for re-voting with the new

information we had from Matt.

CO-CHAIR CORA-BRAMBLE: I have no problem with re-voting, I just want to make sure if there any other comments or questions directed at Matt? Colette?

MEMBER EDWARDS: A question I had had to do with the likelihood the plan of updating some of the questions in light of what had been mentioned before you were on call, Matt, about some of the questions that have been released by IOM.

And the other is the likelihood of the surveys coming with, I won't say a caveat, but a recommendation that at least the first time out of the gate it be use for internal use as opposed to public reporting. Before it starts getting into a scenario of pay for performance or anything like that. Because I would have some concerns about that.

DR. WYNIA: Is it appropriate for me to reply now?

CO-CHAIR CORA-BRAMBLE: Sure, you

can reply.

DR. WYNIA: Okay. On the last point raised about, not the IOM standard. The instruments don't ask in what way the data are being recorded. So what we can gather from these surveys is whether patients believe they've been asked about their race, ethnicity, language.

And we can ask whether they've been asked in a way that is sensitive. We are not asking them what type of categories are being used for example.

There was a second point that you made and I can't remember what it was now.

CO-CHAIR CORA-BRAMBLE: Colette.

MEMBER EDWARDS: The question had to do with the way that it is going to be used. Internally versus --

DR. WYNIA: Yes, so we actually already recommend that sites not report these publically right off the bat. And it's not difficult to make that recommendation and

virtually everyone is happy not to report things the very first time they ever measure them.

So that's been the standard already. We wouldn't preclude someone from publically reporting. Our requirement I terms of the contractual requirement when someone says that they want to use the tool is that if they were to publically report they have to report all of the scores and not just the ones they like the best.

MEMBER EDWARDS: Thanks, and I'm also voting for re-voting.

CO-CHAIR CORA-BRAMBLE: Yes, we will re-vote, that's for sure. I just want to make sure that I cover everybody's comments. Dawn?

DR. WYNIA: Did Marshall have a question also that I've forgotten?

MEMBER CHIN: I just suggested why not just make it simpler in terms of removing some of the staff questions and just having

three or four patient questions to get at the domain?

DR. WYNIA: Yes, I guess what we're getting from the staff questions are issues around training and the appropriate collection methods. So the patients we're really just asking them whether the data were collected and were they collected in a way that was sensitive.

From the staff we can gather information about whether training is taking place. And whether the organization as a whole sees data collection and analysis as an important task for the organization.

CO-CHAIR CORA-BRAMBLE: Okay. Dawn.

DR. WYNIA: They are a little different.

MEMBER FITZGERALD: I only had a couple of comments with regard to perhaps why the voting was different for this one than the others.

One was the recognition that this one did not correlate with any of the measures that were sort of used to impart the high degree of rationale behind using the measure was the first one.

And then the second one had to do with at least in my opinion, going back to Marshall's comment about this staff collection questions are very subjective and I'm not sure that they're really capturing what it is we think.

It's not a do you collect information it's how often in the last year did you collect and that's a very subjective question.

If I were going to ask staff questions I'd want it to be sort of the more objective measure of actually collection of race/ethnicity to be of value.

And the issue the questions that talk about training there's really only two staff questions that relate to training on

data collection. So again, I would have I think the consideration of what those questions look like may have played a role in some of the voting.

CO-CHAIR CORA-BRAMBLE: Okay. Let's just, no maybe we won't vote, Liz, go ahead.

MEMBER JACOBS: I want to go back to something that Marshall said, with what you were saying, Dawn. It's that I actually think they did not find correlations.

Not to criticize your science, Matt, but you didn't find correlations because they weren't the right things to use to validate the impact of these measures on what's happening in terms of quality.

And that they are extremely important and have great face validity given what we know from Romana's work and things that Mara just said.

So I just want to throw out there that I think given their importance that maybe

there weren't these correlations but it's not because they aren't important but because they were just measuring it against the wrong standard or the wrong reason for converting validities. So I just want to throw that out there.

DR. WYNIA: Yes, I just want to, after what was said, I believe that we also did not use the right criteria for conversion validity there.

It is not at all clear that an organization that does better at collecting race/ethnicity data, which is what we're hopefully measuring, will by virtue of that activity hold greater trust and be seen as providing higher quality care.

CO-CHAIR CORA-BRAMBLE: Okay. Around the table, I cannot see the name tags, so I can't. So Romana, then Grace.

MEMBER HASNAIN-WYNIA: I just want to kind reiterate Dawn's comment and also Marshall's in terms of Marshall's relating to

going back and simplifying the questions. But Dawn's in particular around the, you know, kind of, what is the frequency of asking staff about the frequency of their data collection, is really not going to provide valuable information.

Because even prior to ten years of work post IOM unequal treatment report. There's a strong tendency to say that we're collecting these data. Hospital, 80 percent of them were saying they were collecting it ten years ago.

So it's not about the self reported are you collecting it. I think to Dawn's point it's much more important to know whether they're being trained. Whether they're collecting the data in a systematic way.

So that particular question at least from my perspective and the work that I have done does not provide valuable information.

CO-CHAIR CORA-BRAMBLE: So I'm hearing sort of consensus around that same point, so is there a different issue to raise opposed to the one that you just raised Romana. Grace.

MEMBER TING: Right. Suddenly had a brain freeze. But I think in terms of linking collection of data to quality that's only one dilemma but I'd also like to see possible validation to the provision of actual language support services.

You can collect the data but is it leading to better quality, one, but two, improved services. Which through other measures we'll also seeing linkages to hopefully outcomes. I think that might add to my comfort level in terms of validity as well. So I'd like to see that in future iteration as well.

CO-CHAIR CORA-BRAMBLE: Okay, Kevin. Last comment.

MEMBER FISCELLA: Yes, I was just

going to say, I think the lack of correlation is pretty good example of divergent validity. So in some ways I think it actually supports it.

CO-CHAIR CORA-BRAMBLE: Okay. Let's go back to voting for that particular measure. It would be Measure 1881. And now that you have the correct name tag, Adeela. I've been calling her Alisa all morning.

MS. KHAN: Okay. Looking at the importance to measure report was the threshold criteria in the importance to measure and report met? Yes or no, and you can start voting now. So we have 17 for yes, and 2 for no.

And moving on to reliability, to what extent was the sub criteria in reliability met? You can start voting. We have one for high, 14 for moderate, four for low.

And looking at validity, to what extent was the sub criterion validity met?

You can start voting. And we have 7 for moderate and 12 for low. So the measure doesn't pass.

CO-CHAIR CORA-BRAMBLE: Okay. Deep breath. We are going to go on to the next set of measures also from the AMA. This one has to do with community assessment and engagement.

MEMBER CHIN: This is a point of order. In terms of the part where we come up with our suggestions for Matt and Andrew. When would we like to do that? Is it good with them both here right now?

CO-CHAIR CORA-BRAMBLE: I think some of the suggestions have been captured. My suggestion is that you actually write them down and submit them. All right.

Okay. Measure 1886, Ellen Wu.

MEMBER WU: That's me. Okay. So this is around measuring community engagement. And it's part of the same set of survey questions that we just talked about. So it's

it's measuring community engagement and it's part of the same survey that we've just discussed.

I actually, as part of our advocacy work we really look at community engagement by a facility. And it has been hard to get a handle on that. So it was really good to see that there are efforts to do so.

So it's essentially how well the facility establishes a relationship with the community groups and provides opportunities for engagement.

I guess only two of us, is that it basically indicates? That two of us voted on this, who were assigned to review the measure. And then there were differences in the results so the average result finding was 77.8. And the lowest was 68.3, and the highest was 83.1.

And it showed that a five point increase in the measure results in more than a

50 percent greater odds that the patient would report receiving high quality medical care.

And there are three questions combined for the patient survey, the survey for the patients. And two questions for the staff survey. Did that make sense?

The three questions that they're using for the patients piece is, did hospital clinic staff help you find community resources? Does the hospital clinic serve your community well? Does the hospital clinic staff come from your community? Those are the three questions they used.

Do you guys want to hear the staff ones? All right. The staff ones, overall how would you rate the hospital clinics level of involvement in the community? And over all how would you rate the hospital clinics efforts to help patients across community resources?

So these questions actually track really well to the measure that they're trying

to get at.

CO-CHAIR CORA-BRAMBLE: The question that I would have is does it really measure community engagement?

MEMBER WU: I think it's really hard to measure community engagement.

CO-CHAIR CORA-BRAMBLE: I agree, I concur. I just want to know, you know, having the staff represent the community is one thing. Community engagement is something else.

MEMBER WU: I totally agree.

CO-CHAIR CORA-BRAMBLE: Okay. Other questions, comments, Kevin. No comments, really? Liz.

MEMBER JACOBS: I was just going to say that, you know, even if it doesn't measure community engagement, I think that the perceptions of the community hospital, that's actually something that the hospital could actually do something about if they see this happening.

Even though it's not community engagement and they might do focus groups and figure out why. So I think actually it could be a really useful measure.

CO-CHAIR CORA-BRAMBLE: I totally agree. We were commenting here whether it's a misnomer whether to say that it's community engagement. I agree, it is an important measure, but is it community engagement? Something to that effect. Other comments, thoughts.

DR. WYNIA: I think my line is open again. I lost you for a while, I'm sorry.

CO-CHAIR CORA-BRAMBLE: That's okay. There's no questions though that, I didn't hear any questions that were directed specifically to you Matt. I think we're moving along okay. Yes, Jerry.

CO-CHAIR CORA-BRAMBLE: I guess what troubles me about this most is when it comes down to, maybe we haven't gotten to the

reporting part yet. Where we actually have an impact on these hospitals.

It's just that the definition of community, and to the extent that a hospital is going to be graded as low in engaging or interacting with this community.

That and the change it would then take to large extent depends on how it defines it's community and it's stakeholders. Right? And it's not simple in this influx, particularly in today's world and you have interacting hospitals and systems.

I just don't know, I love the concept of community engagement, and that's fine. But as a performance measure I think it's going to be problematic.

Because even the shapes of communities, it's not geographical. I don't know, this bothers me from a performance standpoint.

CO-CHAIR CORA-BRAMBLE: Okay. Liz.

MEMBER JOHNSON: Just to add to that, Jerry, now that you bring that up, you'd have to know who's answering this question, right? Because you need to know what community they're representing. You're right, I actually didn't think about that point, but that's a really good point.

So I was thinking if you went to Cook County Hospital there's several different communities that frequently go to that hospital and you'd have to know who the patient was to say, okay we're not engaging with this community.

Maybe doing great with the Latino community but if you're only measuring the African American community and it's low you might be doing well in their perception. And that is an issue. That's a problem with the measurement. I agree.

CO-CHAIR CORA-BRAMBLE: Okay. Comments form either the workgroup or the group at large? Any other comments before we

vote? Donna.

MEMBER WASHINGTON: Just looking at some of the items, perhaps it is a misnomer, I would have expected to see items related to community member involvement, in key stakeholder committees. In patient resources and so forth.

And so it sort of goes back to an issue I raised earlier about how hospitals will use this information. Maybe it doesn't matter that it's a misnomer to look at the items and perhaps target their interventions to the items but it just seems like a missed opportunity.

CO-CHAIR CORA-BRAMBLE: Agreed. Okay. Let's get ready to vote.

DR. WYNIA: Is this a time when I might say something?

CO-CHAIR CORA-BRAMBLE: Sure.

DR. WYNIA: I just wanted to be clear that sites do receive the data back with stratified analysis if those are appropriate.

So that if for example, there were differences in the perceived level of community engagement according to ethnic or racial groups. That's something that it can be looked at. Using data and often is.

And in terms of community members on committees, that is addressed, it's not addressed in these surveys. So there's a work that the sites also do and unfortunately those go to a much more qualitative.

And so they don't get incorporated into the scores. But in terms of quality improvement you get the score but you also get this qualitative data which do include issues about having community members on committees and so on.

CO-CHAIR CORA-BRAMBLE: Okay. Let us then vote.

MS. KHAN: So again, was the threshold criterion, importance to measure and report met? Yes or no. You can start voting now. We have 17 yes, and two no.

Reliability, to what extent was the sub criterion in reliability met? High, moderate, low, insufficient. You can start voting. And we have 15 moderate and four low. Zero highs, and zero insufficient.

And validity, to what extend was the sub criterion validity met? You can start your vote. We have eight moderate, ten low, one insufficient, zero high. So we stop, the measure doesn't go forward.

CO-CHAIR CORA-BRAMBLE: Okay. Deep breath, the next one, 1892.

MEMBER CHIN: Denice, Romana and I were just talking and we wondering if we're being consistent as a committee. For example, why was this one not passed? I'd just curious in terms of the main actuators who are calibrated consistent with criteria across the different sub scales.

(Off microphone comments)

DR. BURSTIN: It failed on validity only. The question is, it would be

helpful as well for the report, to explain how the committee thought this one was particularly less valid than the other ones with the same methodology.

CO-CHAIR CORA-BRAMBLE: Okay. Comments from the group?

DR. NISHIMI: I mean, I guess what would be useful is those that voted low, why they felt it was low. That's really the crux of the matter here, as opposed to moderate.

MEMBER WU: Okay, my response doesn't have anything to do, it's a little related, but can I? Okay. Well from the discussion it sounds like that people thought maybe that it was named incorrectly or that there were other questions that could be asked to get at.

But if feels like it's just because a measure could be better, does it mean it's not good. The conversation didn't sound like it was a bad measure, it's that it could be improved.

Which doesn't mean it can't work as it is. That's just my perception of the conversation that that happened.

CO-CHAIR CORA-BRAMBLE: I would invite the people that voted low on that particular validity score to speak out. It not as helpful for me hear those that were in favor, rather those that rated it low.

All right. Mary, and then Kevin.

MEMBER MARYLAND: I just have a process question. I understand why we're asking the question. But I question the process of having a person, there's a reason that we vote the way we do.

CO-CHAIR CORA-BRAMBLE: Agreed.

MEMBER MARYLAND: And I don't know whether there's another way to get at it anonymously, but this is a process issue.

DR. BURSTIN: And this is a process response, because this is a great questions. I mean we're actually only using these clickers because they're easier and

faster.

I mean before you used the clickers you raised your hands. You knew exactly who voted low. So it's actually not a process issue that you would reveal who you are. You would have just seen your hand up and the old days.

It's just that with the clicker there is a bit of anonymity, nobody should feel forced to say why they voted low, but if some people would like to share that insight I think the committee would value it.

CO-CHAIR CORA-BRAMBLE: But you know what, just going back at one's, our individual and collective experience reviewing grants there is a level of anonymity, number one. And number two it's atypical to have the measure author, the person who wrote the measure sort of actually knowing who voted.

That I can understand, Mary, your concern. And if the group feels more comfortable raising those issues anonymously

in writing I think that's very appropriate. But I would have to, I do understand what Mary is saying.

MEMBER MARYLAND: I understand what you're saying but I'm saying this is just an artificial process, usually the developer is sitting right here as you are raising your hand. So our process is full transparency.

CO-CHAIR CORA-BRAMBLE: That's a whole different ball game.

MEMBER CLARK: So let me I'd like to make a suggestion. I'm not one of the ones that voted low but, think we do have secret voting I'm not sure we should appeal each negative vote after each vote. Because then maybe we should appeal the positive ones as well.

But one thought might be to go through all of these measures and then if we'd like to revisit the ones that we did not approve. That might be an appropriate time to have some discussion around them. But to

appeal every vote, I mean I would wonder why we we're voting.

CO-CHAIR CORA-BRAMBLE: You had a pointer and then you put it down, and we want to hear it.

MEMBER FISCELLA: I voted insufficient, I'm not sure that the criterion was very strong for this in terms of unexpected correlations. But I just thought the data were insufficient that's why I voted down.

CO-CHAIR CORA-BRAMBLE: Okay. Romana.

MEMBER HASNAIN-WYNIA: No, I mean, I think, just to reiterate I think it was the kind of inconsistency and just wanting to make sure that we're all kind of casting our vote with the same knowledge base, in a sense. Of what we're actually casting our vote for.

And that's I think, I don't know how to address that issue. But that inconsistency is.

CO-CHAIR CORA-BRAMBLE: But what I would say is I think each of us has the responsibility to look at the criteria and to vote accordingly. You can't push that any further, I mean we all are responsible and accountable in terms of how we're voting.

And I do agree with the concern of, you know, are we going to revisit it every time a measure is voted down. I do have concerns about that. Because then let's not vote.

DR. BURSTIN: I don't think that was the intent of the discussion.

CO-CHAIR CORA-BRAMBLE: No, I'm reflecting what Luther said.

DR. BURSTIN: Yes, agreed.

MEMBER JACOBS: I think it's somewhat confounded by the fact that it's not like we're appealing the vote. But legitimate questions are raised given that is it the same instrument that we've been discussing all morning, why is this one, I think it's more

along those lines.

Are we being consistent in how we're applying the evidence. I personally, I did actually rate it low. And the reason is why I did is for those issues that we brought up.

In terms of this information may not truly reflect valid information about how the organizations are perceived in the community because you don't know.

There's not a random sampling. You don't know which communities are being reflected. I know it has validity in terms of it somehow.

Actually it's related to these outcomes but that could be that people who also believe or trust those organizations are going to say that they're also good to their communities.

So there are some measurement issues around this. But that's why I voted low, because I wasn't sure that it would have

base validity in terms of actually measuring how this organization was truly interacting or being perceived by the communities it serves.

CO-CHAIR CORA-BRAMBLE: Other comments, Grace.

MEMBER TING: Right, and I was also one of the ones that voted low because I didn't know how valid it would be to apply this particular measurement or cause a lot of different organizations in different regional settings, urban versus rural.

Is it fair to compare it if let's say they're in an urban setting that has a lot of resources and therefore can make those referrals versus where there may be not. I just didn't see that that would be a fair measure.

CO-CHAIR CORA-BRAMBLE: Colette, and then Donna, and then Ellen.

MEMBER EDWARDS: And mine is really more of a process question. Kind of following up on what Ellen had said. If we're

in a situation, and I'm not saying that that was the case with this measure.

But let's say that the reason that people had voted no was because of the way it was named as opposed to if you looked at it with a different name people would have been comfortable with it.

What do you do with something where it seems to have value and how do you put that forward or have it be considered?

