

RTT-SIG OIRA Questions

(Original questions in black text; responses in blue text)

Overall/Supporting Statements

1. Supporting Statement A references the use of semi-structured interviews with State officials. However, in reviewing the State Protocol, the instrument appears very structured and does not include many probes. Please explain what ED was referring to in SS A?

The RTT-SIG state interview protocol (and also the district interview protocol) should be described as “structured;” we’ve made that change. The protocol uses a structured format – that is, it asks structured questions and offers specific response categories – to ensure systematic coverage and consistent data collection on the topics of interest across all respondents. The collection of such systematic, consistent information will, in turn, support the reporting of objective, quantifiable implementation information – e.g., on the prevalence of particular policies or practices across states and districts – as well as reliable assessments of change for the six reform areas that are the focus of the RTT-SIG evaluation.

The structured format used can mistakenly give the impression that the RTT-SIG interviews will provide few opportunities for elaboration or open-ended responses. This is not the case. Many of the RTT-SIG questions include “probes” (i.e., a “SPECIFY:___” follow-up) for interviewers to gather additional detail on the closed-ended response category chosen (see, for example, questions SC13, DA13, TL51, TA9, and CH22 in the attached revised version of the state instrument). Each question also includes a “notes” field for interviewers to succinctly capture details or explanations about state-level reforms or policies that are offered by the respondent (for instance, why a particular response option was chosen or the nature of reforms being implemented), even if the question does not probe explicitly for such details. The descriptive, qualitative information gathered through these “probes” and “notes” will serve as the source of illustrative examples used in evaluation reports.

2. What existing *state* data collections are there, and what other alternatives were considered outside of those proposed in this data collection?

IES is also carrying out the Integrated ARRA Evaluation, which includes the collection of data from all states (in spring 2011 and spring 2012). We carefully considered whether the RTT-SIG evaluation could coordinate its collection of data from states with the Integrated ARRA Evaluation, but after closely comparing the draft instruments from both studies, we decided that it would not be feasible for a number of reasons. First, the timing of the data collection for the two evaluations is different (with just one year of overlap), and the RTT-SIG evaluation is slated to have two more years of data collection beyond that planned for the Integrated ARRA Evaluation. Second, our detailed comparison of the state data collection instruments for the two studies showed that, while there is overlap in the *broad* topic areas covered by both instruments, there are extensive, substantive differences in the questions asked within each topic area and the level of detail sought by the instruments. To address the research questions laid out in the evaluation’s Performance Work Statement (PWS), the RTT-SIG instrument

generally asks more detailed questions than the Integrated ARRA Evaluation and offers finer-grained response options. Third, the RTT-SIG instrument was designed to support implementation data collection in the context of an impact evaluation. Hence, it was designed such that the *same* questions are asked of *all* states, so that the resulting implementation data can be used to compare policies and reform efforts in RTT and non-RTT states and to help interpret impact findings. The ARRA survey did not use this approach for some of the topics covered, so that data could not support these important comparisons between RTT and non-RTT states. Finally, the Integrated ARRA state data collection is through a survey, whereas the RTT-SIG instrument is collected via an interview. Data generated from two different formats may not be comparable (see paragraph below and response to question 1). However, given that the evaluation seeks to document change consistently over time, if we were to use Integrated ARRA survey data this year but then have to switch back to RTT-SIG interview data for the same questions in subsequent years (because this year is the final year of Integrated ARRA data collection), then it will be difficult to ensure a consistent measure of the types of change we wish to identify over time. We have documented our analysis of the two study instruments in a crosswalk table and have summarized the findings from the detailed crosswalk analysis. We are happy to share these two documents with OMB if requested.

