

WWC Recertification for Standards Version 4.0 Office Hour #2 Transcript January 26, 2018

Hello, everyone, and thank you for attending today's webinar, What Works Clearinghouse Recertification Training Office, Hour 2. The webinar will begin with an introduction from Chris Weiss, Senior Education Research Scientist at the Department of Education's Institute for Education Sciences. After that, the recertification team will respond to the questions about the recertification process and the Version 4.0 Group Design Standards.

Before we begin, here are a few housekeeping items. At the bottom of your audience console are multiple applications which you can use. You can expand each widget by clicking on the "maximize" icon at the top right or by dragging the bottom right corner of the widget panel. If you've accessed the audio through the teleconference line, you may experience a slight delay. If possible, we encourage you to listen through your computer or device speakers. We encourage you to submit questions throughout the webinar using the Q&A widget. You can ask questions at any time. Because we are recording this, the audience is in listen-only mode. That means the only way to ask questions is through the Q&A widget, so please do that. We will try to answer as many questions as possible. A transcript of the webinar is available at the What Works Clearinghouse website for download. So, with that said, I now introduce Chris. Chris, you now have the floor.

Alright, thank you, Brian. Hi, this is Chris Weiss from the What Works Clearinghouse at the National Center for Education, Evaluation, and Regional Assistance at the Institute of Education Sciences, U.S. Department of Education. It's my pleasure to welcome you here to the second office hour of the WWC Recertification Training. We are happy you could join us to hear more about Version 4.0 of the What Works Clearinghouse Group Design Standards and Procedures. And we hope this session will help you address questions that you may have—or questions you don't know you have yet—and provide more information about our new standards and procedures. With that, after my brief introduction, it is my pleasure to turn this over to Neil Seftor at Mathematica Policy Research. Neil?

Thanks, Chris, and thank you all for attending today. This is the second of two planned office hours as part of the Version 4.0 Standards recertification effort. The first set of office hours were on Monday, January 22nd. You can access the transcript, and the webinar will be made available, on the WWC site. But if you registered for that webinar, you can still access it through the link that you used to register. Similarly, with this webinar, the webinar and transcript will be posted, but you can always use the link that you used today to view it on demand. Both of these are follow-ups to the recertification webinar that was on January 12th that will also be made available on the WWC site.

So, today, we will be trying to answer any questions that people have from both the webinar on recertification and the cluster design standards module. We have had a few questions submitted ahead of time, so we will go through those first. Dana Rotz will be answering the majority of the questions, and she will be helped out by Alison McKie and Elias Walsh. We will go through the submitted questions first. And then, if there are additional questions that are submitted during the webinar, we'll try to answer those as well.

So, we have a few questions that were submitted. Some of them were kind of general questions or about the process, but some were related to changes in the standards, so I think we will address those first.

So, the first question, Dana, is about missing data. The question is that: There seems to be some confusion between the Standards Handbook and the missing data standards slides about whether, in the case where baseline data are sometimes imputed but outcome data are always observed, the baseline sample sizes for the sample with observed baseline data are required to assess equivalence. More specifically, in cases where some baseline data is missing or imputed, is the number of observations for the sample with observed baseline data required to assess baseline equivalence?

Those sample sizes are required for two purposes, estimating the pooled standard deviation and performing the small sample size correction to the baseline effect size, or in this case, the largest plausible baseline effect size accounting for the missing data. So, they are required in those two ways. We will look into the clarity of that in the Handbook. And thank you very much for pointing it out.

Yes, thank you. We are always trying to make sure that things are clear and transparent, and sometimes, it is hard to do with some of these trickier issues.

The next question is: In listing the data required to assess baseline equivalence, when some baseline data is imputed or missing, the Handbook says an estimate of the baseline difference based on study data is required. What does “based on study data” mean?