CO-CHAIR CORA-BRAMBLE: I think that can be done with a comment. But if we're evaluating it as it stands. So comments, just like the question was raised earlier.

MEMBER EDWARDS: But how do you then get it considered in a different round, is what I'm saying.

CO-CHAIR CORA-BRAMBLE: It would be reconsidered without a problem.

DR. BURSTIN: Okay. So the developer always has an opportunity to provide more information. If they were to reflect on

this and say, you know, the title isn't quite right, they could submit it back to you for your consideration. Before or after the comment period. Again, we're still quite early in this process.

CO-CHAIR CORA-BRAMBLE: The comment I would make in terms of us as a group and in terms of calibration, bear in mind we voted twice on this measure. And twice it was voted down. It wasn't that particular one? My apologies, go ahead.

Who else had a comment? Go ahead.

MEMBER WASHINGTON: So I was one of the ones who voted low, and across the measures I'm moderately concerned by the low response rate by the opportunity for selection bias. And I thought it was just a moderate concern with many of the others.

But with this measure is of particular concern because it really is, when we're talking about community engagement the lack of separation of patient versus staff

responses and the lack of stratification by race group if, you know, for better for worse --

If that's how we're defining community, really questions the validity of what's given. You know, we don't really know what these, who these responses represent. So that's why I voted no on this particular one.

CO-CHAIR CORA-BRAMBLE: I just want to revisit it that the comment that I made I terms of a measure, not necessarily this one.

We as a group in terms of our own internal calibration have looked at a particular measure and have been consistent in terms of our assessment of that given measure.

And we voted that measure down twice, not this one but a measure. So in terms of us as a group, in terms of our assessments of the measures I think there should be a level of collective confidence that if we decide to for whatever reason, no

this one does not meet the bar. Then it is what it is. Period.

Okay. Anything else? Yes, Luther.

MEMBER CLARK: I guess it's a similar comment. I guess there's a little discomfort in not approving this measure. But maybe one of the things that might be helpful to the developers, if the group wants to do that.

If a measure does not pass perhaps we could address the question, is this a measure that's salvageable. I mean, is there something that could be done that we think would make it meet the criteria that would be comfortable with or not.

And provide that feedback so that it could come back. Rather than have the developers necessarily appeal every measure that is voted down.

CO-CHAIR CORA-BRAMBLE: Any other comments? Colette, do you have a comment?

No. Okay. My suggestion is that we do one more before we break for lunch. The next one would be 1892.

MEMBER FITZGERALD: I believe that's me. I hope it's me. If it's not me I studied for the wrong quiz.

So standing between you and lunch is my discussion of this measure so I'll try to be brief but as comprehensive as I can.

So this is a measure of individual engagement domain of the same tool that we've been talking about all morning. So in terms of impact and description of why it has a high impact to the community.

The sources are largely the literature around better effective communications. And service provision and language that's understandable to the patient et cetera.

Not specific to this measure or it's research in particular, but just generally model of effective communication

promoting better health outcomes from a patient population perspective. And many many references made to that.

With regard to it's reference to other measures of the evidence et cetera, again I think it falls upon the idea that this is, you know, measuring what is commonly understood as an important aspect of clinical care.

As it relates to variability again, there is discussion that the scores on this particular measure relative to the outcome measures we've previously discussed is high. It does correlate with each of the three measures we discussed previously.

The overall sense of quality of care provision, confidentiality of medical records. And the notion of good effective honest communication with patients. So that probably addresses the issues around the first element.

With regard to the evidence for

reliability and validity, again, there fairly similar to the previous discussion. I will point out that this measure unlike some of the other measures has a much more robust set of questions for the patient.

I believe there are something 15 questions that are patient specific questions. Many of them in my opinion sort of mirroring what one sees commonly in a CAHPS survey, in terms of overall satisfaction with care.

Availability of the appointments, schedule setting, did the doctor respond to questions, et cetera. So if you want to know the specific ones I can list them.

But they fall into that general language of effective communication between either the doctor or office staff and followup, et cetera.

From the staff survey perspective we get into issues again that overlap with some of the other leadership questions. My senior leadership welcomes a friendlier

environment for patients. Senior staff intervenes when patients feel they haven't been respected.

And then some general overall ratings of the quality of the hospital in terms of treatment towards their patient population, is what the staff survey elements include.

And again, unlike some of the other measures the survey items specific to this one, there are a number of staff questions associated with this as well. Some eight or so questions related to that.

Let's see what I want to touch on besides that. I don't think there's anything else that's very different about this particular measure relative to what we've discussed previously in terms of the reliability or validity of the measure.

The same psychometric testing et cetera being utilized. So unless there are any questions or if I haven't covered

something you're particularly interested I I'll turn it over.

CO-CHAIR CORA-BRAMBLE: Thank you. Any questions from this particular workgroup and then we'll take questions and comments from the group at large. Too close to lunch huh?

All right. Okay, then let's vote.

MS. KHAN: So again in points to measure was the threshold criteria importance to measure and report met? Yes or no, you can start voting now. We have two people missing. We have 18 for yes and one for no.

Moving on to reliability, to what extent was the sub criterion reliability met? You can start your vote now. One for high, 17 moderate, one for low, and zero for insufficient.

Moving on to validity, to what extent was the sub criterion in validity met? You can start your vote now. We have zero for high, 16 moderate, three low and zero

insufficient.

So we can move forward, was the criterion of scientific acceptability of measure properties met? You can vote yes or no. So we have 15 yes, and four no.

Moving on to usability, to what extent was the criterion usability met? You can begin your vote. We have one more. We have one for high, 15 for moderate, one low, two insufficient.

And feasability, to what extent was the criterion feasability met? You can start voting. Zero for high, 17 moderate, one low and one insufficient.

And overall suitability for endorsement, does the measure meet all NQF criteria for endorsement. You can vote yes or no. We have one person missing. We have 14 yes and four no. So it passes.

CO-CHAIR CORA-BRAMBLE: Okay. What we're going to do is we're going to give folks a break now. We're going to break for

lunch. I'm going to ask people to come back at 1:15, rather than making it a total working lunch.

I'll give folks, you can have a break and then let's start a little bit earlier because we have a little catch up to do. So 1:15 sharp, let's be back in the room so we can get started.

All right. Thanks so much.

(Whereupon, the foregoing matter went off the record at 12:44 p.m. and went back on the record at 1:29 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:29 p.m.

CO-CHAIR CORA-BRAMBLE: And no, I did not plan the alarm. No, that was not all part of the big deal. We got to be outside for a few minutes, but it was not my plan.

Let us regroup. We are at Measure 1894, and that discussion will be led by Donna Washington.

MEMBER WASHINGTON: So 1894 is another one of the domains from the CCAT. This is the cross Cultural Communication domain.

In the enumerator statement, they describe that as the component of patient-centered communication, we're an organization to create an environment that's respectful to populations with diverse backgrounds.

This includes helping it's work force understand socio-cultural factors that effect health beliefs and the ability to

interact with the health care system.

And this measure's scored on three items from the patient survey and 16 items from the staff survey.

The distribution of scores with respect to the importance to measure and report in the performance gap was sort of all over the board, and one high, one moderate, one low from the members of the work group that scored this.

And I think part of the distribution is explained by the evidence that they put in. They sort of used boilerplate language and used the same evidence base in all of the nine domain statements rather than making it specific to the domain at hand.

And so cross cultural communication, for example, has a huge depth and breadth of literature supporting it's importance. It's just not reflected here, and so that's probably what we're looking at.

But it is highly important in my

opinion. In terms of the evidence, then, it looks like that's also sort of split between moderate and low. And likely based on the fact that it's just a one multi-site study.

In terms of the scientific acceptability, looks like there's more consensus, reliability and validity.

We addressed most of those issues this morning with the other domains and there really isn't a whole lot more to add about this one in particular other than looking specifically at the results of the testing that they report.

The Cronbach Alpha for the patient items was just in the acceptable range. It was .59 in contrast to high numbers for some of the others. And that's probably related to only three items being in that domain.

It was higher for the provider items. And so there was a range of opinion about whether it meant scientific acceptabilities criteria.

In terms of the usability and the feasibility, really the discussion I would have about this is similar to what we discussed earlier in terms of it not really being correlated with specific actions that healthcare systems can take.

So it's not really clear how patients might interpret the results, how healthcare systems might use the results. And the link to the website just mentions paid consultants as the next step.

CO-CHAIR CORA-BRAMBLE: Concise, sweet, to the point. Inviting feedback from the rest of the workgroup members. Okay, from the committee at large.

Either you're all on a roll or you're asleep. All right. Ms. Kahn, let us vote.

MS. KHAN: Okay, so importance to measure and report was the threshold criteria and importance to measure and report met, press one for yes, two for no. And you can

start voting now.

So we have 19 for yes and zero for no. I'm going on to reliability. To what extent was the sub-criteria and reliability met? You can start voting. We have 15 moderate, four low and zero for high and insufficient.

And going on to validity, to what extent was the sub-criteria and validity met. You can begin your vote.

And we have 13 for moderate, six for low and zero for high and insufficient information. So the measure will go forward.

Voting on overall scientific acceptability, you can start your vote now. And we have 14 yes and five no. Moving on to usability, you can start your vote.

And we're missing one person. Oh, there we go. And we have two for high, 14 moderate, two low and one for insufficient.

And feasibility, you can start your vote. And we have zero for high, 17

moderate, one low and one insufficient.

And lastly, overall suitability for endorsement. We have one person missing. And we have 14 for yes and five for no, so the measure passes.

CO-CHAIR CORA-BRAMBLE: Okay. Measure 1896.

MEMBER CLARK: Thank you. So this is another of the AMA's CCAT tool kit domains in the numerator's statement.

An organization should determine what language assistance is required to communicate effectively with the population it serves, make this assistance easily available, and train it's work force to assess and use language assistance resources.

The score calculation was based on a minimum of 50 staff responses and 100 patient responses.

And as in the others, there were two components to the target population, staff, both clinical and non-clinical and

patients.

In terms of importance of the measure, of course LEP and disparities are national priority goals.

In terms of performance gap and opportunity for improvement, understanding and improving communications is one of the keys to addressing disparities, which is an important national health policy goal.

The body of evidence composed of one multi-site study which involved two phases, and we've heard some of this.

The first phase was for psychometric testing and to refine and simplify the tools.

And in the first round surveys also included our standard items about quality and trust in healthcare which were used to assess the constructs of validity of the two kit domain.

And following the first round of field tests, nine of the 13 organizations

agreed to perform reassessments using the refined tools to assess variability and performance within and between organizations.

So again, in terms of methodology, 100 responses to the patient survey and 50 to the staff survey were required.

And if there were sub-groups, a minimum of 50 surveys from each of these to be compared would be required. And it might, in some cases, necessitate over sampling.

In terms of reliability, the domain of language services showed internal reliability in the excellent to very good range.

The patient survey component which consists of 15 items from the patient survey displayed an internal consistency of .83. And the staff survey component which consists of the 16 items from the staff survey displayed an internal consistency of .96.

In terms of validity, and again, some of this came up earlier, but unlike most

of the other CCAT domains, the language services did not display a strong correlation to patient reported trust and belief in privacy.

In a couple of the examples, is to demonstrate this, the multi variate analysis showed that a five point increase in the measure score result in a ten percent lower odds that the patient would report receiving high quality medical care.

While the same five point increase would result in a slight increase that patients would report a belief that medical records are kept private.

And multi variate analysis also showed that a five point increase in the measure score would result in no measurable change in patient's belief that an error in their care would be hidden by the healthcare organization.

So although the domain of language services was not found to be correlated to the

same indicators of health quality as some other CCAT domains, we know that numerous other studies have demonstrated that improved language services do have a positive effect on quality of care.

And you can see the scores there, there's some mixed numbers and low response rate. So I will pause there for discussion or questions.

CO-CHAIR CORA-BRAMBLE: Okay, I think the group has found it's stride.

MEMBER CLARK: Yes.

CO-CHAIR CORA-BRAMBLE: Any questions, Liz?

MEMBER JACOBS: Yes, Andrew this is for you. I'm guessing you did this, but I just want to make sure. Some of these items should be reverse coded.

Like, for instance, how often have you used a child under the age of 18? You wouldn't want a higher score on that to be good use of out of, you know. I just wanted

to confirm that.

MR. JAGER: Yes, those ones are reverse coded.

MEMBER JACOBS: Okay.

CO-CHAIR CORA-BRAMBLE: Any other questions?

MEMBER O'BRIEN: I don't know why I decided to ask this now instead of a lot earlier. But in terms of how the validation was done, and you're looking at the correlation between the score and then one of the three measures of trust, et cetera.

The endpoint outcome is all measured on an individual respondent, a survey respondent, I guess the patient. What the explanatory variable, was that how a hospital unit or a clinic had measured on the survey?

Or was that how a patient had responded to the survey, meaning that were you showing that, you know, different patients depending on how they report in one part of the survey predicts how they respond to

another question?

Or is it when you aggregate results across multiple respondents from the same hospital and average them to get some overall assessment of hospital performance, is that what predicts how patients will respond?

I mean, I think the latter isn't the more relevant one. You want to know how well this measure, which is ultimately, you know, it's administered at a survey level one at a time.

It's really the aggregate result where you're averaging within a hospital or clinic and it's whether that score can predict patient's responses on other items that they care about.

MR. JAGER: Yes, I don't know if Matt is on the line.

DR. WYNIA: Yes, I'm here.

MEMBER O'BRIEN: Did the question make sense?

CO-CHAIR CORA-BRAMBLE: Matt, do

you want to address that question, or do you want it simplified --

(Simultaneous speakers)

DR. WYNIA: -- I would be happy to. Sorry, this is just one of the challenges of doing stuff over the phone.

If I understood the question correctly, what we were looking at in the validation studies is hospital level performance, not individual performance. In other words, not the correlations within one particular survey.

MEMBER O'BRIEN: Okay, thanks. That answers, it's good.

DR. WYNIA: Is that what you were asking?

MEMBER O'BRIEN: Exactly, thanks. Yes. That's what I was hoping for.

CO-CHAIR CORA-BRAMBLE: Yes, that addresses the question. Thank you. Any other questions from either the workgroup members or the committee at large? Okay, let's prepare

to vote.

MS. KHAN: And importance to measure and report, you can start voting. So we have 19 yeses and zero nos.

Looking at reliability, you can start your vote. So we have one high, 17 moderate, one low and zero insufficient.

And moving on to validity. You can start voting. So we have zero high, 13 moderate, six low and zero insufficient.

So we're going to go forward and vote on scientific acceptability of the measured properties. Okay. So we have 15 yes and four no.

Voting on usability. So we have two high, 13 moderate, three low and one insufficient. And feasability. So we have zero high, 16 moderate, two low and one insufficient. So the measure will pass.

CO-CHAIR CORA-BRAMBLE: Okay.

MS. KHAN: Sorry, we're going on, I'm jumping the gun here. Overall suitability

for endorsement. So does the measure meet all the NQF criteria for endorsement?

(Off the record comments)

MS. KHAN: Now. We have two people missing. One more. Whatever. Okay, so we have 15 yes and four no. So the measure will pass.

CO-CHAIR CORA-BRAMBLE: This next one on Measure number 1898 will be the last of the AMA submitted measures. Health literacy domain of communication, climate, assessment toolkit. Jerry Johnson.

MEMBER JOHNSON: Yes, well I have the pleasure of doing this last one of this measure we've all come to know and love.

CO-CHAIR CORA-BRAMBLE: That's the first time I've heard the word pleasure all day.

MS. KHAN: You know it.

MEMBER JOHNSON: I'll try to maintain that love for as short a period of time as possible. But in any event, this

domain is health and literacy.

And the numerator is stated in a way that annoys me in that it says an organization should consider the health literacy level of its populations and use this information to develop a strategy for clear communication and so forth.

And so you have a two part numerator. And it's just the way they state it. I think instead of just saying that the numerator is the measure on the literacy domain of the CCAT, it doesn't say that. But I think that's what it means.

So this domain is measured by 13 items from the staff survey and 15 items from the patient component. And the same issues related to performance gap and importance that have been described previously today hold here, too.

A lot of citations that are more general than directly focused on health literacy, but there are some.

As a whole, the group of us who reviewed this were split pretty much 50/50 in our thinking about whether they address the important issues sufficiently.

One of the main criticisms against it being important was the lack of stated evidence that they reviewed about an impact on morbidity and disability and mortality.

I think that's too stringent a criticism. It's a whole two, but that was one of the reasons.

As for the evidence base, again, it's the same study, it's the same evidence base as before. This one study, with the kinds of outcomes measures that are trust and quality of care.

The reliability and the validity testing, we've heard about before. So validity of these 13 of 15 questions, either as a composite or as a total score is correlated with those trust items and with the quality items.

And a five point change in the overall score was shown to move in the same direction as a change in quality in almost all of the different 13 organizations where this study was tested.

And that's about it. And then we have usability and feasibility which are exactly the issues, I won't reiterate, that have been discussed before today. I don't think this is any different in that regard. It's the same survey.

CO-CHAIR CORA-BRAMBLE: Okay, Liz?

MEMBER JACOBS: Oh, I'm sorry. That's actually from the last one.

CO-CHAIR CORA-BRAMBLE: Okay. Any comments, questions from either the workgroup or the committee at large? All right, Kevin?

MEMBER FISCELLA: This question is asked in both the adult and pediatric surveys, is that my understanding?

MEMBER JOHNSON: Are these questions asked, I think so. There was an

example, they're pretty much the same.

One thing that I might add, one thing, I like the face validity of the questions for the most part. They actually make sense.

I would say of the 13, maybe 11 or 12 of them actually make sense to me and one doesn't. And most of the others do, too. So maybe even more so than some of the other domains.

CO-CHAIR CORA-BRAMBLE: Okay, any other questions or comments? All right, let's vote for the last AMA measure.

MS. KHAN: Voting on importance to measure and report. We have one person holding out. There we go. So we have 19 yeses and zero nos.

Moving on to reliability. We have one high, 16 moderate, one low, one insufficient. And going on to validity. We have zero high, 15 moderate, three low and one insufficient.

So scientific acceptability of the measure properties. So 15 for yes and four for no. Moving on to usability. Two for high, 15 moderate, one low, one insufficient.

And feasibility. So zero for high, 16 moderate, one low, and two insufficient. And finally overall suitability for endorsement. We have one person. Yes, that's okay. So we have 15 yes and three for no. So the measure will pass.