In addition to considering whether it might be possible to coordinate data collection with the Integrated ARRA Evaluation, we also considered whether a survey format for collecting data from states might be a possibility instead of the proposed interview format. After careful consideration, we determined that the interview format is better suited for the aims of the study, for several reasons. First, we wanted to be able to capture more nuanced, descriptive information about the reforms. To accomplish this, we have built in “probes” and a “notes” feature into our data collection procedures (see response to question 1). Second, we feel that an interview format will result in high quality data, as interviewers will be able to clarify any questions from respondents as the interview is being conducted to ensure that the respondent understands the questions as intended. Third, the experiences of colleagues on the Integrated ARRA Evaluation suggest that a survey approach may not end up reducing burden. We have been told that the survey format on the ARRA Evaluation proved to be quite challenging and ended up requiring extensive, unanticipated follow up with respondents who had questions about the survey or who left many questions unanswered. (With an interview format, respondent confusion could have been more efficiently and completely addressed during the interview, which would have reduced burden on respondents by reducing the need for follow-up.) An interview format is also more flexible and amenable to the fact that – in some states – multiple respondents will need to respond to our state interview; this issue would likely be more challenging to address in a survey format. Finally, to facilitate an efficient interview (that reduces the amount of reading and repeating the interviewer needs to do), respondents will be provided with copies of the questions and responses during the interview so that they can more easily follow along.

3. Are there ways of incorporating data that are currently being collected through other RTT/SIG related evaluations already?

As laid out in our response to question 2 above, we carefully considered whether we might be able to coordinate data collection with the Integrated ARRA Evaluation. We also carefully considered whether

we might be able to coordinate data collection with the Study of School Turnaround (SST), another study currently being conducted by IES. We ultimately decided that coordinating with the SST would not be feasible for several reasons. First, the SST study is focused only on School Improvement Grants (not Race to the Top). Second, the SST sample would not allow us to meet the goals of the RTT-SIG evaluation. In particular, it is critical that the RTT-SIG study sample include districts, schools, and states in which the planned regression discontinuity approach is feasible. The SST sample was not designed for this purpose, as it is a set of case studies in roughly 35 SIG schools and 6 states. The RTT-SIG evaluation has a much broader sample to reflect its broader scope. Finally, the focus of the SST data collection is very different than what is planned for the RTT-SIG evaluation. The RTT-SIG evaluation focuses on concrete requirements, actions taken, and reforms implemented by states, districts, and schools. The SST, in contrast, focuses on the change *process* and the way in which the SIG reforms are perceived and rolled out in states, districts, and schools. Their data collection is, therefore, focused much more on the details of implementation.

To reduce burden on respondents, we plan to incorporate data that are already being collected through other avenues whenever possible. See the response to ED Branch's question 3 for additional details on these plans.

4. Given the lengthiness of the state interview, and it is being proposed to be conducted orally through phone, are there considerations to either 1) scale back the length of the interview (from a methodological standpoint, in order to improve reliability) or 2) conduct it on the web instead?

Because the RTT-SIG evaluation covers a wide range of policy areas (data systems, standards and assessments, teachers and leaders, turnaround, charters, and state capacity for reform) that must be examined in considerable detail, we expect that multiple individuals in a given state (or district) will need to provide information. We also feel that an interview format will be more flexible and amenable to this multiple-respondent scenario than a survey format.

We do not plan to administer the full state-level interview protocol as a single 4.5 hour interview. The RTT-SIG state interview protocol comprises six separable modules corresponding to the six policy areas noted above. We expect to administer these modules on separate occasions to different respondents, with none of the individual modules taking more than 60 minutes. The planned module approach—in which each respondent might be designated to answer questions for one or two modules—was used during our pilot-tests, and it worked quite well. Respondents found the length of the telephone interviews acceptable, and we usually only conducted one or two modules at a given time.

The questions included within each module were carefully chosen to collect the minimum amount of information needed to answer the research questions specified in the evaluation's PWS and/or to directly reflect priorities expressed by ED and IES staff. In early rounds of revisions to the instruments, the study team made substantial cuts to the content of the instruments to reduce burden on respondents.