“Based on study data” means the baseline difference could be either the baseline difference calculated using observed data only, or the baseline difference calculated using observed and imputed data. So, if you look at Slide 53, that provides some further clarity on the precise measures and statistics that are required in order to assess baseline equivalence.

[Reference Slide 53 from the January 26nd, 2018, recertification webinar]

Ok, thanks.

The next question is related to baseline equivalence and assessments. The question is: Are standardized assessments conducted in different grade levels always considered to be different assessments, even if they are designed on a vertical scale or otherwise aligned? And what about parallel forms or computerized adaptive tests?

If the test has documentation that indicates the direct comparisons of different versions of the test, or across different grades, can be made, then these can be considered the same test. If you are in this situation, you should consult with review team leadership if you are uncertain, and the team’s content expert may be asked to weigh in on that decision.

Ok, thanks. We have a question related to the slide that is already up. The question is whether you can clarify the difference between Option 1 and Option 2 on this slide?

Right. So, either the data from Option 1 or the data from Option 2 is required in order to assess baseline equivalence, but not both. Essentially, what this boils down to is, when you have baseline data which are imputed, you can estimate the largest plausible baseline difference accounting for that imputed data by either taking the baseline difference calculated using the observed data or by taking the baseline

difference calculated using both the observed and the imputed data. So, in the Handbook, these are, I think, formulas C1 and D1, use those different statistics. So, that is why you can use Option 1 or Option 2 in order to assess that largest possible baseline difference.

Thanks. We have a couple of more general questions. Not specifically related to the changes in standards, but are useful to address. One question is: Can a study be rated “Does Not Meet WWC Group Design Standards” for failing to report statistical significance correctly?

We would not change the rating of a study based on how the statistical significance is reported. The calculation of standard errors, p -values, and significance levels does not affect the study’s rating. It only affects whether the WWC reports a p -value or significance level associated with a finding. So, in some cases, if there are issues in how the standard errors, p -values, or significance levels are reported, the WWC will adjust the statistical significance of a finding using a post-hoc adjustment. For example, to account for a mismatch between the unit of assignment and the unit of analysis for cluster assignment studies, there could be a post-hoc adjustment made. Or to account for multiple comparisons among the main study findings.

We will not correct for other issues using a post-hoc method. For example, if the authors use regression imputation but do not adjust their standard errors for the missing data, in those cases, the WWC will simply not report a p -value or statistical significance associated with finding.

Ok, thanks. We have another follow up on the slide you have up, which has a lot of information. Someone is still a little unclear about which approach is Option 1 and which approach is Option 2?

Right, so if we look at the third column here, Option 1 uses the baseline means by sample using the observed or imputed baseline data. So essentially, there you are taking the baseline difference that’s calculated using observed and imputed data, and then, you are using the formulas based on that. It is called g sub XI, I believe, in the Handbook. Option 2 uses only the observed baseline data. That’s the g sub XR option.

Then, this is a little bit different here, for the case where the outcome data is sometimes imputed and the baseline data is sometimes imputed, in the final column. There, Option 1 is a slightly different sample, whereas Option 2 starts with the g XI, the baseline difference using the observed or imputed data, whereas Option 1 starts without using the imputed baseline data.

Thank you. We have a couple of questions related to the recertification test and the process. One user has not received their credentials to access the test and wonders what they should do?

Yes, so, if you are in this situation, you should please contact the WWC recertification team at WWCRecertification@mathematica-mpr.com. So, that email address is right here.

[Reference Slide 63 from the January 26nd, 2018, recertification webinar]

If, for some reason, you forget that email address, it is posted on the WWC website, on the recertification page, but you can always contact the Help Desk. That will help you get in touch with whoever you need to get in touch with to get you set up with an account.

Right, the next question regarding the recertification test is about one of the questions on the exam, a bit of clarification. So, on the recertification exam, the last sentence of Question 2 states, “In addition to

the effect size associated with the intervention, the authors report unadjusted means and standard deviation of the outcome measure for the analytic sample.” Then the question this person has is: Were the data provided separately by study groups or together?