CO-CHAIR CORA-BRAMBLE: Okay, the next two measures are from the Agency for Healthcare Research and Quality.

DR. WYNIA: Madam Chair, would it be okay, I'm going to get off the line.

CO-CHAIR CORA-BRAMBLE: Oh, okay. Thank you.

DR. WYNIA: Yes, if I may, I would just like to really thank the committee for what I know was a lot of time and energy and deep thought put into looking at a set of measures that is not always easy to fit into

the traditional performance measurement framework.

We sent you a challenge, and I really appreciate the effort that you've put into looking at these.

And I hope that Marshall Chin's earlier comment about maybe sending us some feedback on some of the domains that didn't pass and how we can make them stronger in the future, we would really appreciate that kind of feedback.

CO-CHAIR CORA-BRAMBLE: Okay, sounds good. Thank you so much, though, for your help.

DR. WYNIA: Thank you.

CO-CHAIR CORA-BRAMBLE: All right.

MS. BRACH: All right, I'm just going to give you a very quick overview of both this, the health literacy measures and the cultural competence measures together.

They are developed based on item sets that are supplements to an already NQF

endorsed measure, the Clinician and Groups CAHPS.

They were developed separately, separate testing. But also there was some overlap in the areas where they coordinated.

The CAHPS development process is a very rigorous one. We first look and see what else is out there in the field in the area. We publish a call for measures in the Federal Register.

We convene stakeholders to tell us what domains they think are important. We do cognitive testing in both English and Spanish, and field testing, which we did with a mailed survey followed by a telephone follow up.

And just for those of you who aren't familiar with cognitive testing, cognitive testing is what gives us a lot of confidence that these measures are actually measuring what we think they are measuring.

And let me introduce Bev Weidmer who is our survey director. And Ron, are you

on the phone? We were supposed to have joining us Ron Hays, our psychometrician. But --

DR. HAYS: Yes, I'm on.

MS: BRACH: Oh, terrific. So Ron, all the hard questions go to Ron. But the cognitive testing, you know, you ask the patients what they think they're being asked and why they gave the answers.

And that identifies where there are problems with our items. And then we refine them and retest them.

There are a large number of measures. We developed two composites based on those. But these are all independent. As supplemental items, we don't have any expectation that anyone will adopt all 27 items for the health literacy measures.

There are all 30 items that you can pick and choose. You could do one of the composites, you could do a set of them that makes sense for your organization.

And they can be reported both at the individual clinician level, aggregated at a group or a clinic practice level. And that's true the composites as well.

And I just wanted to take the last minute to bring this back to disparities, which is why this call for measure went out and this panel's been convened.

What you're looking at is the disparities in health literacy as measured by the National Adult's Assessment of Literacy.

And you can see that on the lowest, the below basic and the basic categories, that we had much higher proportion of Black and Hispanic Americans suffering from limited health literacy than White Americans.

And similarly, this is from our National Healthcare Disparities report thanks to Ernie Moy, we have shown that there are disparities in reported communication measures from our CAHPS core items.

So that, you know, just in case

you were wondering why we were looking at these in a disparities call, it really gets right to the heart of some of the disparities that we see.

So I will just leave you with a quote from an article that was published this month in Health Literacy that Assistant Secretary Koh led, that really both of these health literacy and cultural competence are very important in addressing health disparities.

So hopefully we are in the right place for that.

CO-CHAIR CORA-BRAMBLE: Thank you, great introduction, Cindy. Thanks so much. All right, we're going to move on to Measure 1902. So Mary Maryland.

(Off microphone discussion)

MEMBER MARYLAND: All right, got it now.

CO-CHAIR CORA-BRAMBLE: I'm sorry, Mary?

MEMBER MARYLAND: Sorry?

CO-CHAIR CORA-BRAMBLE: Before you continue.

MEMBER MARYLAND: Oh, yes. You want to pass out that correction.

CO-CHAIR CORA-BRAMBLE: We had a little late night cut and paste error. So in case you're wondering why on the health literacy measure the one --

MEMBER MARYLAND: Yes, I got it, thanks.

CO-CHAIR CORA-BRAMBLE: -- slurry of evidence is all about cultural competence, it's because we accidently --

MEMBER MARYLAND: Cut and paste in the wrong place. While she gives that out, let me just tell you a little bit about what this measure is.

So, first CAHPS, it's actually Community Assessment of Healthcare Providers and Systems. So as we talk about it, recognize that it's from the consumer's point

of view and it's talking about both providers as well as the systems in which they get care.

We are specifically looking at five items in terms of health literacy and three items in relationship to medication administration.

The five questions in relationship to health literacy are specific in terms of what the emphasis is and what they're asking folks to look at.

And in medication administration, it's also talking about medication safety. So did the provider tell you about how to be compliant in taking your medication?

Did they tell you in a language that was easily understandable to you? And did they tell you about the side effects of the medication?

So the medication ones had the better specificity in terms of what they wanted you to do. So it's basically around how well did the provider communicate with

you?

And with that, I'll just go through the list, one other thing. So this measure was evaluated in two facilities. One in New York in the Bronx. And the one in the Bronx was a Medicaid health plan.

And the one at the University of Mississippi was an outpatient medical center. So both outpatient facilities.

And just by way of definition, Healthy People 2010 defined health literacy as the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions.

So that's the frame of how this came around. And, again, I iterated that it was from the consumer's point of view.

The comment was made that the Federal Register was used to solicit comments for this. And it's not unusual, but that call did not reveal anything, no new measures.

And there's typically a low response, even though that's part of what's typically done to get additional information.

So looking at the responses very specifically, and in my group for whatever reason there were six of us, but only two of us weighed in. So it will be short and sweet.

So importance of the measure to report, 50/50, one yes, one no. That makes it really simple. In terms of impact, guess what, one yes, one no. Not yes/no, but one high and one low.

In terms of looking at the performance gap, it was 50/50, but it was one high and one medium. And looking at the evidence, there it was 50/50, one yes, one no.

Health outcomes, six of us agreed that this was not a health outcome, so I guess that was good. In terms of quantity, two of us 50/50. The one high and one low.

Quality, one medium and one low. And consistency, one medium and one unable to

determine, inconclusive.

So part of the issue is that there was only the one study at two sites. So that limited the usability and the relationship of health outcomes was not described.

So thinking about the scientific acceptability of the measures, one yes, one no of the two people. In terms of reliability, one high, one medium. Validity, one medium, one low.

And the specific issues were in relationship to psychometric properties, which I'll tell you. And the reliability was tested in two populations, neither rural is one of the comments.

And a biased selection sample toward low English proficiency individuals. Both the facility in the Bronx for the health plan, as well as the University of Mississippi Medical Center were both fairly low English proficiencies. So that was the other comment.

In terms of feasibility, 50/50,

one high, one medium. It said that we would need some additional surveys to be able to identify. And the sampling strategy was well reasoned, but the assumption is that it holds true for the entire group.

And the five questions on health literacy were a subset of a larger item. And I'll tell you about the reliability scores for those in just a second.

So in terms of the five items, the subset, and that subset came from an original set of items that was 17. And it was decreased because it was felt that the 17 were too long.

So those five items, just in case you're wondering were they the right five items, those five items accounted for 90 percent of the variants. So there was fairly good comfort that this one was the appropriate set of five items.

And the reliability estimate for those five items was .79. We also had, on the

set of three items for the medication administration, I think .84, if I remember correctly.

So the other thing I think is important to know is that this is currently being utilized by MEPS and my acronyms, I had to look up what that was since I had no clue.

But it will definitely improve the data set because it is the group that is responsible for large numbers of health insurance plans, and it is, oh I lost it.

(Off microphone discussion)

MEMBER MARYLAND: Yes, somebody have the information about what MEPS is?

MS. BRACH: It's the Medical Expenditure Panel Survey, which is a nationally representative household survey that is fielded by AHRQ every year.

And several of these measures from this item set were included in the 2011 fielding of MEPS and will be included two more times between 2020 to produce measures for

Healthy People 2020 health literacy objectives.

MEMBER MARYLAND: The other thing about MEPS is that it also includes cost data, which in this environment we were really interested in that. So that's the brief summary. And my other teammates, anything?

CO-CHAIR CORA-BRAMBLE: Actually, I have a question. The issue that you raise regarding the English proficiency, you were raising it as a confounder in terms of literacy? Is that --

MEMBER MARYLAND: In my opinion --

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER MARYLAND: -- it is a confounder because we don't know outside of that limited English proficiency, how the measures would have performed.

CO-CHAIR CORA-BRAMBLE: Any questions from the group, either the work group or the committee at large? Liz, and then Marshall.

MEMBER JACOBS: My question is regarding what is this global physician rating scale, and why is it that you would think that it would show validity if it predicted actually global physicians?

MS. BRACH: I'm sorry, Liz. Can you get a little closer?

MEMBER JACOBS: What is the global physician rating scale you used? What's on that, and why did you think that that would actually validate this measure?

MS. BRACH: Right. This is a core item from the clinician group's cultural competence.

It asks the patient how they would rate their provider overall on a scale of one to ten.

So what we were trying to do there is seeing to what extent were these items that measure the health literacy practices of the clinician and the group seem to be related to the patient's overall assessment of the

provider.

MS. BRACH: It's that one item question?

MEMBER JACOBS: Yes.

MS. BRACH: Okay, I know what you're talking about. Okay, thank you. Although I didn't want a question about that.

CO-CHAIR CORA-BRAMBLE: Mary, your response to that?

MEMBER MARYLAND: It's not a response to that. It's actually just an additional piece of information.

All the other items in this scale were Likert, always, never, in that manner as opposed to this being zero to ten rating your provider.

CO-CHAIR CORA-BRAMBLE: Marshall?

MEMBER CHIN: Just a point of information. Can you read the actual question from the scale and the question in that validation, global question and then just to repeat the liability and validity data?

MEMBER MARYLAND: Sure. So the first one is, I want to say question nine, it is.

And the question says, "In the last 12 months, how often did this provider give you all the information you wanted about your health?" Likert, never, sometimes, usually, always.

The next question, "In the last 12 months, how often did this provider encourage you to talk about all of your health questions or concerns?" Same Likert.

Question 14, "In the last 12 months, how often did this provider ask you to describe how you were going to follow these instructions?"

And this is referring to medication. No, this is referring to instructions about how to manage that health problem. And same Likert.

The next one is 20, "In the last 12 months, how often were these instructions

about how to take medications easy to understand?"

And there should be one more. Twenty-eight?

Twenty-six? Did I skip one, sorry. Yes. "In the last 12 months, how often were the results of your blood test, x-ray or other test easy to understand?"

So the one above it says do they give you that information, and this is asking did you understand the information.

And then I think the last is 28? "In the last 12 months, how often did someone explain the purpose of a form before you signed it?" And the question above it is did you sign a form in your office?

MS. BRACH: That's not part of that scale, that's a separate item. So the first five that you listed are on the scale.

MEMBER MARYLAND: Okay.

MS. BRACH: So the scale ends with the blood test one.

MEMBER MARYLAND: Okay.

MS. BRACH: So that the rating --

CO-CHAIR CORA-BRAMBLE: Ellen?

MEMBER CHIN: Yes, then the reliability, yes, validity data.

CO-CHAIR CORA-BRAMBLE: Anything else, Marshall? Mr. Win for next.

MEMBER CHIN: Yes.

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER CHIN: Was the validation question. And then what is the reliability and validation data? Thank you.

(Off microphone discussion)

MS. BRACH: Okay, so the data comes from this field test that Mary eluded to.

So for the five item composite that she just spoke to, we did an internal consistency measurement. And it came out to .79.

Is that okay? And then did you want the correlations? We did a regression on

the global rating, and the alpha was a 6.77 at a .001 key value. You're looking --

CO-CHAIR CORA-BRAMBLE: I'm sorry. I can't read your face, I don't know if that means yes, no? I don't know what it means.

MEMBER CHIN: I can't interpret that, those last numbers. I didn't understand that.

MS. BRACH: At the 6.77 is the regression coefficient so that that's the --

MEMBER JACOBS: So just for clarification, you're saying that a higher measure on the health literacy measure was significantly related to a higher score on the global physician rating?

MS. BRACH: Exactly.

MEMBER JACOBS: I think that's your question, right?

MEMBER CHIN: Yes.

MS. BRACH: It had a very high confidence level.

DR. HAYS: Yes, and you know, just

to clarify, let me see, we've got an echo. I heard that on the previous caller.

If you look in the document, they have correlations of each item with the global rating and those range between .42 and .61.

MS. BRACH: Right, that's each of the items separately, not the composite.

DR. HAYS: Right.

CO-CHAIR CORA-BRAMBLE: So Liz and Marshall, does that address your questions, both of you? I see nods now, we're good. All right, Ellen?

MEMBER WU: So this isn't specifically about this measure, but a general comment which I had, I think, at the first meeting brought up that there's a concern that CAHPS is actually only implemented in English and Spanish.

So we're losing feedback from a lot of populations. And hopefully this committee, our efforts and QF's work can really work with NCQA in making sure that the

other translated versions of CAHPS are used out in the field.

CO-CHAIR CORA-BRAMBLE: Good point. Cindy, do you have a response to that or any comment?

MS. BRACH: No, I mean it's something that we struggle with. Some items, for example from our hospital CAHPS and some of these items have been taken up and adapted for hospital which we're going to be publishing shortly.

But the issue is really because we do such a rigorous job in psychometric testing that these items are actually co-created in English and Spanish.

So when we develop it and we're making changes to an item, we think about what is that going to mean for the Spanish translation? And sometimes it makes things, you know, very difficult in Spanish.

And so we have to adjust it so that they're sort of we're metering them

against each other. And just to do that in additional languages, you know, triples, quadruples, et cetera, the expense and effort in producing the measures.

CO-CHAIR CORA-BRAMBLE: Okay.

DR. HAYS: But, there are examples in, for example, California where we've translated into Asian languages and other languages depending on the application where it's needed. That's always a possibility and has been done.

CO-CHAIR CORA-BRAMBLE: All right, I have Kevin, yes, no, you? And then Mary, yes? Oh, Romana, yes?

MEMBER FISCELLA: Just a comment and a question. I guess the comment is, I think, that this is probably going to be, at least for now, state of the art measurement of these key constructs, so I'm very enthusiastic about them.

My question is, is there or will there be a national normative data for these

measures as they are for the cores?

MS. BRACH: Yes, you're talking about the National CAHPS bench marking database?

MEMBER FISCELLA: Right.

MS. BRACH: Unfortunately not, because these are supplemental measures. And so we have not been able to get enough folks who are fielding the same supplemental measures to constitute a reliable database for that.

So right now the MEPS measures are going to be the only ones that will really have national bench marking data for the items that we've incorporated into MEPS.

MEMBER FISCELLA: Is global incorporated into the MEPS data?

MS. BRACH: Only three items.

MEMBER FISCELLA: Only three items?

MS. BRACH: Not the whole item set. But one other potential source of future

data is that we're about to field test a health plan version of this that also includes these items.

And health plans are much more likely to, rather than at the clinician and groups level, to have more data that could be compiled to produce that kind of measure. You know, so I'm hoping in the future.

CO-CHAIR CORA-BRAMBLE: Mary, and then Romana.

MEMBER MARYLAND: So this information just speaks to the diversity of language. And this is from our last census in 2010.

And so it says, "Of the 281 million people in the United States 5 and older, 55.4 million, or 24 percent report speaking a language other than English at home." So that's one in five.

After English and Spanish, which Spanish is 34.5 million speakers, the next most prevalent languages are Chinese at 2.5

million, followed by Tagalog at 1.5 million, one of the Filipino dialects.

French 1.4 million, Vietnamese 1.2 million, and German, 1.1 million, and Korean 1.1 million. And the largest group in terms of age of all of these folks is 78.3 million were between 41 and 64, but there are 32.6 million speakers 65 and older.

CO-CHAIR CORA-BRAMBLE: Okay, thank you Mary. Romana?

MEMBER HASNAIN-WYNIA: I just wanted a point of clarification based, Mary, on your summary. So was this only tested in low-income Medicaid, LEP? Both in the Bronx and at the University of Mississippi?

MS. BRACH: No, I'm sorry.

MEMBER HASNAIN-WYNIA: Okay.

MS. BRACH: That was a misstatement. The Mississippi actually was 100 percent in English. The respondents in the Bronx, about 42 percent of them, I believe, responded in Spanish and the rest in

English.

And those included multi-lingual groups who, you know, either got assistance in filling it out in English or were able to fill it out in English. Does that answer, yes.

(Off microphone discussion)

MS. BRACH: Yes, it was.

CO-CHAIR CORA-BRAMBLE: Okay, the folks that are finished speaking, just put your name tags down so that I'll know that you're finished. Kevin? Liz?

MEMBER JACOBS: That raises another question, Cindy, which is that did you find differences between the two sites in the performance of the measure, given that they're very different populations?

MS. BRACH: We did have similar response rates in both. But I'm not sure, did we compare?

(Off microphone discussion)

MS. BRACH: Yes, I understand.

MS. WEIDMER: Yes, and we did

compare. We didn't have enough data to really adequately examine.

I mean, because basically the study was powered to overall have sufficient power to be able to measure, but we only had half as many at each place.

CO-CHAIR CORA-BRAMBLE: Mary?

MEMBER MARYLAND: So those original, I think our correct sample size was targeted to be 1,200. They did 601 was the total. And so the response rate was about --

MS. BRACH: Yes.

CO-CHAIR CORA-BRAMBLE: Kevin?

MEMBER FISCELLA: What was that mean level? I'm sorry, I missed that. What mean level is that?

MS. BRACH: You're right, you didn't see it.

MS. WEIDMER: It's consistent with the CAHPS which are, we aim for a sixth grade reading level. I should clarify, we don't rely on, you know, the Flesch-Kincaid or other

word measures of reading level because they're not very accurate.

But in the cognitive testing, about two-thirds of the respondents that participated in cognitive testing, both in English and in Spanish had high school or less. And over half had less than an eighth grade education.

So we really, really aim to task the measures with patients with very low levels of education, very low literate.

MS. BRACH: Right, and in the cognitive testing, we sort of simulated the mail administration of the survey by having them read the questions themselves and fill it out, but think out loud so that we could understand.

And then we probed them afterwards. But we did half like that and then half sort of simulating the telephone where we read the questions to them.