The RTT-SIG state (and district) data collection instrument was designed for administration as a telephone interview, with experienced interviewers guiding respondents through the questions for their designated topic area(s). The interview format is expected to maximize response rates, minimize respondent burden, and reduce the need for time-consuming follow-up, by enabling interviewers to address respondent questions about the terms used in the instrument, the wording of particular questions or response categories, or other issues. This was deemed important because the policies and reforms of interest are complex and relatively new. Hence, there is no widely accepted terminology to refer to or describe them. Our pilot-tests of the RTT-SIG instruments confirmed that respondents occasionally had questions about the terms being used in the interview and/or the category that best fit their state (or district). The ability to answer respondent questions “on the spot” helped ensure that the desired information was collected without requiring extensive follow-up to obtain answers to skipped questions or resolve logical inconsistencies in their responses. Overall, because of the complexity and breadth of data we are collecting in this evaluation, we feel that the interview format offers the best combination of ensuring high response rates, high quality data that is complete and consistent, and limiting respondent burden.

5. Please explain how the module system works in the case of these surveys and whether different modules will be answered by different people. Who determines who answers these modules? It seems like the reliability of this data across the states depends on making sure the right people are answering the right questions.

As noted in our response to question 4, different modules may be answered by different respondents. During the recruiting process, the study team asked our contacts for the names of the individuals in the state (and in the district) who are in the best position to answer the types of questions included in each module. Prior to conducting the interview, interviewers will confirm that the previously-identified respondent is still the best person to address the questions. If needed, the interviewer will send the protocols to help the contact determine who should participate in the interview. To ensure that the study team collects the most complete and accurate data, the study team will share the appropriate module(s) with the designated respondent ahead of the actual interview, highlighting questions that may require the respondent to look up information or check with others in their state (or district) if they do not know the answer, and encouraging the respondents to do so ahead of the interview. If the respondent does not have some of the information sought during the interview, the study team will follow-up to collect these additional details later.

In contrast, we expect the principal (or another school administrator) to complete the entire school survey. In this case, the principal will be encouraged to consult with other school staff as needed. Because the school survey is considerably shorter than the district and state interviews, and also covers a more limited set of topics than the interviews, we do not anticipate principals and other school administrators having major problems with answering all of the questions on the survey. This expectation was confirmed by the results of the pilot testing.

6. Why do the different surveys use different historic school years as (I assume) the basis for comparison? In the state its 2007-8, while in the district its 2009-10.

Different years are used across the instruments to ensure that the pre-intervention period is appropriate for the particular survey. The state survey, which is the data collection instrument that focuses on the evaluation's RTT component, uses 2007-2008 as the baseline period, because that is the school year before ARRA was signed into law in February 2009. The first round of RTT grants were awarded in March 2010, so we considered using the 2008-2009 school year as the baseline, but ultimately decided that using the 2007-2008 school year would be the most conservative approach to take, because the February 2009 ARRA legislation laid out the reform areas of focus for RTT, so after February 2009 states may have already started to react to and make changes in response to the legislation. Therefore, we decided that questions about the 2007-2008 school year would be most likely to capture the true pre-intervention (i.e., pre-RTT) conditions in states.

In contrast, the district interview and school survey are focused on the evaluation's STM component and a sample of districts that received SIG grants to implement STMs in some of their schools. These awards were generally made in the summer and fall of 2010, so the relevant school year to reflect the pre-intervention period for these districts and schools is the 2009-2010 school year.

State Survey Instrument (how is this "semi-structured"?). Note that most of the comments are related to the fact that this is apparently being done over the phone. The burden estimate says this should take 4.5 hours. Why isn't this being done on the web?? It would save time and money.

Please see above responses to questions 1 and 4.

SC3: They should consider making the answer categories easier to recall, particularly if this is over the phone. How about "To what extent (a lot, some, a little, none at all)..." Also, this is a pretty long list of constructs they are testing. Did all 12 of these emerge from some qualitative work? Same with other questions that use this scale (i.e. SC5, TL48)

The response categories for question SC3 (and all similar questions, like SC5 and TL58) were changed along the lines suggested above – to great extent, moderate extent, little extent, or not at all – based on pilot test results.

Before the study team conducts interviews, we will send all respondents a copy of the appropriate module(s) from the instrument ahead of the interview (see response to question 5 above). Respondents will be able to follow along with their copies of the appropriate module(s) during the telephone interview. During pilot tests of the instruments, we used this approach and respondents reported that having copies of the instrument was helpful.