You can assume that the means and standard deviations of the outcome measure were provided separately for the analytic intervention and comparison groups in this case. Apologies if that was not clear.

Thank you. Actually, those were all of the questions that were submitted prior to the webinar. I’d like to give people a few more minutes to submit questions now, where we have time to answer more questions if you have them. Just to remind you, the original recertification webinar that covered the changes in the standards, along with the cluster module for online training, and these two office hour webinars will all be available to you on the WWC site along with transcripts. Then, you should hopefully have login information to access the recertification exam.

Because the transcript and webinar are going to be made available for the previous webinar, we won’t be covering all of those questions again. But all of them will be, the questions and answers, will be available. If you have follow-ups based on those questions, please feel free to let us know.

There is a question about when the transcript will be available. I am not sure how long it takes the transcript to come from, I think it comes from ON24, but I believe it will be up next week, if I am correct.

I think you might even be able to access, through ON24, the actual recording as well, but I am not positive about that.

Yes, the recording will be available approximately one day after the webcast using the same audience link.

Right.

Okay, so we have another question, back on our favorite slide, Slide 53. The question is: Where are Option 1 and Option 2 instantiated? Where are they used? Are they conducted in the SRG, or are they done elsewhere?

[Reference Slide 53 from the January 26nd, 2018, recertification webinar]

The SRG will essentially ask for the required data in order to assess baseline equivalence, and the reviewer will input all the information needed. Then, the SRG will output an effect size, which corresponds to the largest plausible effect size accounting for the missing or imputed data. While it is great that we understand the formulas, there is really no need to be able to use them in order to calculate precise baseline differences. The SRG is set up to do that for you.

Right, it prompts you for information to be able to answer all of these questions, so that you know what you need, and when you put it in, it does the calculations for you.

We have a question about cluster designs. If students are assigned to conditions individually and then placed into tutoring groups, is this a clustered design, since treatment is provided in groups, but random assignment was not at the cluster level?

In classifying the design, we focus on how sample members were assigned to the intervention and comparison conditions, rather than how the intervention was delivered. In this case, random assignment was done on an individual level. So, we would consider the study to be an individual level. I am not sure if it explicitly said random assignment design. If it was randomly assigned, we would consider it an individual-level RCT, and if they were non-randomly designed, it would be considered an individual level, quasi-experimental design study.

We have one other question that came in here.

We have another question related to when to think of samples in a cluster design. The question is: Is it appropriate to determine whether sample members were excluded at the attrition step for an RCT, and the joiner and representative steps for a QED?

I am not quite sure about the meaning of this question. For RCTs, you are going to assess cluster-level attrition, and you are going to assess individual-level nonresponse. Also, whether there are joiners. Then for QEDs, or cluster RCTs that don't pass any of the requirements—any of those three requirements that I just mentioned about low cluster-level attrition, no risk of bias from joiners, and limited nonresponse. For QEDs and RCTs that don't pass those bars, or compromised RCTs, you will assess baseline equivalence, and you would also assess representativeness of clusters at baseline and outcome, potentially. I am not sure if that gets to the question, maybe, I don't know. Elias, do you have a different read of this?

I'll just add that in cluster-level assignment studies, you do want to think about the sample exclusions. One design that I know the WWC has seen before, for example, is when a study assigns clusters to conditions but then focuses on clusters, let's say on schools. But then within the schools, focuses only on students with some demographic characteristic, like low-achieving students. That's okay. The end of the cluster-level assignment section of the standards describes how to think about sample exclusions like that, limiting the sample within clusters to certain students.