MEMBER JACOBS: You couldn't

capture, then, a group of people who really couldn't read well at all? Right? Or did you, by these telephone?

And if you didn't, how do you think that, I mean, it impacts the utility of this measure if it's about health literacy but then people have to read it to fill it out.

MS. BRACH: We did what we call a mixed mode administration. So anybody who did not complete the mail survey after several attempts was called multiple times at different times of day to try and get them to fill it out over the telephone.

MEMBER JACOBS: Is that how CAHPS works now? So I know a lot of hospitals and health organizations use CAHPS. So will administration have to change to be able to do that?

MS. BRACH: CAHPS right now, supports three kinds of administration. One is mail only, one is telephone only, and one is the mixed mode.

And so it could be that a practice that was using this and did it only by mail would miss out on people who, you know, anything written just automatically goes into the trash.

I mean, one thing which we do try and capture on CAHPS is asking a question whether or not people had any help in filling out the survey and what kind of help did they receive.

And we find that, I believe, and correct me if I'm wrong, that a small proportion of people fill it out with help.

So, you know, it's not perfect and I would certainly recommend anybody, you know, to use the mixed mode administration. But it's more expensive and some organizations clearly are not going to find that feasible.

MEMBER JACOBS: Okay, thank you.

CO-CHAIR CORA-BRAMBLE: All right, thank you so much, Cindy. Any other questions from the group? Okay, let's get ready to

vote.

MEMBER YOUDELMAN: Can I just mention one thing for voting since Dennis and I are now unmuted that we may need to change the total number so that they know if everyone's voted?

CO-CHAIR CORA-BRAMBLE: Good point. Twenty-one?

MS. KHAN: All right, 21.

CO-CHAIR CORA-BRAMBLE: Twenty?

MS. KHAN: Okay.

CO-CHAIR CORA-BRAMBLE: Any other conflicts, was that it? Our add is 20. Okay.

MS. KHAN: Importance to measure and report. We have two people missing, so if you guys could enter it one more time. There we go. So we have 20 yeses and zero nos.

Going on to reliability. So we have seven high, 13 moderate, zero low, zero insufficient. And going on to validity. We have five high, 14 moderate, one low and zero insufficient.

And overall scientific acceptability of the measured properties? I have 18 yes and two no. And going on to usability. We have six for high, 14 moderate, zero low, zero insufficient.

And feasability? We have three high, 17 moderate, zero for low and zero insufficient. And overall suitability for endorsement. We have 20 yeses and zero nos. So the measure passes.

CO-CHAIR CORA-BRAMBLE: Okay, excellent. The second AHRQ measure would be number 1904, and Norman Otsuka? Yes.

MEMBER JACOBS: Just, I want to say I have a conflict, so I'm not going to participate in the discussion part of the, just prior to it happening.

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER OTSUKA: All right, great. Thank you for the opportunity to review this cultural competence item set. Editorial, I liked it.

As a clinician, I thought it drilled down to the real issue of confidence and trust.

And although it's not a health outcome, I think in some respects, it is related to health outcome and adherence and how patients respond or interact with the physician.

The review of the literature is quite compelling. And there is differences in trust and confidence based on race in a physician/patient relationship.

One of the comments from my colleagues was that the citations didn't represent the full body of evidence.

But nonetheless, what was presented was pretty, I mean, thorough and I think the disparities that were seen were quite compelling.

With that being said, it was tested, it was tested in two large samples in New York and California.

One of my colleagues suggested that there may be some bias in that, but I can't think of two more diverse populations than New York and Los Angeles to test this in.

So I thought that it was adequately tested. Another concern there was that the questions did not measure cultural competence, but it measured elements of culture, bias, prejudice and language competency.

And we can discuss that if you wish. But nonetheless, my colleague still rated it a moderate rather than the high.

Usability, I mean, I didn't quite understand my colleague's comment. But, you know, I thought it was very thoughtful questions and items that could easily be answered without any issue.

Feasability, when my colleague suggested that most of the elements of the questions aren't gathered on electronic health records, but I mean, I don't think that's an

issue for this item set.

Frankly, I thought it was well thought out and the literature as well as their testing bear out that there are disparities with confidence, trust and communication based on race in a physician/patient relationship.

And my editorial is that I liked it and, well let's open to discussion. Thank you.

CO-CHAIR CORA-BRAMBLE: Okay. Questions for Norm or for Cindy? Marshall?

MEMBER CHIN: So I have the same question, but the rest of the committee didn't have access to this particular information.

So if you could state what the actual questions were, Norm. And then the reliability and validity data in the correlation question.

MS. BRACH: They are in the numerator's specification.

MEMBER OTSUKA: Yes, they're in

the back.

MS. BRACH: You don't have that?

MEMBER CHIN: No, we don't have that. I think, except for the people who were sent it, the rest of us don't have it.

MEMBER JACOBS: Is it on that thumb drive?

MEMBER CHIN: It's on the thumb drive and --

(Off microphone discussion)

MEMBER OTSUKA: I can read it all, but --

(Off microphone discussion)

MEMBER TING: Yes, so it's a little long, it's a lot. But, you know, a lot of it is prefaced by in the last two month, did you feel that you could tell this provider anything, even things you might not tell anyone else?

Do you feel that you could trust this provider with your medical care? Do you feel that providers always told you the truth

about your health, even if it were bad news?

Do they care about your health as much as you do? Do they care about you as a person? Do they talk too fast, they use a condescending, sarcastic or rude tone or manner?

Do they interrupt you when you are talking? And there were some questions regarding do they ask you about use of complementary medicine? You know, acupuncturists, herbalists, so on, so forth. Things of that nature.

CO-CHAIR CORA-BRAMBLE: Jerry?

MEMBER JOHNSON: Yes, I like these questions. I think the range of these questions cover the domains that we find when we read about cultural competence.

With the one exception of no questions that I can discern that ask anything about causation of illness or the patient's view of why he or she is sick.

And some of that literature, that

explanatory model question is considered to be really important. But was it in there and fell out because it tested out, or was it just never in there?

And then the other domain that I'm not quite sure that I would have considered for a cultural competence kind of a survey would be some questions having to do with help seeking behavior in the extent in which providers understand the kind of help seeking behavior that one group uses versus another.

So I don't see those two, which when I think about a list of domains that would make up cultural competence, they would include those two.

MS. BRACH: The second one, though I might argue with you the patient is not going to be the best source of information on that. So, you know, we were certainly focusing on the patient experience and what they could report back.

MEMBER JOHNSON: I don't

understand that. Why the patient would not have a sense of how he or she seeks care.

MS. BRACH: How the doctor -- oh, I though you were saying how the --

CO-CHAIR CORA-BRAMBLE: It's how they're seeking care.

MS. BRACH: -- doctor understands how I seek care.

(Off microphone discussion)

MS. BRACH: So you would be asking the patient does your doctor understand how you seek care --

MEMBER JOHNSON: Exactly, yes.

MS. BRACH: -- in some way. That's hard for the patient, I think, to assess whether the doctor understands or not.

MEMBER JOHNSON: We have a different view on that one.

MS. BRACH: Okay.

MEMBER JOHNSON: Yes.

CO-CHAIR CORA-BRAMBLE: But why not ask the patient directly in terms of their

own health seeking behavior? Isn't that what you're eluding to, Jerry?

MEMBER JOHNSON: Yes. I think if the relationship is a meaningful --

MS. BRACH: You have to turn on your mic.

MEMBER JACOBS: No, I'm saying if the relationship is an effective relationship with effective communication, the patient should have a sense that this doctor or this nurse actually kind of understands my network of help seeking behavior, that's all.

MS. BRACH: Right. Well, that is something that I don't think the cultural competence team did even seek to develop items about. I can imagine that it might be quite difficult to get to a cognitive testing.

MEMBER JOHNSON: What about the causation issue?

MS. BRACH: You mean health beliefs, I would -- asking about what my health beliefs are, why, you know --

CO-CHAIR CORA-BRAMBLE: What do you think caused the illness?

MS. BRACH: -- the hind-end questions and stuff. And Bev was on the cultural competence team. So I'm going to --

MS. WEIDMER: You know, we did a fairly extensive literature review leading up with trying to identify what domains were the domains to prioritize for inclusion in the item set.

You know, like any project, we're limited in what we can include. It was already a fairly extensive item set as it was.

And that was not one that kind of surfaced to the top in terms of what should be prioritized either in the literature or from expert input and from stakeholder input.

That was not one of the domains or topics that we felt and they felt should be prioritized for inclusion in the item set. So that's not a very satisfying answer, but that's essentially why we didn't include it.

CO-CHAIR CORA-BRAMBLE: So the feedback, at least, from some of us in the group is that it would certainly be, I don't know who the experts were, but those two questions that he raised are very key as it relates to, you know, measuring, if you will, cultural competence.

All right, around the table. Lourdes and Mara all the way around. Yes? I'm sorry, Romana and then Mara.

MEMBER HASNAIN-WYNIA: Maybe we touched on this already, but I didn't get a chance to review this measure. I couldn't access it for some reason.

So what I'm struggling with when I'm looking at the items on here is how are these cultural competency measures? So for example, "In the last 12 months, how often did this provider use medical words you did not understand?"

Or, "In the last 12 months, how often did this provider show interest in your

questions and concerns?" Unless you stratify them by language or --

MS. BRACH: Exactly.

MEMBER HASNAIN-WYNIA: -- how would these --

MS. BRACH: No, you're absolutely right. And in fact, they're were even other measures on help promotion which we've since booted out of the set for that reason.

We did keep those around communication and those are overlap with the health literacy items. Those came from the health literacy item set.

But because we know that there really are disparities in those reports that we felt for people who are going to just look for cultural competence measures and are going to look to this item set, that it was important to include them there.

But you're absolutely right, that for those to really be measures of cultural competence, you would need to stratify them by

race or ethnicity, which we did in our analysis.

(Off microphone discussion)

CO-CHAIR CORA-BRAMBLE: -- about race or about stratifying by race and ethnicity or were you questioning the actual question as to whether they were measuring cultural competence? What were you doing?

MEMBER HASNAIN-WYNIA: Right, I was questioning. I mean, I understand the stratification piece, because that would be the next piece of making these akin to cultural competency questions, or having the ability to look at them through that lense, I guess, the cultural competency lense.

What I was struggling with is, when I read these questions, these did not come across to me as cultural competency questions. So in some ways, I guess, you know, I'm struggling with how these questions will be perceived if this measure passes.

And we label them as cultural

competency questions. So for the end users, there's a little bit of a disconnect. I mean, there's the disconnect for me. So I guess, you know, are these cultural competency questions?

CO-CHAIR ANDRULIS: Well, it comes back to Jerry's point about are they targeted to that. It's almost, in some ways, more patient-centeredness rather than cultural competence.

CO-CHAIR CORA-BRAMBLE: Yes. So it's the same issue we addressed with one of the other measures where is the title right? Does it capture what's in the body of it?

CO-CHAIR CORA-BRAMBLE: I was going to let Mara speak, but it's okay.

MEMBER JOHNSON: Okay. I better be quiet.

CO-CHAIR CORA-BRAMBLE: All right, you have the floor, Jerry. Go ahead, go ahead.

MEMBER JOHNSON: No, I was going

to say this is a tough one. What really is cultural competency, what's a question that is in that general domain and what's not?

For the most part, I like these questions and I thought they were. I mean, how are you going to ask about, for example, of the two examples that you gave, I thought the last one was trying to get at whether or not the provider was respectful.

Or whether or not the perception of the patient was that the provider was respectful. And I would view that as in communicating as one example. And then, of course, there's the complimentary alternative medicine question.

So it looks like they just went through a list of domains and says do you perceive that the provider is taking actions in these domains? That works for me as long as the domains are relevant.

MEMBER HASNAIN-WYNIA: I agree with you. I think the questions are fine.

But I think you said it, that these are questions about patient-centered communication or communication quality, quality of communication almost.

It kind of takes me back to some of the measures that, you know, some of the instruments that Debra Roter and Mary Catherine Beach around the quality of communication.

Is it more, you know, provider dominated versus patient? Is the patient asking? And so to me, those are more related to patient-centered communication.

MS. BRACH: It's only when you get to stratification --

MEMBER HASNAIN-WYNIA: Exactly, exactly. I think the questions are fine.

CO-CHAIR CORA-BRAMBLE: Mara, and then Donna.

MEMBER YOUDELMAN: And I think a lot of it is sort of the first, you know, fifteen, well it's going to keep going, I

guess, like 20 questions which really are more about patient-centered care, which is important to know.

But a provider can be respectful without being culturally competent, necessarily. I mean they could, you know, take you on time. They could, you know, answer some of the questions, but cultural issues might not have come into play.

And so, I guess, that's my concern here is that it doesn't sort of get to the, you know, were your cultural beliefs identified, discussed, addressed?

You know, how that impacts sort of treatment and care, because that's really getting to the meat of the issue rather than did they use a, you know, condescending tone, to me.

I mean, I think the second half of the questions, which you get into the interpreter and the language services certainly is more related to cultural

competency.

And it's almost like there's two pieces of this. Don't smile at me, Cindy. I mean it's almost like there's an interpreter competency subset, and then the rest kind of came in from the health literacy to make it.

MS. BRACH: Well, there is a language access subset. But there is also a discrimination, you know, questions. There are also trust questions.

So I mean, I think that Romana has made, to me, anyway, it resonates the most to me, that some of the communication items from the health literacy item set that were brought over to here because they felt that providers need to get this right with all diverse populations.

And if they don't, that's a problem. But those, to me, have the less cohesion with what we think of as cultural competence.

MEMBER YOUDELMAN: Right.

MS. BRACH: But I would say, you know, trust and discrimination and these other domains in addition to language access are also very much squarely in the realm of cultural competence.

MEMBER YOUDELMAN: Do you --

CO-CHAIR CORA-BRAMBLE: Okay, let me get the rest of the comments around the table. Somebody was speaking?

MEMBER YOUDELMAN: Yes, I just wanted to ask one more question. Do you expect or has experience been that if a provider does literacy, they also do cultural competency? Or is it really they take either or?

MS. BRACH: Are you talking about the item set?

CO-CHAIR CORA-BRAMBLE: When you say when the provider does, wait a minute. Hold on just a second, Cindy.

MS. BRACH: Sorry.

CO-CHAIR CORA-BRAMBLE: What do

you mean by when the provider does literacy?

MEMBER YOUDELMAN: My understanding is that they're sort of optional subsets of CAHPS.

So when the office or the provider or whoever is deciding to do CAHPS, are they picking we're going to add on the literacy piece, we're going to add on the cultural competencies, we're going to add on both.

CO-CHAIR CORA-BRAMBLE: I see.

MEMBER YOUDELMAN: Like, I'm wondering if they're sort of being seen almost, even though they're two subsets, are they really being taken as one?

MS. BRACH: Right. This is not an evidence based answer. It is sort of an informed speculation. There are two things.

One is they're competing measures because people are worried about the length of the survey. We certainly can't do all of even one of these item sets every, you know, time. So that to some extent, people are going to.

I think that, in general, people have in mind a certain quality improvement area. I'm going to work on health literacy. I'm going to work on disparities and cultural competence.

And they will look around for measures in those areas. So I don't think that somebody who's focused on disparity quality improvement is necessarily going to go through the health literacy item set and say oh, what looks good here? But as I say.

CO-CHAIR CORA-BRAMBLE: Okay, let me take the rest of the comments around the table. Donna, you had yours up and you put it down, because then I have to go around the other side.

MEMBER WASHINGTON: No, you addressed most of my points. But I do have one other question. So it looks like it's a larger data set that some of the, multi set of questions, there are two composites within it.

And some of the items that aren't

included in the composite more specifically address cultural competency. And so that's my quick read of this measure now.

And so the question is can some of those items that specifically address more of the issues directly related to cultural competency be pulled out? Did you conduct testing on those to see if they stand alone as a scale, for example?

MS. BRACH: Right. We actually tested seven different domains. The language access is not included as a composite measure because some of the items that constituted the composite had been removed.

So that what we did testing on for the composite isn't the end that we have here. So we did not put that one forward.

The other ones didn't hold up. So we had one that was an equity, you know, a discrimination one with those two items. And, you know, when we did the reliability testing, it didn't pass muster.

So we've only put forth those composite measures that, you know, had a scientific evidence base to stand behind. I would have liked to be able to offer those.

CO-CHAIR CORA-BRAMBLE: Okay, Kevin?

MEMBER FISCELLA: Yes, at the risk of belaboring a point, I completely agree with Jerry here that I think failing to ask the patient whether the provider inquired about their culturally specific beliefs, explanatory models, practices, health care use is really fundamental to cultural competency.

This is really the essence of, I think, what it means, particularly if you're not a member of that group. But really for everybody.

So I think in the future, I would really encourage some work on this. I think it's quite doable. I think you can develop items around it.

And, you know, I like the other

items. I agree there's lots of overlap between patient-centered care and cultural competency, and that can probably be teased out a little more. But I think there's really need for further work here.

MS. BRACH: And I really appreciate these comments because I really agree with you. I mean, I was not involved in the initial part of the development of this item set.

And I'm a stalwart believer of the climbing questions and how important that is. And we do have opportunities to do, you know, further testing and adding and stuff. You know, so we will definitely pursue that.

CO-CHAIR CORA-BRAMBLE: Okay, so Mary, Lourdes, and then Grace?

MEMBER MARYLAND: Thank's so much.

CO-CHAIR CORA-BRAMBLE: And then we're going to vote. Go ahead.

MEMBER MARYLAND: My comment is perhaps a way to frame the question that just

came up. Are we looking at areas of diversity versus areas of disparity rather than areas of cultural competence?

And so that might explain why we have some questions that are looking at where there has been known disparity in outcomes.

But in terms of looking very specifically at what's critical in terms of a patient's health seeking behaviors and what the provider understands, as you look at discharge planning, and we're all supposed to be doing that now to prevent re-hospitalization, it is critical that we understand what drives the patient into the healthcare system and at what point.

And if we can't answer that question, we can't permit those unnecessary re-hospitalizations.

CO-CHAIR CORA-BRAMBLE: Lourdes, then Grace.

MEMBER CUELLAR: Again, not to belabor the issue, but you know, I think

asking a patient what do you call your problem gives insight into their knowledge base of their disease process.