Just as we would in an individual-level RCT, the reviewer should think about whether that exclusion is exogenous, meaning whether it's based on characteristics that were clearly determined prior to the intervention. If the answer to that is "no", if it is possible that the exclusions were based on something that may have been affected by the intervention, that has implications for calculating attrition, calculating, you know, all these things. It will lead you to a case where you might have compromised random assignment, or where some of those exclusions may count as attrition rather than being a subgroup, where you do not have to count the lost sample as attrition. There is a lot there. The Handbook lays it out. There is a new graphic in the Handbook that walks you through sample exclusions, whether they count as attrition, whether they're subgroups, whether they compromise random assignment. All of that applies to cluster-level assignment studies as well.

I think that addressed the question, which was really trying to figure out when to consider exclusions. But if there are still some more follow-ups, please let us know.

We have another question related to attrition. It's asked in relation to the cluster module, but it is a general question. Can you clarify how exactly the WWC distinguishes attrition from nonresponse and whether the difference matters?

So, in a cluster-level study, we consider individual nonresponse and cluster-level attrition. Cluster-level attrition is calculated in the same way it was essentially calculated under Standards Version 3.0, where you look at how many clusters were randomly assigned, and then you look at how many clusters are present in the analytic sample, and you compare those two numbers to determine cluster-level attrition. Individual-level nonresponse has to do with what proportion of the individuals who are in the cluster are providing data for the analytic sample.

Nonresponse is actually only assessed when any joiners included in the analytic sample do not provide a risk of bias. For example, if late joiners are included in the analytic sample, and late joiners pose a risk of bias, then we don't assess individual-level nonresponse. We skip in the review process to the assessment of baseline equivalence. But in the case where joiners included in the analytic sample do not pose a risk of bias, we are going to calculate individual-level nonresponse by using the individuals present in the clusters that have not attrited at the earliest possible point in time after all joiners have joined the clusters. So, in this sense, attrition is at the group level, and nonresponse is really about representativeness of the individuals in the group, after those last joiners that don't pose a risk of bias have joined the study.

Yeah, and I will just add, there is a semantic reason to use the term "nonresponse," related to what Dana just described, which is, attrition is always relative to a randomly assigned sample. Nonresponse may not be relative to the randomly assigned sample; it could be at some point after random assignment, after joiners have joined the sample, after some individuals have already left the sample. The use of the term "nonresponse" acknowledges that it is not always measured from the point of random assignment.

Thank you both for clarifying that. We have addressed the questions that have come in so far. I would like to give people another minute or two to see whether there are any other questions. I did check on the WWC recertification page, and the recertification webinar has been posted, or I'm sorry, the transcripts and slides have been posted on that page, so they are available, and as I said, the ones from the two office hours will be posted as soon as available.

I don't see any other questions. Are there any other things we need to remind people of? Elias or Dana?

I will just mention that we have had a couple questions about the SRG, the Study Review Guide module, which will be a requirement for completing your recertification. That module is not available; it is still in development. But we do expect it to be available at some point, and at some point, you will be able to check off that last step. But do not look for it on the recertification site right now. You will not find it. It is coming, and we will let you know when it is available.

Right, and the SRG itself will hopefully be made available in the next couple of weeks on the website. Hopefully all those remaining materials will be up soon so you can finish the process. We still haven't received anymore – Oh, sorry, go ahead, Dana.

Yeah, I would just take this opportunity to emphasize that this is not your last opportunity to ask questions. You can continue to send any questions that you have to the recertification team at WWCRecertification@mathematica-mpr.com. We will do our best to respond to you via email.

I think that wraps it up for today. As Dana says, if you do have questions or if you think of more, please feel free to send them in. At this point, I will turn it back over to Brian to conclude the webinar.

WWC Recertification for Standards Version 4.0 Office Hour #2 Transcript

Thank you. This concludes the webcast for today. The on-demand recording will be available approximately one day after webcast using the same audience link that was sent you earlier. The transcript will also be available at the WWC website later this month.

You can submit any feedback to the team through the contact us form on our website: [WhatWorks.edu.gov](https://www.whatworks.edu.gov). Thank you, and have a great afternoon.