So something as simple as that. Asking questions related to transportation, a big one. Who is the primary decision maker, or who makes the decisions related to healthcare is very essential.

And then there's a lot written around religion and religious beliefs and how that effects things. But some of these questions also lead to the whole question of fatalism.

So I think while I agree, I like some of the questions, I think there is a lot that has been missed.

CO-CHAIR CORA-BRAMBLE: Okay, Grace?

MEMBER TING: Great, and since Cindy's taking feedback, and really to piggyback off of what Lourdes has said, WellPoint, over the past five years has done a

lot of consumer market research particularly around this area regarding behavioral and decision making drivers.

And we've mapped out trying to figure out what is "cultural competency." And we sort of distilled it out to five major domains that covered a lot of what Lourdes just said.

And we call them the five F's. And these are, in no particular order, food, family, faith, which we group to be both religious and spiritual belief, cultural belief. Food, faith, fear, finances and there's one more.

CO-CHAIR CORA-BRAMBLE: You said it, fear.

MEMBER TING: Sorry, five groups. Food, fears, family, faith and finances, yes. So what we've found is that when a provider or when a health center approach communication from these five domains, and certainly not every domain hits every single communication

point from a very culturally specific component addressing the needs of that specific population, the message tends to be a lot more effective.

So you know, to I think Lourdes' point on who makes the decision with your family, you know the family dynamics. The food, is it culturally appropriate. What do you believe about your disease and so on.

All those play into it. So in the future, it would be great to have some questions that reflect that. But this touches some of it.

But I do agree it focuses more on health literacy and patient-centered communication.

CO-CHAIR CORA-BRAMBLE: Okay. Mara?

MEMBER YOUDELMAN: I agree with that everyone said about -

CO-CHAIR CORA-BRAMBLE: Use your mic.

MEMBER YOUDELMAN: Sorry. I agree with everyone completely about the need to sort of go further and get more specific on the cultural issues.

But I also think that as-is, it is a really good step that, you know, is moving forward because, in part, a lot of the language, you know, interpreting measures towards the back.

But also because, as people have said, if folks are only going to do one or the other, you can get to some of the, you know, more indirect cultural competency through the patient-centered care.

So I just make that pitch as, you know, I'm looking at it as sort of good right now, can be a lot better.

But I think it's one of those where, you know, it is important to think about getting it moving forward and approved at this point.

MS. WEIDMER: Can I just add

something? I just wanted to mention that the CAHPS surveys, as part of there's a whole host of supplemental item sets that we have as part of CAHPS.

And we do have an item set on shared decision making that has been tested numerous times. It's not part of cultural competence, but it is a CAHPS item set.

I just wanted to throw that out there just so you know that it is something that we have been working on, although it didn't include it in this item set.

MS. BRACH: Well, it was originally included in this item set, and parsed out because it was being handled elsewhere.

MS. WEIDMER: Yes, so some of it is, you know, there's competing CAHPS item sets with, you know, content that overlaps.

And so we have to make decisions about where do we include them. But some of these things have been included in other

supplemental items.

CO-CHAIR CORA-BRAMBLE: All right. Any other comments, questions before we vote? All right, let's vote.

MS. KHAN: So, importance to measure and report. We're missing someone.

CO-CHAIR CORA-BRAMBLE: We have 19 this time.

MEMBER JACOBS: I can't vote.

MS. KHAN: Oh, yes. Okay. So we have 18 yeses and one no. I'm moving on to reliability. One more person. We have one high, 17 moderate, one low and zero insufficient.

And going on to validity. So we have 16 for moderate, three lows, and zero high and zero insufficient. So voting on overall scientific acceptability of the measure properties.

(Off microphone discussion)

MS. KHAN: If you can go ahead and start voting. So we're missing three people.

We've got one more. Can everyone just enter their vote in one more time? Yes, it should be 19.

Oh, here we go. So we have 17 yes and two no. And going on to usability. We have three high, 15 moderate, one low, zero insufficient. And feasibility? So we have two high, 17 moderate, zero for low and insufficient.

And overall suitability for endorsement? So we have 17 yes and two no. So the measure will pass.

CO-CHAIR CORA-BRAMBLE: Okay, so it is exactly 3:00. We have finished 11 of the measures. We have four left to go. We're going to take a 15 minute break, and regroup at 3:15.

(Whereupon, the foregoing matter went off the record at 2:59 p.m. and went back on the record at 3:13 p.m.)

CO-CHAIR CORA-BRAMBLE: All right. The first of our last four measures is Measure

number 1821. This is one of four measures submitted by GW. So we're going to start off hearing from the GW team. They didn't like us.

DR. REGENSTEIAN: Hello, everyone.

(Off microphone discussion)

DR. REGENSTEIAN: I wanted to look directly at you, Mara.

(Off microphone discussion)

CO-CHAIR CORA-BRAMBLE: Okay, it's all you.

DR. REGENSTEIAN: First of all, I wanted to thank everyone for considering these measures for endorsement. These measures --

CO-CHAIR CORA-BRAMBLE: Could you introduce yourself?

DR. REGENSTEIAN: I'm so sorry. I'm Marsha Regensteian, and I'm from George Washington University.

MS. WEST: Cathy West from George Washington University.

DR. REGENSTEIAN: And we are in

the Department of Health Policy where we had the pleasure of running a project called Speaking Together, which was a quality improvement project funded by the Robert Wood Johnson Foundation.

And that program, many of the features of that program have been included in a subsequent quality related program called aligning forces for quality.

And I just wanted to thank the committee and also just give two seconds of background, which is that when we started thinking about doing quality improvement and language services, we realized that there weren't really a set of measures for healthcare providers to guide their quality improvement work.

And so we developed this part of that program, piloted it and then tested a set of measures that today you'll be reviewing four of them that try to get to some key components in the delivery of language

services.

They're all process measures and they get to, first of all, demand for language services. So what patients in a hospital setting indicate that they prefer to get healthcare in another language.

If they have that preference, did they actually receive healthcare in that language? If they get an interpreter, does that service come in a timely fashion?

And then finally, if interpreters are providing qualified, trained services, are they using their time productively and efficiently? So with that, thank you.

CO-CHAIR CORA-BRAMBLE: Okay, thank you so much. So Mara, you're up.

MEMBER YOUDELMAN: Great. And I will mention, while I was not involved in the measure development of this, so it's not a direct conflict, I was on the National Advisory Committee for this Speaking Together project.

And the National Advisory Committee helped select the ten sites that ultimately participated. So it's not a direct conflict, but I did want folks to know that.

So with 1821, the measure is patients receiving language services supported by qualified language services providers. As I think folks who reviewed this one agree that the evidence base of need is high.

There's significant research that's been documented by the Institutes of Medicine in the Unequal Treatment report. And lots of other research and literature articles about the barriers that limited English proficient patients have in accessing care due to language.

And that having interpreters or bilingual staff who provide services directly in a non-English language can improve access, improve safety, efficacy and overall quality of care.

Other research base at this point

to support this was the Joint Commission which adopted hospital standards on accreditation.

Again, I should just disclose that I was a subcontractor to the Joint Commission and helped in that project to develop the measures and co-authored the roadmap that came out with that.

But their new standards do require that staff must be competent to do the jobs that they're expected to do in the hospital, and that the hospital must effectively communicate with limited English proficient patients.

In addition, there are a number of NQF preferred practices on providing language services and providing qualified and competent interpreter resources.

And those were part of the project that preceded this one, which a couple of us were on that panel for.

The measure itself is sort of a point in time measurement. And when the

hospitals were doing it in the Speaking Together project, they were doing it comparative monthly. So the higher the number, the better the quality.

And so the number was, you know, how many patients actually got language services by a qualified provider, whether that was an interpreter or a bilingual staff member at initial assessment and at discharge divided by the total number of individuals in the hospital who identified a language other than English and the need for language services. Am I right, Marsha? Okay, just making sure.

So it was tested in ten hospitals during the Speaking Together project, and it's also used in the Aligning Forces for Quality project going on right now.

In addition, I think one thing that I don't think was mentioned in the materials is with the requirement that was adapted as part of the HITECH Act for electronic health records, that the definition

of meaningful use does require that hospitals or provider offices who are getting incentives and funding to implement electronic health records, one of the requirements for meaningful use is to collect language data.

So I think that also shows the feasibility because as more and more providers are adopting electronic health records and are actually getting federal funding to do that if they're Medicaid and Medicare providers, they certainly are going to be collecting this data.

And so then it's just taking the next step of, you know, there should be documentation in records for risk management issues and legal issues about the provision of language services.

So it's just a next step to assessment. So I think I will leave it at that.

CO-CHAIR CORA-BRAMBLE: Okay, comments, questions from the rest of the

committee members first, I mean from the workgroup and then from the committee at large. Romana?

MEMBER HASNAIN-WYNIA: Yes, I'm part of the workgroup. So this is a question for the measure developers.

And you know, the thing that I struggled with here was the notion of qualified interpreters and the evidence base for qualified interpreters.

And the very limited number of qualified interpreters, I think, may be based on some of the work that Mara has done.

I think we're maybe at about 200. So how do we reconcile that in this measure?

MS. WEST: When we started it out, we told them to use whatever their hospital's definition for qualified interpreter is because when the Joint Commission walks in the door, or they have a CMS survey, they will ask them what qualified is for their institution.

There was an absence of that. You

know, even now we can't tell them what qualified is.

MEMBER YOUDELMAN: And I'll just clarify it because I think Romana eluded to, one of my other hats is I chair the Certification Commission for Healthcare Interpreters.

And over the last three years, we've actually developed a certification program for healthcare interpreters in three languages and a competency assessment for interpreters in all other languages.

That didn't exist at the time that Speaking Together was initiated and was preceding through. There also was a second organization that does certify interpreters.

So I think as the field also develops, there will be more recognized. And there even are now, there are even more recognized standards of what is a competent interpreter then there were when this measure was developed.

And then in conjunction with the Joint Commission standards, that a lot of hospitals, at least, are starting to think about requiring credentialing or certification as the evidence base for the Joint Commission.

CO-CHAIR CORA-BRAMBLE: Let me go around the table. Let me have Ernie and then Lourdes, and then Dennis. Yes, Ernie. Oh, and then Kevin. Go ahead.

MEMBER MOY: So thank you for raising the issue of whether it was a qualified provider. I thought also standardization would help.

I thought it was a good measure, but that would be something that's helpful, and I don't know if there are other alternatives other than certification, which might be a pretty high bar.

And how are you going to get bilingual staff and providers to actually go out and get certified when it's not their main job?

But there are other things that might be available, like, you know, specific testing for a level of language proficiency in a different kind of language that might be acceptable that's lower than official certification.

The other thing I had a problem with this measure is it seems to switch back and forth between preferred language and limited English proficiency. And those are obviously not the same.

And I think you mean preferred language other than English. But the LEP kind of slipped in there and you might want to fix that.

CO-CHAIR CORA-BRAMBLE: Okay, thank you Ernie. Kevin?

MEMBER FISCELLA: I'm unclear exactly on what the numerator and denominator is for the measure.

(Off microphone discussion)

MS. WEST: The denominator is all

patients who have identified needing a language other than English for healthcare.

And the numerator is all patients who got initial assessment and discharge instruction in that visit.

MEMBER FISCELLA: Where are the data coming from?

MS. WEST: The hospitals creates a system to collect the data. The denominator comes from screening. Screening, asking the patients what their language preference is for healthcare. So that creates the denominator.

And then the numerator is, if you're the patient, did you get interpreters at those two points in time?

MEMBER FISCELLA: Based on self report?

MS. WEST: The hospitals document receiving --

MEMBER FISCELLA: The hospital actual documentation?

MS. WEST: -- delivering the

service.

MEMBER FISCELLA: On that, okay.

CO-CHAIR CORA-BRAMBLE: Lourdes, and then Dennis.

MEMBER CUELLAR: Excuse me. Overall I like this measure. However, I'm not as worried about the interpreters or translators as I am about the proficiency for bilingual staff.

We're actually struggling with this in my own organization and we used a measure to test them. And we had many native speakers who were born, raised and trained in South America and Mexico who didn't pass the test.

And so the level of the testing for this proficiency is a question that's come up, at least in Texas. I'm just telling you that's an issue.

I mean, to what level? I mean you want to be able to communicate with a patient. But some of the questions are so high level

that even native speakers are not passing the exam.

CO-CHAIR CORA-BRAMBLE: Okay, thank you. Dennis?

CO-CHAIR ANDRULIS: Yes, I guess I agree, it's very important. There's no doubt about it.

What I struggled with when, as I was one of the reviewers of this is I would have liked to have seen, even if it were just out there for review and presentation, more of a focus not so much on the importance of interpreters, but on the issues around qualified.

The operative word here is qualified interpreters. When I look at this measure, I'm thinking okay, that's the point that we're supposed to be getting at. It's not that there isn't an interpreter needed.

And I guess what I struggled with was when I read this, when I looked at what was written, I said there's not much

discussion around what are the issues around bilingual versus full time versus part time interpreter?

What does it mean to qualify? What are the ranges, what are the experiences in terms of qualified? You know, what seems to have worked? What role does the existing organizations play?

Are there issues to resolve within those organizations? How accepted are the issues related to those organizations now as they try to expand their scope?

What prevents them from being expanded? All these and other points around the issue of qualified, because I kept on coming back to that word, they weren't there.

And I had difficulty to try to then get my hand around what was missing and what it meant in terms of something I agree with, you know, intuitively and by face validity.

But I was struggling to

operationalize it in the context of qualified.

CO-CHAIR CORA-BRAMBLE: Liz.

PARTICIPANT: I actually have a question for you.

CO-CHAIR CORA-BRAMBLE: No. I can't answer it.

PARTICIPANT: She wants you to go ahead.

(Off microphone discussion)

CO-CHAIR CORA-BRAMBLE: Microphone.

MEMBER JACOBS: Oh, sorry. I just happen to know the literature very well. And it turns out that I was part of a review where I reviewed the literature and looked at whether people got interpreters or not and whether they were qualified or not.

And that qualification was like, did they mention in the paper that there was some training or testing? So it was very vague. It was defined by the investigators.

And those interpreters like that

and it's like I think there were 30, I can't remember exactly how many of these papers that were actually outcomes based rigorous research showed that these interpreters as the investigators called them qualified.

And we had some minimum standards around it. Very minimum standards, where it actually showed impact on outcomes. The other types of interpreters didn't.

So even if it's sort of left vague like this, we have evidence that this vague definition of qualified or professional or staff is much better than any other thing that you do in terms of using family, friends and that sort of thing.

Don't get me wrong, I have some issues around I wish we measured this better and did better at it.

But there is evidence that even using this sort of you define what qualified is actually does have a positive impact on outcomes and reducing disparities for -

CO-CHAIR ANDRULIS: Is there any sense of what the range in the term qualified? I mean, is this sense of what has constituted from anything from -

MEMBER JACOBS: Yes. And there's lots of people who could answer that question for you, yes. But I don't know. What would probably be better is to hear from you what your range of qualified was, though.

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER JACOBS: I'm going to guess.

CO-CHAIR CORA-BRAMBLE: No, I agree. I think that we're privileged in having people that are part of this committee that are really experts in these areas.

But I want to make sure that as the measure is drafted and presented, that, you know, those that are the authors of it can sort of share that same nuanced perspective.

DR. REGENSTEIAN: Great. So I want to say a few things about the measure,

but I also want to acknowledge Liz and others because we developed these measures really sitting at their feet.

I mean, we drew heavily from the experts in the field who have both the real understanding of the literature, but also practice this at the bedside and so know how messy it can get.

For us, and you know, Cathy and I have talked about this for years, this is the measure that we care about. It's the most.

This is the one that counts the most for us because until hospitals started looking at this measure, they didn't even think about recording at the patient level whether someone was receiving a service.

So the Speaking Together collaborative was about 18 months long. They probably spent 17 and a half months of it wrestling with these very issues.

This was the hardest measure for them because of all these issues because they

had to define what qualified meant to them.

What they were going to do about testing their bilingual providers, because you know, some of the hospitals we worked with are considered the premiere hospitals in this area. And they don't really test very much.

They go through a testing process that sort of, not certification or testing in the field, but whether their interpreters feel that the new interpreters that they're hiring are qualified.

And they also wrestle with how they deal with bilingual providers. Most do not test. And the testing that occurs is most often not of the caliber that experts in the field would feel comfortable about.

So what this measure was so helpful in doing for them from quality improvement was really addressing all of the ways that they currently classify people, because they're implicit.

These decisions are so implicit in

terms of who gets to be an interpreter, who gets to interact with the patients. So, you know, I absolutely agree with all of these issues.

We actually struggled with the term qualified interpreter because we wanted to have some designation or qualified provider. And we did leave it to the point of the hospital because they had liability and they were doing the quality improvement.

In terms of training, we had as a threshold that we said was a 40 hour training period because that was what we felt the field had said in training programs would sort of be a minimum amount of training that an interpreter should have.

But we really, again, left that designation up to the specific hospital. And they had to document that the people who are providing this service did, in fact, meet their internal qualifications. Do you have anything else to add on that?

MS. WEST: The other thing they did, as a lot of the hospitals have external agencies where they get interpreters from, and it was the first time they had reviewed contracts to see what those qualifications from those agencies were.

And to make sure that they met their hospital's own minimum qualification standards.

CO-CHAIR CORA-BRAMBLE: Okay. Around the table, Romana, you start off.

MEMBER HASNAIN-WYNIA: So you know that I, too, am very supportive of this work and, you know, the efforts that you're pushing forward through developing these measures.

So what I'm struggling with is, you know, based on the first question that I asked which was how do you define qualified. And Marsha, you said well, we're leaving it up to the hospitals.

So what I didn't see in the measure, and maybe I missed it, was any kind

of a what's the bare minimum? Is it 40 hours, is it there? I mean, because I didn't see it.

I worry about that partially because, you know, though the Speaking Together hospitals represented a diverse group of hospitals, they were still hospitals that were doing work related to language services and had, I assume, some systems already set up.

Whereas, we know that a lot of hospitals around the country are not quite there. And so if we don't specify some sort of a base, if you will, for what qualified means, I'm worried that it's going to be left to interpretation.

So even the response about the 40 hours of training is something that adds a little bit more of a parameter to the term qualified.

CO-CHAIR CORA-BRAMBLE: Okay, Mara and then Liz.

PARTICIPANT: Well, actually Grace

was next.

CO-CHAIR CORA-BRAMBLE: Oh, she put it down.

MEMBER TING: Yes, it's kind of the same thing.

PARTICIPANT: Mic.

MEMBER TING: Oh, sorry. It is kind of the same thing. I struggle with the lack of parameters, myself.

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER YOUDELMAN: And I think, in part, that was because of, as Marsha said, where the field was even just a couple years ago.

And so I also think it's important to some degree, even though this is one measure, to bifurcate interpreters versus bilingual providers, because with interpreters, we do have National Code of Ethics, National Standards of Practice, and actually just released a year ago, National Standards for Training.

Now, none of those are mandatorily enforced because it's not like the federal government has adopted them. But the field has sort of moved.

And also, with the credentialing, both us and our competitor have said minimum 40 hours of training.

I think the profession's going to move beyond that, you know, as things proceed. But that's sort of at least the bare minimum recognized for credentialing right now.

Ernie, you're absolutely right, that there are different levels. And so we're never going to have full certification for every language. We can't, because the psychometrics with the AHRQ folks here.

You know, the cost of developing an oral exam to test interpreting skills is an incredibly expensive task. So there do have to be alternatives.

So we offer, like a credential to test knowledge, but then still leave it to the

hospital or the provider to test language proficiency. Credentialing isn't something you're really going to see for bilingual providers.

So Lourdes, you're right that there still is some sort of figuring that out. But I do also think there is been greater recognition that there has to be some assessment of provider's language skills.

And I think the Joint Commission standards have moved the field forward in that regard, because staff does have to be competent. And so how do you know that they are competent to provide services in Spanish or Mandarin if they haven't been assessed?

So there's not quite as much as I think all of us would like, but I think this is a good start, and it certainly helps move the field forward to have some requirement both for the language collection, and then the assessment. I mean, sorry, the documentation of provision of language services.

CO-CHAIR CORA-BRAMBLE: Liz?

MEMBER JACOBS: So I was also a reviewer of this measure. Not a surprise. And I want to reflect on what's happening in the room, which is that we're talking about trying to make organizations or assessing whether organizations have gone to zero to 60 in like one minute.

And the truth of the matter is still, I do this work all the time. I go around the country, you all know this. I mean, as a physician, I teach other physicians.

They still don't use even an interpreter on the phone or an interpreter who's a staff member. And they're using family members.

I can tell you a story from yesterday about it. And what you're reflecting is that you're just even asking them to assess who are they getting to interpret?

And qualified, and like I said, there is evidence behind this that even if it's not, I would love to see everyone have like the most professional interpreter.

But on the other hand, we have evidence that something minimum is still better than something bad, or nothing which also happens.

And that we're going to reduce disparities. We also know that from work, that if we actually start to get people to increase the number of times in which they offer people these interpreters, that you're calling qualified and maybe you want to use a different term, minimally qualified or something like that.

Maybe we want to change that. Maybe you want to take out the bilingual providers because that is harder to assess than the interpreters.

But I really think that this measure could go a long way to reducing

disparities based on what we know even though we don't have these great measures of what qualified is.

And even as they're defining themselves, organizations are going to have to start saying who are we using, why are we using them? Is this language lying? Do they actually really test their interpreters? Some of them don't.

And then they'll start looking and they'll say oh, they don't, so I'm taking them off the plate and now I'm using this language service.

So I just think this could really move organizations in a direction that if they met this standard, it has a high likelihood of reducing disparities for these patients. So that's my passion --

CO-CHAIR CORA-BRAMBLE: Thank you. Dennis, and then we're going to wrap it up.

CO-CHAIR ANDRULIS: Yes, nobody's arguing, I don't think, about the importance

of this. It's so vitally important. But I think it comes back to what Romana was talking about that I really believe in, too.

Is there sufficient confidence that you could create at least that base and that base would accompany any issuance related to the guidance in some way, shape or form?

That there would be at least a minimum to start with to give the field a sense of not just the concept of qualified, but that there is actually something attached to the term qualified that would be sufficiently acceptable as at least a minimum.

CO-CHAIR CORA-BRAMBLE: Mara?

(Off microphone discussion)

MEMBER JACOBS: Sorry, so you're asking them to actually beef up what they say is qualified? I'm confused as to what you're saying.

CO-CHAIR ANDRULIS: I'm trying to get to the point where could we offer some wrap around guidance to this measure to

describe what constitutes a minimum acceptable base of qualified.

So any provider out there who's not part of Speaking Together goes qualified, oh at least I've got a sense of the ballpark now. Rather than saying qualified, who knows what qualified is?

CO-CHAIR CORA-BRAMBLE: That's a fair statement. Okay, Romana? Oh, let me start down there and then I'll work my way back. So Mara, Romana and then Jerry. Liz, you have a --

MEMBER JACOBS: I'm sorry.

CO-CHAIR CORA-BRAMBLE: -- Okay, it's okay.

(Off microphone discussion.)

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER HASNAIN-WYNIA: All right. So this conversation reminds me a little bit about, you know, it's a little bit analogous to collecting data on race and ethnicity.

So, you know, we set a bare

minimum, right? We said bare minimum, collect the OMB. But ideally, collect more granular ethnicity because that's how you can do quality improvement within your organization.

So there's a bar. You know, whether we agree with it or not, it's, you know, but it's the bare minimum bar.

And what I'm struggling with, and this is what Mara and I were having this little side conversation about, is that if it is completely left, I mean I understand what you're saying, Liz. And I agree with you to a certain extent.

But if the definition of qualified is left for each organization, each hospital, physician practice to interpret, then in that context, and this is really speaking in hyperbole, and I recognize that, you know, my grandmother can be an interpreter for me if I had limited English proficiency, right?

I mean, that would be an ad hoc, you know, because we know there are different

kinds of interpreters. So you can get an interpreter through a community organization. There are contract interpreters, there are in-staff interpreters.

There's, you know, telephonic interpretation, there's video monitoring. There are all different modalities for providing interpretation.

I guess, you know, because maybe Marsha, I'm kind of pushing on this because you offered the 40 hours. And I guess I'm just kind of, I don't think that that's kind of out of the realm of reality to set a bar that is a bare minimum.

CO-CHAIR CORA-BRAMBLE: So Mara, that point is acknowledged, it's actually very similar if I hear you correctly, to the one that Dennis made. Is it the same sort --

MEMBER HASNAIN-WYNIA: Yes.

CO-CHAIR CORA-BRAMBLE: Okay, all right. Other perspectives, other comments.

MEMBER YOUDELMAN: So my counter

to that has multiple parts. I think one, existing civil rights law and the guidance that comes from the HHS office for civil rights defines and discusses what it is to be competent in interpreting.

Now, is that as enforced as it should be? No. But it is federal guidelines that is out there that does discuss what is competent. Again, does it get to the level of how many hours of training? No.

But you know, I won't get into the issues of 40 hours bare minimum. There's lots of reasons why you would want 60 or 100 or, you know, actually to specify.

It's not just 40 hours, because 40 hours of medical terminology might not be great. You would want ethics and standards of practice. But I think we do have some guidance from the Federal Office for Civil Rights.

I think we do have the Joint Commission standards, which again, it doesn't

talk about qualify, but it does, I forget exactly what the wording is.

But, you know, that staff really do have to be competent. And so that also is pushing the field. So if someone is going to be interpreting in the hospital, that they do have to have the relevant skills.

Or one, the Joint Commission can come in and hold it against them for accreditation. Two, the Office for Civil Rights could come in and investigate them and find them in non-compliance.

And then three, I do think we are starting to see the recognition in the field. So I do agree that qualified is vague.

But I think what we have seen in the development with credentialing and certification is what skills an entry level interpreter must have.

We did a national study on this in order to develop our credentialing and certification.

So, you know, we surveyed interpreters of what they're doing on the job and what the tasks are and what the knowledge skills and abilities are, and then based credentialing on that.

So I do think we're moving in that direction. And I think by again, sort of pushing this envelope and making folks think of what is qualified, which they should be doing for Title VI compliance for risk management already.

And now for Joint Commission accreditation that, sort of those three along with this standard really do sort of set the stage for getting folks thinking about this more.

CO-CHAIR CORA-BRAMBLE: So Mara, let me make sure I understand you because I heard two people say that they wanted a more explicit definition of what a qualified interpreter, what is sort of the bare minimum.

You offered some sort of a counter

argument. Am I understanding you to say that you do not think that it needs to be further clarified? I just need to be very --

MEMBER HASNAIN-WYNIA: So I would be --

CO-CHAIR CORA-BRAMBLE: -- you know, there's sort of passion on both ends here.

MEMBER HASNAIN-WYNIA: I would be fine if it is defined. I don't think that there's going to be agreement as to what the definition should be right now.

And the second piece is, I think more importantly, I wouldn't want to see the measure fail because we can't agree on a definition or we can't go back to these guys because the project is over and say test out what the definition should be.

CO-CHAIR CORA-BRAMBLE: I hear you. I'm just trying to clarify that there are some members of this workgroup that feel that further clarification is needed. I

acknowledge your point.

DR. NISHIMI: If I can jump in here and throw something out for you to think about, both the committee and the developers.

The project's over. You know, they can't go back and test. But as the measure stewards, it is within their power to alter the specifications to reference, you know, footnote qualified, and say pursuant to the Joint Commission, blah, blah, blah. And pursuant to the OCR, blah, blah, blah.

That adds a degree of specificity to the specifications. I shouldn't have used those both. And may take care of some of the concerns that we're hearing that qualified standing alone is problematic.

So is that kind of a footnote something that the developers are willing to do?

DR. REGENSTEIAN: Very thrilled to do that.

MEMBER HASNAIN-WYNIA: And I just

want to say that that's exactly. I wasn't saying the 40 hours.

I just wanted something. And that, at least from, you know, since I've been speaking up about this, would satisfy kind of the vagueness of the qualify term at this point.

CO-CHAIR CORA-BRAMBLE: Well, that was easy. Footnote, that was easy.

MEMBER YOUDELMAN: But my question is then, do we get to vote on it today to approve it pending a footnote, or does it have to go sort of on --

DR. NISHIMI: No, they just agreed to make that change. So we vote it with that change.

MEMBER YOUDELMAN: Okay.

CO-CHAIR CORA-BRAMBLE: So we will vote on the measure with the understanding and assumption that they will amend it and include that footnote --

MEMBER YOUDELMAN: Got it.

CO-CHAIR CORA-BRAMBLE: -- to clarify what a qualified interpreter is or to give us some sort of guidance to that effect. Is that accurate?

DR. NISHIMI: Yes.

CO-CHAIR CORA-BRAMBLE: Where did my consultant go? All right.

Any other comments, thoughts. So I tell you, this is what happens when you have all these fabulous experts sitting around the table that have done great work in this field for many, many years. Other comments, thoughts, perspectives? Mara?

MEMBER YOUDELMAN: I hate to ask this, but just in the sense of clarity we've talked a lot about the interpreters.

Do folks think they need a footnote for the bilingual providers, or are we taking the bilingual provider out of this?

DR. REGENSTEIAN: The Joint Commission guidance also address. I mean --

CO-CHAIR CORA-BRAMBLE: Could you

speak into the mic?

(Simultaneous speakers)

PARTICIPANT: Mara, speak into the microphone.

DR. REGENSTEIAN: Sorry, I think you could have the footnote that applies to both, right? There's some guidance from Joint Commission.

CO-CHAIR CORA-BRAMBLE: Okay, Jerry? He moved it forward --

(Simultaneous speakers)

CO-CHAIR CORA-BRAMBLE: -- there you go, so I had to interpret the non-verbal. Go ahead.

MEMBER JOHNSON: No, I'm going to move from the level of expertise to just trying to understand how you operationalize the part about the initial and then the discharge encounter.

So the data are collected, and even how you define those, to the extent that they're important. If they're not so

important just let me know.

But for an organization that wanted to meet these criteria, I'm just trying to think how you would know which part of this record to look at.

Does the initial assessment mean in the emergency department? Is it in an administration office when the person is checking in. You know, this is when they're up on a floor.

And the discharge is the last conversation with a doctor or a nurse or a home, and where do you find that recorded?

MS. WEST: We have a specification manual that defines all the terms. The initial assessment is the first encounter with a provider who's qualified to treat the patient to assess and treat.

That could be the doctor, that could be the nurse. It could be a midwife, it could be a PA and it could be a nurse practitioner.

So we spell that out and we tell them it's the first one for that encounter in the healthcare system for the first person who is qualified to do that.

So that's not the receptionist at the desk. It's not the ward clerk. It's not those people, it's people that are qualified to assess and treat.

MEMBER JOHNSON: Can I comment on that because still when I think about the real world, that still leaves variability. And it's puzzling because the purpose of the encounters can vary a lot even within a given area.

So the first nurse may not be getting as much information, even basic information after the second nurse or the third nurse and the first doctor with the second or third doctor.

So just saying the first professional who takes some assessment data, it's just hard for me. I don't know if that's

what you're trying to get at. But why don't we leave this alone because I think it may be too much detail.

CO-CHAIR CORA-BRAMBLE: But I think your point from a clinician's perspective or from a front line sort of provider, it's valid in terms of how do we operationalize it.

And sometimes there's a gap between those that write the measures and write the policy, and then those of us who are tasked with implementing it. So I think it needs the feedback, that's all.

MEMBER JOHNSON: That definition of initial that you just gave me, I would just say even though it sounds clear to you, when I think about what happens in a hospital from the time a person walks into the door until they -- that definition of initial, it did not answer the question for me.

CO-CHAIR CORA-BRAMBLE: So I think it's valued sort of feedback for those that

wrote it. Yes?

DR. REGENSTEIAN: Again, this is an area that we spent a lot of time thinking about. Our goal was not to identify the most important time. It was to identify important times.

And you know, there are trade offs. So if you have a patient who comes in through the ED, is eventually admitted, goes through days, tests. There are countless times when an interpreter could be necessary.

So we thought, what are among the most critical times. And that first initial assessment where you initially get information from the patient is important. Whether it's as important as the next interaction is debatable.

But it is when you get information that has clinical significance. Likewise on the discharge component, which are kind of combined in an ambulatory visit.

But it's to get the front end and

the back end where communication is very important. And it doesn't have anything at all to say about other times in a clinical experience that also would be important.

CO-CHAIR CORA-BRAMBLE: Okay, we're going to take one more comment from Mary, then we're going to vote. Yes?

MEMBER MARYLAND: So on the issue of interpreters, the Joint Commission says that it should be implement a language plan that establishes access at every patient point of contact. Period.

MEMBER YOUDELMAN: Right, I mean, if I can just respond. Like, this is an issue Marsha and I had many conversations about of the expectations of Title VI, of patient-centered care, of everything else is that you do provide the interpreter at every point of contact.

It isn't just beginning and end. But at least to get the field moving in this and have something that you could concretize.

At a beginning level, those were the two points in time that they identified as most important to get the ball rolling.

I think, ultimately, for lots of reasons including risk management and everything else, you should be documenting it at every point.

But this measure was sort of more limited recognizing that we have to get it started, and then, you know, you move forward.

CO-CHAIR CORA-BRAMBLE: Liz?

MEMBER JACOBS: Yes, I was just going to say from a practical measurement standpoint, that would be really hard to do. Like, how often a patient gets an interpreter.

I know as a researcher who's tried to actually document that and had like a research staff actually trying to do that, it was even hard to do.

So I actually think it is the most important times and it's much more practical to do than trying to do it at every point.

And the hope is, is that if you're documenting that, you're getting people to think about doing it every time.

I mean, it's like where you can shine the light given the limited resources of an organization.

CO-CHAIR CORA-BRAMBLE: Okay. Sir?

MEMBER O'BRIEN: Before we vote, can I just hear again what the footnote idea is because to me, that's very important if we're saying that --

CO-CHAIR CORA-BRAMBLE: It was a clarification in terms of what is a qualified interpreter, if I heard that correctly. Some sort of --

(Off microphone discussion)

CO-CHAIR CORA-BRAMBLE: Right, so that it's not left up to each individual provider or hospital to decide what's qualified or not. Some sort of quasi-objective measure.

DR. NISHIMI: Right. So, where the word qualified appears, footnote. See Joint Commission blah, blah, blah. Or blah, blah, blah.

MEMBER O'BRIEN: So if you are of the opinion it's fairly important from a validity standpoint to have an operational definition that can be implemented on a measure, as I'm leaning that way.

If it's going to have the word qualified in there, you need to be able to say conceptually what are you talking about and how are you operationalizing that?

We're voting on the validity of the measure before knowing what we're really voting on. And if it's being done in the future, how do we know what we're saying yes or no to? I mean it seems like the actual what goes into that footnote would be fairly important.

CO-CHAIR CORA-BRAMBLE: But you know, I don't think that's a huge issue. If

you want to see it in writing, somebody can draft the sentence.

You know what I'm saying? It's not, maybe not us but maybe you all. I understand what you're saying, but I don't think it's that complex. Liz?

MEMBER JACOBS: I think what I hear Sean saying is he's not sure he wants to vote on a measure where he actually knows what qualified means.

And that if we're putting these footnotes in, that it could be actually variable how people define it. Is that your comment?

MEMBER O'BRIEN: Me, personally, I'm not the one to judge whether the wording is right for how to define qualified. I'm just noting a gap between what I think should be in the specifications in the measure versus what is there.

Other people have made comments. I think Dennis was emphasizing that a lot,

that you could say something else without the word qualified in there.

That's basically, when you have qualified in there, that really adds emphasis to that particular component. That you're not just talking about what proportion got something.

You're talking about something a little more specific. To me, actually, you know, I think some concepts are maybe inherently difficult and imprecise to define.

But when I hear qualified, that connotes to me something that's relatively concrete that the qualifications are often, by law or by an accrediting agency.

You know, usually you hear that word and you think oh, that means something. That's something concrete.

CO-CHAIR CORA-BRAMBLE: Yes, I want to hear what Marsha has to say.

DR. REGENSTEIAN: Well, you know, I agree with you. But the Joint Commission,

the National Quality Forum, and NCQA have all developed guidance in this area without a specific definition of qualified.

So the field hasn't caught up yet. And these definitions, which is why we didn't define it, because we would have been setting standards for the field about practice in a way that we thought was beyond the scope of our work.

So the focus of this measure, again, is really, you know, documenting whether the patient got a language service at all. And if they did, what kind, from whom? Was it from their brother in law?

Was it from an interpreter who's hired there? Was it from a bilingual provider? And then for the organization thinking about these things, I agree 100 percent for the need for more description.

But even in the Joint Commission and the Office of Civil Rights and class standards, you don't get that specificity.

You do get guidance that you should have interpreters all the time when you need them from people who are qualified. And there's really no more specificity, unfortunately than that.

CO-CHAIR CORA-BRAMBLE: So am I hearing that you are willing to amend it, or not? I just mean --

DR. REGENSTEIAN: Oh, I'm very willing to amend it because the guiding principles are embodied for hospitals, you know, the most relevant kind of guidance is Joint Commission and the Office of Civil Rights.

And these are bodies that they recognize as being relevant to this issue. I think it strengthens the measure.

But for those of you who sort of are interested in having much more specificity about what that means, unfortunately, it's not defined in the field.

CO-CHAIR CORA-BRAMBLE: Okay, so I

want to hear you, Liz. But I also want to hear, because you raised this issue about wanting to see or understand what you're voting on.

I'm hearing they're willing to amend their measure, but I mean, I guess they can come up with the language. I just need us to give closure as it relates to this specific issue. So maybe you could think about it while Liz gives her remarks.

MEMBER JACOBS: Yes, I'm just repeat something I said earlier, which is that, again, there's research that shows even when there are these variable definitions of what qualified is, it's not these other things.

And that's actually the most important thing, that it's not like an ad hoc interpreter, it's not the janitor. You know, that sort of thing and that actually enhances outcome.

So even though it's somewhat a

black box as to what's happening, we know that that black box is better than just letting not actually knowing whether people get qualified interpreters or not.

So while the measurement may be imprecise, it's sort of like these questions around, you know, self rated health. They predict mortality, morbidity, like all these things.

It's like why? We don't know, but something about people's own perception of their health actually is related to their health and healthcare outcomes. It's the same in this sort of situation.

So I know that the data's not, while it may be imprecise, it's better than what's happening now and that this measurement. So I just want to put that out there.

CO-CHAIR CORA-BRAMBLE: Okay, Sean. Oh, go ahead.

MEMBER YOUDELMAN: In the

standard, actually Marsha and company, cite the Joint Commission way. So I think you're right. We're not going to be able to say necessarily, you must have 40 hours of training, because that's not accepted.

But what the Joint Commission says is, "The hospital defines staff qualifications specific to their job responsibilities."

And then there's a note to that standard saying, "Qualifications for language interpreters and translators may be met through language proficiency assessment, education, training and experience."

And then, "The use of qualified interpreters and translators is supported by," blah, blah, blah in Title VI of the Civil Rights Act. So you sort of get to it by getting the concepts.

But I don't think you can actually say, or this group should say, the standards for an interpreter must be 40 hours of training.

You know, you could say they need to have been trained in and assessed in terms of code of ethics, standards of practice, that type of thing because those are recognized.

I mean, I chair CCHI and I wouldn't want you to say that they must be certified, because we're not there yet, and we're not going to be there for a lot of folks.

CO-CHAIR CORA-BRAMBLE: So I'm hearing that the actual measure already has a citation that we were contemplating in terms of whether it needs to be added.

MEMBER YOUDELMAN: It has --

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER YOUDELMAN: It quotes the Joint Commission, so I think you could take that text from the Joint Commission standard and sort of adapt it into a footnote that qualified interpreters, you know, should be assessed, or are determined through language proficiency, assessment, education, training,

and experience, including the Code of Ethics and Standards of Practice. Something like that.

CO-CHAIR CORA-BRAMBLE: Okay. Liz, did you have something else to say.

MEMBER JACOBS: Oh no, sorry.

CO-CHAIR CORA-BRAMBLE: Okay. So you see, the privilege of being the chair is that I have read articles of some of the individuals that sit around this table.

And it's just wonderful to hear them debate and discuss these various issues. Are we ready to vote?

MEMBER O'BRIEN: Yes, I mean to me, I think it's fairly important to have an operational definition of a measure that's concrete.

And if the way you operationalize it is to say there's some flexibility in how it's interpreted and just, you know, basically say that, maybe that's acceptable.

I don't know. But it sounds like

there is some kind of, we're voting on the measure with the idea that there will be some text added to make it more concrete.

CO-CHAIR CORA-BRAMBLE: It may not specify ours, but it will specify, you know, these are the current guidelines in terms of what is a qualified interpreter.

That's how I understand it. Yes? Let us vote, ladies and gentlemen, distinguished colleagues. Let us vote.

MS. KHAN: So on importance to measure and report. I believe everyone is eligible this time, correct? Okay, so we're looking for 20. Right? There's 20.

(Off microphone discussion)

MS. KHAN: Oh, so it's 19. So we have 19 yeses and zero nos.

(Off microphone discussion)

MS. KHAN: And reliability? One more person. There we go. And we have one high, 16 moderate, one low, and one insufficient.

And validity. We have two high, 16 moderate, one low and zero insufficient. And the scientific acceptability of the measure properties? I have 17 yes and two no.

Usability? I need one more. Okay. We have two high, 16 moderate, one low and zero insufficient. And feasibility? We have zero high 17 moderate, one low, and one insufficient.

And lastly, overall suitability for endorsement. So we have 17 yes and two no. So the measure passes.

CO-CHAIR CORA-BRAMBLE: Okay. Let's go on to the next one. Measure number 1824, screening for preferred spoken language for healthcare. Romana is our presenter, yes.

MEMBER HASNAIN-WYNIA: Okay. So this is Measure number 1824, screening for preferred spoken language for healthcare.

A brief description of the measure is that this measure is used to assess the percent of patient visits and admissions where

preferred spoken language for healthcare is screened and recorded.

There were seven of us who were on the assessment team. But only five of us, for the most part, scored this measure.

You know, I think Mara summarized much of the evidence that's also been presented for this measure as well.

In terms of the impact, the lack of organizational information on patient primary language and screening for preferred language feels disparities.

And the measure addresses a specific recommendation that was actually put forth by the Institute of Medicine in it's standardization of race, ethnicity and primary language data for healthcare quality improvement.

In terms of the impact, you can see that, you know, three of us voted high, one voted medium, one voted low. Screening for interpreter need is clearly a necessary

first step to getting language services to patients who need them.

Though it was pointed out that screening alone doesn't guarantee getting the language services. This is purely a screening measure. It's not guaranteeing that just by screening, the language services are going to be provided.

It was also pointed out that this is not a good disparities measure. So not a disparity in asking for language need. English speaking patients aren't asked, either.

So it's not necessarily a disparities measure. So my kind of minor sidebar in this is that the measure itself is not a disparities measure, it's an important measure for assessing disparities.

So I think that we have to be very clear about that. It's a necessary first step to be able to assess disparities at the organizational level.

Again, in terms of the evidence, organizations such as the IOM, the Joint Commission, we've talked about this, NCQA. Mara raised a issue of the HITECH and meaningful use.

All of these larger bodies have asked for recording of either primary language or screening for language need. Let's make sure I hit all the points here.

There's also sufficient evidence that there is a performance gap in terms of organizations screening for preferred language.

The measure developers cited two national surveys and another study that showed that there is, you know, a great deal of variation in terms of healthcare organizations, hospitals in particular screening for preferred language.

In terms of scientific acceptability, I didn't see this, and if I missed it, I'm sorry. The developers didn't

really provide evidence of screening variation across the different settings.

So what I mean by that is variation in the inpatient setting versus the ED versus the outpatient setting.

And again, this kind of speaks to the fact that, you know, we're in some ways the evidence hasn't really kind of caught up with what we all recognize as a need, in some ways, to garner the evidence.

It's kind of a chicken/egg. You know, the chicken/egg scenario. So I just want to point that out. There is strong face validity, but there's no formal testing.

You know, just again, this measure is very straight forward. As I said, it has face validity. The measure measures what it sets out to measure. There are no exclusion criteria.

It's really hard for me, at least, to picture another more direct way of finding out whether people are being screened for

language services. It's just a very straightforward measure.

(Off microphone discussion)

MEMBER HASNAIN-WYNIA: In the Speaking Together hospitals where this measure was tested, it was pilot tested initially in two hospitals. It's something that hospitals can definitely do without undue burden.

There is a question about training staff to screen. And, you know, there clearly may be some variation. But again, the burden on the organization is relatively minimal.

In terms of usability, again, it was useful in the Screening Together learning collaborative. The measure is at the core of the organization's ability to identify language needs of it's population.

You know, I'll just bring this up again. It remains questionable about the generalizability. Again, the Speaking Together hospitals are a self selected group.

But I think that there's enough

variation in the Speaking Together hospitals that because the measure is so straight forward, you know, again, I don't think that it's going to create an undue burden or a huge variation in how healthcare organizations collect this particular measure.

There are protocols that exist for screening. This represents an early first step in helping organizations recognize the language needs of their patients.

I think there are questions about the readiness for public reporting. And that's pretty much it.

I mean, again, I just want to reiterate that, you know, of the measures that I reviewed and read, to me this was one of the most straightforward measures in the group.

CO-CHAIR CORA-BRAMBLE: Thank you. Questions, comments from anyone in the group. Yes, Donna?

MEMBER WASHINGTON: Just a point of clarification because I didn't read all of

the details. How is this operationalized? For example, if preferred language is recorded in an electronic health record, then will every single subsequent visit count?

MS. WEST: If it's recorded on a visit, the hospitals can decide if they're going to ask on every subsequent visit, or if they're going to allow a certain amount of time to go by and ask them to verify it as they do with their insurance and that sort of thing.

It's unlikely that if a person is speaking Korean in December, that they're going to be speaking a different language in June. So that's where that premise comes from.

MEMBER WASHINGTON: So the hospital decides how often to measure it or what visits count?

MS. WEST: They ask on every visit. If they have fields that are already populated, when you come into the hospital,

some of your information is already populated onto your screen.

Your insurance information, your address and that sort of thing. Once a person is asked if that information comes into the field for preferred spoken language, the hospital can decide if they're going to ask the patient again, and the field will come up blank so that they have to ask.

Or they can choose to keep it pre-populated as they do your insurance and all of that and ask if anything changed. But it's unlikely that a person who's --

MEMBER WASHINGTON: Okay.

DR. REGENSTEIAN: But in terms of counting the measure, if it is in the health record, it counts as screening the patient for language services. So they get credit for that.

CO-CHAIR CORA-BRAMBLE: Okay, Kevin and then Grace.

(Off microphone discussion)

CO-CHAIR CORA-BRAMBLE: Turn on your mic.

MEMBER FISCELLA: A question on that upper. So if somebody documents it in a single encounter, that would still count, even if the person has had, you know, another ten encounters there and it's really lost in those encounters.

Nobody's going to go back and see it. As opposed to being in that data field that gets carried forward that includes insurance, age, sex and that sort of thing.

DR. REGENSTEIAN: Cathy can answer this, too. So some of the hospitals don't have information systems like health records that follow patients throughout everything in their system, in which case it would not count.

But if you have an electronic health record where if you come in the ED and then you have an outpatient visit, if that appears in the health record, then that

appears in the health record.

And we don't require, in terms of counting the measure, during the testing periods, we did not require them to ask again and verify.

So if it doesn't appear in the record, then they don't get credit for it, even if they might have asked six months ago. But their language has to be documented.

MS. WEST: For that visit, for that encounter.

DR. REGENSTEIAN: Right.

MEMBER FISCELLA: But I mean, as a provider, if I document it in text within my EMR, would that count?

MS. WEST: If whatever you documented shows up in what you document it shows up in your subsequent time that --

(Simultaneous speakers)

DR. REGENSTEIAN: You know, it counts for that encounter. The next time someone came in or went to a different

physician, if it didn't show in the record, it would not count.

MEMBER FISCELLA: Right, that's what I'm asking. Would not count. Okay, thank you.

CO-CHAIR CORA-BRAMBLE: Okay, Grace.

MEMBER TING: I think this is more of a general comment. And certainly, I think in a face-to-face care setting, screening for and provision of language and language services is so critical to quality.

I would really like to see some way, maybe in the future or in a future iteration that this measure be reflected to include additional stakeholders like health plans, because we, as health plans, should be screening for language services, too.

And maybe it doesn't have quite a direct, you know, quality impact. But it certainly has a lot of access and sort of benefit. High level of understanding and

impact.

And I find that a lot of these measures, whether it's interpreter services or how literacy doesn't have enough of a tie in to health plans, which definitely has a role to play in all this, too.

CO-CHAIR CORA-BRAMBLE: Thank you, Mara?

MEMBER YOUDELMAN: I completely agree, Grace. And I think there's a lot of reason to expand the measures to do that.

In large part, you know, one, it's a customer service and two, it's an access that if someone calls the health plan because they're trying to find a provider or the coverage of a service or something like that, they are going to need language services there, as well. So I completely agree and support and see what we can do.

MEMBER TING: Right, and I think it would help us with language concordance linking, you know, the right member or patient

to the right providers and so on and so forth, yes.

CO-CHAIR CORA-BRAMBLE: Okay, Dennis. And then we're going to vote.

CO-CHAIR ANDRULIS: Not to belabor this point, but I think the Affordable Care Act may also facilitate this, move this along because there are the requirements, for example, for the exchanges are on class. So that it's a natural opening.

CO-CHAIR CORA-BRAMBLE: Okay.

MS. KHAN: So, importance to measure and report? You can go ahead.

(Off microphone discussion)

MS. KHAN: We need one more person. So we have 20 yeses, zero no. And reliability? So we have nine high, ten moderate, one low, zero insufficient.

And validity? We have seven high, 13 moderate, zero for low and zero for insufficient. Scientific acceptability of the measure properties? You have 20 yes, zero no.

And usability? We need one more person. So we have ten high, nine moderate, one low and zero insufficient. And feasability. Eleven high, nine moderate, zero for low and zero insufficient.

And overall suitability for endorsement. So we have 20 yeses and zero nos. So the measure will pass.

MEMBER JACOBS: Thank you. This is patient wait time to receive interpreter services, also submitted by George Washington.

This measure is used to assess a percentage of encounters where wait time for an interpreter was 15 minutes or less.

And the numerator is the number of interpreter encounters in which the wait time is a fifteen minutes or less for the interpreter to arrive.

And the denominator is the total number of interpreter encounters stratified by language. They did the same study that we talked about before to actually look at the

use of this measure.

And they did find that there is actually a variability across the sites, and a variability across languages. So it could be used to actually assess whether you're having a problem overall with interpreters or individual languages.

Looking at the criterion by which we are ranking these things. There's impact and opportunity for improvement.

And I would say it's not exactly clear what reducing wait time for interpreters would do in terms of actually improving care. A lot of patients wait a long time.

And there really isn't evidence more than anecdotal that actually waiting for an interpreter somehow delays or inhibits adequate or quality care.

So I didn't find that there's very good evidence for that. There is opportunity for improvement, especially across some languages.

And it was useful for QI for some of these organizations because they realized for instance, for Vietnamese and Chinese speakers, they actually weren't getting interpreters there in a timely manner.

And they did something to improve that. But that was, like, more internal and we don't know how globally it would impact culturally competent care or disparities.

So I would say the evidence, the quantity for the importance of this is low. The quality is not very good. And I can't really comment on consistency because the quality was so low.

It's potentially important to measure and report, but I don't feel like there's a case made for it, and I'm not sure it's very feasible.

It's open to all sorts of measurement issues around interpreters not wanting to show that they show up 15 minutes late.

Who's doing the measurement of, and maybe you guys will want to address this in your comments, who's actually measuring when the interpreter call is called and when they show up?

So I wasn't surprised there was high variability across the organizations in this study because it could be some are just doing a better job of actually getting adequate measures and others not.

And I also felt that the usability and feasability of it was difficult from that standpoint unless you have some electronic system by which you actually follow your interpreters.

Like your interpreters log in when they've been called to an appointment and then log in when they get there and time it that way.

So overall, I really felt this was not a measure that was really ready yet for our endorsement because the lack of evidence

that it would actually have an impact, that knowing this information would be important for improving quality or reducing disparities in this population.

Our overall review of this actually reflects that. Is this right? Yes, so you can see, like, for instance, the evidence we have, like one high, one moderate, one low, two insufficient.

The quality's moderate, low, insufficient, consistency, so and only one person voted that it met importance.

And so I think, actually, overall we all felt this wasn't quite ready and we don't have enough evidence yet behind it to endorse it. I'll end there.

CO-CHAIR CORA-BRAMBLE: Nice job. Comments, questions? I see a quizzical look on Marshall's face. Go ahead.

MEMBER CHIN: Well, I guess a question for you, Liz and the rest of the committee members.

You know, if a lot of the evidence for the prior ones was also more sort of face validity, if you took like the IOM Pillars of Quality in terms of time limits and patient-centeredness being a couple of pillars, would that be sort of the same type of criteria?

MEMBER JACOBS: I guess so, but we don't know how this impacts care. I mean, I have patients who wait an hour because we're waiting on a lab result to do something, and they're English speakers.

And in fact having a measure like this might actually encourage people to use the wrong interpreters because they want to actually reduce this measure because they're more timely accessible.

So I just feel there are so many issues where it's open to bias, it could encourage inappropriate use of interpreters and I think that English speakers actually wait for these things, too, wait for all sorts of reasons.

So I'm not sure how this is going to help us measure. If we measure this and there are demonstrable changes, is it really a disparity sensitive measure?

CO-CHAIR CORA-BRAMBLE: Kevin, and then Colette?

MEMBER FISCELLA: Yes, I'm inclined to agree with Liz on this. I think this may be a case where less may be more. I mean, I would be really happy if those last two measures were really hit and we really did a good job on that.

Without, at least at this point in time, adding this third measure with all of the issues associated with it. Perhaps down the road, but let's start with first things first.

CO-CHAIR CORA-BRAMBLE: Agreed. Colette? Oh, I'm sorry, go ahead.

MEMBER EDWARDS: I guess the only thing that I would say, and I don't know that 15 minutes is the right amount of time, but

I'm kind of inclined with Marshall because everybody waits too long in the healthcare system.

But if every time if I have to wait for an interpreter it's two hours, then I start not showing up. And then I end up eventually in the emergency room.

And then the way that we measure that this is important is the person in the ICU who didn't need to be there. I mean, I understand all the concerns, but I would push back a little with that.

CO-CHAIR CORA-BRAMBLE: Okay.

MEMBER FITZGERALD: I would just say that I hate time measures.

CO-CHAIR CORA-BRAMBLE: Well. Succinctly.

MEMBER FITZGERALD: If only because we've worked with a lot of the CMS measures that have to do with time to a particular procedure and the documentation necessary to establish when the clock starts.

If this goes to it's ultimate point of public accountability, it would really scare me that the vagueness here is just simply 15 minutes and there isn't the exact, you know, well when does that start?

And it really would lead to a level of specificity for a measure that I think is kind of like apple pie. But maybe that degree of specificity wouldn't be relevant for this kind of measure. So that would be my opinion.

CO-CHAIR CORA-BRAMBLE: Mara, oh. Yes?

DR. REGENSTEIAN: Can I make a comment?

CO-CHAIR CORA-BRAMBLE: Sure. By all means, join us.

DR. REGENSTEIAN: I can't disagree, you know, with your assessment of the evidence at all. But I will just give you one snippet of background on this.

And that is that when we did our

visits with healthcare with physicians, nurses, we went to hospitals, we talked to a ton of people.

The single biggest complaint about the delivery of language services was not the unavailability of interpreters, it was the wait times.

And what we felt was that the wait times were causing patients to go ahead and physicians and other healthcare providers to go ahead without the interpreter because of the perception of a wait time even more than the reality of a wait time in some cases.

So that's why we did this. The other thing is, we originally had five measures. We used five measures for Speaking Together, we used five measures for Aligning Forces for Quality Language work.

And the fifth measure that we didn't submit was another timeliness measure. It was how long does the interpreter wait for the physician or nurse and the encounter to

begin.

And that's just because the field, at the ground level, this issue of timeliness seems to be such a big issue. So, you know, I don't know what other opportunities we will have in the future.

Hopefully we'll have more disparities measures, but this issue of timeliness is a really big deal. And I know patients wait a long time, but people wait longer sometimes for an interpreter, and it can mean the service just doesn't happen.

CO-CHAIR CORA-BRAMBLE: I would add one point as a clinician, that the encounters with interpreters, that by itself takes longer.

So if we start measuring the time to get to the encounter, the counter argument can be, well you know what, it takes us twice as long to see patients when we use interpreters.

So I do think we need to be

careful with this in terms of the impact on the field. Mary and then Mara.

MEMBER MARYLAND: So I wonder if a criterion could be thinking about either patient acuity or delayed treatment outcome, because that's really what the time piece relates to most significantly.

So if it's an emergency department and you need a couple of stitches in your finger, and you wait a bit, that may not be a big deal.

But if you're in the emergency department and you've got a precipitous delivery, that's a big deal. So I don't know that we can just do this without some qualifier.

MEMBER YOUDELMAN: I have a couple thoughts on this. The first is the Office for Civil Rights has said that you can't expect LEP patients to wait unnecessarily when you're treating English patients at the same time.

So there is, again, going back to

federal civil rights laws and expectation that, you know, there not be an unnecessary delay.

Second, we have seen some states, and the one that comes to mind at least is New York which actually has set timeliness standards for interpreters because they do recognize that the waits do affect access and also care.

That people have waited so long that they, as I said, leave and then end up in other situations. And so, you know, I guess from my perspective, I understand the concerns that folks have.

But responding to Liz, yes, we all may wait an hour for a lab result. But if everyone's waiting an hour for a lab result, that's fine.

But what I don't want to see is the English speaking patient gets the lab result in 15 minutes and the Mandarin speaking person is waiting an hour and a half because

you can't get an interpreter.

The way technology is right now, you can get an interpreter in about 180 languages if you use a telephone line in under a minute. I mean, it's pretty amazing.

Yes, if you use staff interpreters, you know, they may be traveling back and forth, et cetera. But I think that the benefits of this is it does show the compliance with civil rights laws.

It is helpful from an equity perspective and it does ensure the access to care that, you know, really is at the heart of addressing some of the disparities.

CO-CHAIR CORA-BRAMBLE: Liz, and then Dennis.

MEMBER JACOBS: Well, I was just going to say, there is some face validity to thinking about this issue. But then your own data actually shows like, at one hospital that 90 percent of the time, the interpreter showed up within 15 minutes.

At other times it wasn't as good, and it depended on the language. And so I actually think that in hospitals that are doing this well, people may not be actually waiting that long. But we don't know that.

I mean, we're kind of assuming this is what happens. And then, the other thing I would say is I think this could be useful, but it's mostly for looking at disparities across language barriers, I think if you ask me, and not necessarily English speakers versus limiting English speakers.

And, I mean, going back to what Kevin said, I do think that previous measures are so much stronger in terms of if we're going to ask people to do these measures, I would much rather see the ones that I -- I mean, not that we should put that in the context of this.

But it's just not as strong as the others, as the previous measures. And we would really be, I think, measuring a limited

disparity issue. But I appreciate your point of view.

CO-CHAIR CORA-BRAMBLE: Okay. Thank you. Let me have Dennis, and then Mara.

CO-CHAIR ANDRULIS: Yes, Mara, the point you raise is really important. It's just not in the same way directly on point with regard to this measure as we try to get specific on wait times.

It's not to say that this requirement about getting an interpreter in a timely manner is appropriate. It's just that in the context of actually coming up with a specific measure that would be endorsed by this group, I think it's a bit separate.

You know, and it's not to say that it's not extremely important, but I just see it as not on point for our discussion here.

MEMBER YOUDELMAN: I mean, I fully agree that the first two are of a different caliber and quality and certainly want to see those.

You know, I'll just continue to play devil's advocate because I finally get to speak this afternoon at least one time. At least at one point, I have to disagree with Dennis.

(Simultaneous speakers)

MEMBER YOUDELMAN: You know, so anyone from the advisory panel who has those would have to disagree some time.

I think it is important in another realm, which I forgot to mention earlier, which is the planning piece. And I know it's not exactly a measurable thing.

But if you're getting your Spanish speakers an interpreter in ten minutes and your, you know, Swahili folks are waiting an hour and a half, what does that say about what you've done for screening for your languages and implementing what the screening says, which is taking the next step and making sure, okay you've screened them, now you get the language services in place.

So, I mean, I do think there is a benefit. I do agree, and you know, that there is a different quantification of this.

But I also think that it is important as a proxy for ensuring that folks are getting the same type of care and access to care as English speakers and that this is sort of what we've got at this point, again.

And you know, do we use it as a way to try to push the field to say, you know, it's not enough that you screen.

You now actually have to provide the language services and provide them in a timely manner to comply with civil rights laws, patient-centered care, equity, principles, et cetera, et cetera.

CO-CHAIR CORA-BRAMBLE: I have two counter arguments. One of them has to do with sort of the push back and backlash that I get when I'm, you know, going on speaking engagements across the country regarding this issue of culturally competent care and what

are some of the hurdles and barriers.

And I have to recognize that there are barriers and that we can push too hard. I think this issue of 15 minutes may be desirable, may even be optimal.

The reality on the clinical setting, I don't know, I find this one harder to actually implement as somebody who leads multiple clinics across the city.

So I would argue that we do need to be careful. And if I had a choice of measures that I think are slam dunks, I would go for the first two and, you know, because I think we can deal with the overkill.

My two cents, and with that I'm done. Oh, what did I start. Those were supposed to be concluding remarks.

MEMBER JOHNSON: The comments that I heard here, I was dead set against the time for reasons that Dawn gave.

But, I mean, there have been some compelling arguments made that, at least in

the mind set of patients and families, this wait time is critical.

So the question becomes what would be a wait time that would be reasonable and would maybe minimize the gaming of the system?

So with that discussion about what was the right time, 15 minutes versus. Fifteen minutes seems awfully short to me.

CO-CHAIR CORA-BRAMBLE: But that's what it says on the clinic stuff.

MEMBER JOHNSON: But I'm just curious. So when you were developing this, was there discussion about that?

MEMBER JACOBS: Oh, there was no discussion, right?

MEMBER JOHNSON: It just seemed like a good number.

DR. REGENSTEIAN: It's funny, Liz, you say that because this probably got less discussion than some of the other measures. So we had this long process, staged process to develop the measures.

And then of ten potential measures, we convened a group of experts, four individuals who ran interpreter services programs and four physicians who were using ambulatory services and therefore interpreter services.

And they were directors of ambulatory services at large health systems. And the 15 minute thing just was like yes, everybody agreed on 15 minutes. There was some debate.

You know, it was sort of does everyone think that this is a good way to at least set a standard internally in a hospital to track the timeliness of services.

So I don't think there was a sense that this is going to be a national standard at that point. But there was something that said, we can reasonably provide these services in this amount of time.

And that's a reasonable wait time to add on to a patient who needs an

interpreter. And, you know, it wasn't based on literature.

There's no literature on this in terms of, you know, what's a realistic thing to wait for an interpreter. And I think that the usefulness has really, I mean first of all, it was paid attention to it.

Second of all, it did highlight disparities across populations within one hospital.

And the third thing was that there was push back from some of the clinical staff to use interpreters because they said I wait too long.

And if the interpreter staff could show that these numbers were reasonable, there was better buy-in in terms of their training for use of interpreters.

But you know, that 15 minute number, it was kind of just everybody kind of agreed with it. And then they were reviewed again by a much broader group, and the people

seemed to find that to be a good number.

(Off microphone discussion)

MEMBER WASHINGTON: Yes, I just wanted to add a couple of comments about unintended consequences. People have already talked about the problems with the evidence around this and with the validity.

But just thinking about some of the unintended consequences, it seems that this would place an even greater burden on healthcare systems that serve a large number of patients that need interpreters.

That they'll really be the sort of highlighted as not meeting the standard and may shift resources in an undesirable way to try to achieve this standard.

And then the second thing in terms of usability, even though we may not necessarily get there given the other criteria, but I'm not sure that patients will understand how to use the results of this.

So people have pointed out, for

example, that the results aren't stratified by particular language.

So if you're someone who speaks a language for which there aren't interpreters that are commonly available and you're reviewing statistics that may reflect, for example, Spanish language interpreters, it's not very helpful and it's actually misleading.

MEMBER FITZGERALD: Just one comment on the explanation around the 15 minutes, which I appreciate the honesty in terms of hey, it sounds like a good number to me.

But are we not, when we endorse a measure then setting a national standard? And if that wasn't the intent of the measure to say 15 minutes, then I have concerns about whether or not it's really an endorsable measure without that critical evidence surrounding that number. So that's just my comment.

MEMBER YOUDELMAN: Again, I don't

want to see this sort of as competing with other standards, because I think we're supposed to be assessing each standard sort of individually at this point.

And then if there's competition or conflict later, we sort of address it. And so that was a little bit concerning from the comment that I heard from you, which is yes, I agree. The other two are great.

But I want to see this one sort of evaluated on it's, you know, independently as opposed to in comparison.

(Off microphone discussion.)

MEMBER YOUDELMAN: I still go back to the technology factor, that it can be done in 15 minutes. It should be done in 15 minutes to ensure equity and compliance with civil rights laws.

Where we're getting the push back is from folks who don't understand what their hospital's, you know, policies are, or don't know how to get to the language line, or don't

know how to sort of get that scheduled in advance.

And I think that's part of the problem. And we get that a lot with language services.

And that's why people still do grab the family members, because they're right there and they don't have to wait or figure out what the code is to call.

So, I mean, I understand what folks are saying and I understand the concerns with it.

On the flip side, and I'm still going to just, that my opinion as an advocate is to push for it is I do want to see something measurable that is showing that we're not asking LEP folks to wait significantly longer than an English speaking person.

Now, it's not you have to see the person in 15 minutes, but that you shouldn't be waiting for, you know, the interpreter for

more than 15 minutes.

So if everyone's waiting an hour and a half to get, you know, triaged, it's not like you're going to see the LEP person in 15 minutes.

But that once you get to that hour and a half, you know, you should get an interpreter within 15 minutes.

CO-CHAIR CORA-BRAMBLE: So we have those that are in favor, and those that are not. It's time to vote, all right? Here we go. Ms. Khan.

MS. KHAN: Okay, importance to measure and report.

(Off microphone discussion)

CO-CHAIR CORA-BRAMBLE: Do we all have to --

PARTICIPANT: Yes.

CO-CHAIR CORA-BRAMBLE: What is your question? What?

PARTICIPANT: Do we all have to vote?

CO-CHAIR CORA-BRAMBLE: Yes, you have to vote.

MS. KHAN: We're still missing two of you, so two of you. One more. Oh, we're going to, okay, so we have nine for yes, ten for no.

CO-CHAIR CORA-BRAMBLE: Okay. The last measure of the day, 1831. And our presenter --

MEMBER YOUDELMAN: Can I just say I apologize, but because of childcare obligations, I'm going to probably have to leave before voting on this. So I'll see you all tomorrow.

MEMBER EDWARDS: This will either be really short or really long. Because we're now --

PARTICIPANT: Longer than 15 minutes?

CO-CHAIR CORA-BRAMBLE: That's the catch, madam, we are at 4:45.

MEMBER EDWARDS: Okay, so it will

be short. My presentation's going to be very short because I mean, this is just like a one step further removed from the concerns that people raised on the previous measure.

So this is the percent of the work time that's spent by interpreters providing interpretation in clinical encounters.

And so the concern here is that providing the services is potentially very costly and potentially scarce. The technology not withstanding.

And therefore, people who are interpreters should actually be spending their time interpreting as opposed to all the other things they may get looped into doing.

It's been established, the hospitals that were involved, because it's the same as all the previous measures.

I would say that, something that I would add that I don't know necessarily has been brought up before, but there was lots of interaction with the field including focus

groups with patients.

This measure has been accepted as part of the AHRQ National Measures clearinghouse. There was a mix of hospital types. There was a mix of languages tested.

And some of the concerns that were raised by the people reviewing were the variability and the types of interpreter services that were available.

We won't even revisit the whole definition of qualified. The quality of the studies that were cited and then the feasibility of data collection.

In terms of the actual findings, the overall score for the hospitals was low in terms of low meaning a low percentage of the time that the interpreters are actually spent doing interpretive work.

And it ranged from ten percent to 73 percent. And that, over time, seven out of the ten hospitals increased by at least five percent.

So the overall conclusion by the group reviewing this as a whole was that this was another measure that was not yet ready for prime time. And particularly not ready for prime time relative to public reporting.

CO-CHAIR CORA-BRAMBLE: Okay, comments, questions? As she said, it could be very short or very long. Yes?

(Off microphone discussion).

CO-CHAIR CORA-BRAMBLE: No, no, no, please.

DR. REGENSTEIAN: Okay, so this is one of the measures that, from a quality improvement perspective, was very important to us because, you know, that measure, the first one we talked about, the one that was the L2, did you get an interpreter.

If you screened and you need an interpreter, did you get an interpreter? Or did you get a qualified, I hate to bring that up again, but did you get a qualified language service, okay?

So for organizations that are beginning to really address this need, that finding, okay, I'm a hospital. I find out that 30 percent of the time, my patients are getting the language services that they need.

The thought could be we need more interpreters. We need more resources. And you know, all the guidance is to get all that you need at all points of care for all patients.

The reality is that's not what's provided. And so I think it's important to have some sense of productivity and efficiency and appropriate use of resources for the people who are doing these kinds of quality improvement projects.

So this is kind of more of an internal measure, or a way that a hospital can track whether it's using it's resources as effectively as possible.

It doesn't have a direct patient care link. But without any measures at all

that do this, hospitals are going to have, any healthcare organizations are going to have a really hard time determining whether they have the right capacity to deal with their patient populations.

So in the field, all of you know, unqualified people are used all the time to do things that they shouldn't do. And really qualified people are used to do things that their qualifications are higher than they need to do.

So people who are very qualified, and in interpreting sometimes call patients for reminders, sometimes will help patients walk through the hospital because they're the ones that speak the language and take them and show them, you know, how to get to a medical test or something.

And so the goal of this is really to track utilization and productivity with that measure, that's the first measure that was approved.

CO-CHAIR CORA-BRAMBLE: Kevin?

MEMBER FISCELLA: Couple thoughts. One is that waste is sort of an inefficiency, a sort of a hallmark of the U.S. healthcare system.

And so I would hesitate to really focus, and I'm not saying it's not important to address, but I'm not sure I would begin with focusing on interpreters in this context.

In addition, I worry a little bit about the unintended consequences here. I mean the solution would be to get rid of your staff of interpreters and contract with one of the language line services to meet that measure better.

CO-CHAIR CORA-BRAMBLE: Any other comment. One thing I would add to this from sort of somebody who does healthcare management a lot is that this is a management responsibility as it relates to the people that are in charge of interpretive services.

So if they are not being used in

that way, I don't know that a national measure is what's needed. What you need is really strong management and leadership of your interpretive services program. That's my perspective.

MS. KHAN: So importance to measure and report? We'll have all 21 people vote this time. So we're missing two people. And we have two for yes, and 19 for no. So we will not go further.

CO-CHAIR CORA-BRAMBLE: Dr. Burstin? It's 4:55.

DR. BURSTIN: Nicely done.

CO-CHAIR CORA-BRAMBLE: Thanks to all of you. Great job. Thank you all, it was a pleasure. It was a real pleasure.

MS. MCELVEEN: Yes, thank you to Denice for plowing us through our measures today. She won't be with us tomorrow, so thank you very much.

So quickly to the group, first if we gave you a thumb drive with materials on

it, we do need those back.

And then secondly, one of the things that we will discuss tomorrow, and to give some thought tonight is any gaps in measurement around disparities in cultural competency.

So again, I know we touched on this a little bit when we reviewed the commission paper, and obviously when we drafted the Call for Measures.

But in light of these new measures that we have submitted, we do want to take the time to get some further feedback from the group regarding that.

The other thing we will be doing tomorrow is reviewing the disparities sensitive measures assessment that we've been working on. NQF staff will be going through the results of that process with the group and getting some feedback.

And I would also like to just quickly get a show of hands of people who will

not be here in person. Who will not be here in person tomorrow?

I know Luther, you will be gone. Anyone else who will not be here in person? Norman, okay. All right.

OPERATOR: We have no phone participants at this time.

MS. MCELVEEN: Breakfast is 8:30 tomorrow morning, and the meeting starts at 9:00. Thank you guys.

(Whereupon, the the above-entitled matter was concluded at 4:52 p.m.)