The list of constructs included in this question directly reflects the areas of focus outlined in the RTT application for states. The items about supports and/or reforms related to English Language Learners were included in response to requests from the Office of English Language Acquisition, who provided some evaluation funds to support this English Language Learners focus.

SC4: How do they plan on asking this question over phone?? Are they expecting the respondents to keep all 7 of these options in their minds and then order them? Without something to look at, recall or ranking tasks should be kept to around 5 options or less. This would be avoided if this survey was over the internet/on paper!

See response to comments about state interview question SC3.

SC6: The wording of this prompt is awkward.

We have revised the wording of the question to read, "To what extent does the state education agency play each of the following roles (a great extent, moderate extent, little extent, or not at all)?"

SC7: Is the for example needed? If so, can it be a follow-up by the interviewer if the respondent needs help? Reading as it is takes the better part of a minute now.

We think it is important to provide an example to ensure that the respondent understands the question. However, we have streamlined the example to reduce the length of time needed to read it.

SC9-11: Why isn't this set parallel? I take it we're asking for the last year and the 2007-2008 year to see what changes are a result of the \$, but why not ask if the state implemented actions in response to 2007 to 2008 reporting? As it is now, you couldn't say that the implementation was due to the money if you find that monitoring did not change over time.

To make the questions parallel, we added a question (SC12) to ask about actions in response to 2007-2008 reporting.

SC17 &18: Are these to be asked of all states or just RTT states?

These questions will be asked of all states (see response to question 2). We need data collection to be consistent across RTT and non-RTT states so that we can make comparisons between the two groups of states. For example, comparing RTT and non-RTT states on SC18 will show whether RTT states were more likely to implement any of the listed changes. We edited the stems for both questions (adding "if applicable") to clarify that all states will be asked to respond to the questions. (These questions are SC18 and SC19 in the current version of the instrument.)

SC19: Same issue as SC4, but even longer. Also, are the open-ended prompts just for the top three or do you expect the respondents to first fill in the blanks and then rank?

Please see response to comment on state interview question SC3. The open ended-prompts are only for the top three barriers selected by the respondent. Interviewers will ask the respondent to select the top three barriers and then will probe for additional details on the response options selected. Pilot test respondents were able to readily respond to this question. (This question is now SC20 in the current version of the instrument.)

SA3: Will the respondents have this data (percent of curriculum standard that conform to the common core) in front of them? If this isn't something they will explicitly know, you'll want to instruct them to look it up beforehand in the intro letter.

As noted in our response to question 5 above, before the study team conducts interviews, we will send all respondents a copy of the relevant module(s) of the instrument. The copies sent to the respondents will highlight specific questions that we think may require looking up data in advance. During pilot tests of the instruments, we used this approach and respondents reported that having copies of the instrument and highlighting specific questions that may require looking up data in advance was helpful.

DA3/4: How are they sequencing these questions on the phone? Are they running through all DA3 then asking DA4 for all that code as yes (1)?

We plan to ask these questions across rows. For example, interviewers will first ask about whether the group currently has access to the data from the state longitudinal data system. If the response is "yes," the interviewer will then ask what type of access the user group has. If a "no" is recorded for whether a group has access to data from the state longitudinal data system, interviewers will skip to the next group. This approach worked well during our pilot tests.

DA6: Same as with SC7...save the ex/ if they have questions. Can the last sentence in the prompt be used just as a follow-up if their answer codes as 3?

We have deleted the example from the prompt and have designated the example text as a prompt for interviewers to use only if respondents ask for help. We also streamlined the text for the question stem.

DA9: Same as SA3. Let them know you'll be asking for this specific info

Please see response to comment on state interview question SA3.

DA14: Save example for respondent follow-up

We have deleted the example from the prompt and have designated the example text as a prompt for interviewers to use only if respondents ask for help.

DA15: Same issue as SC19, but even longer still! How are respondents supposed to keep these substantive statements in their heads?

Please see response to comment on state interview question SC3.

TL24: Recall issues here too.

Please see response to comment on state interview question SC3. (This question is TL35 in the current version of the instrument.)

TL27: Save example for respondent follow-up

We think it is important to provide an example to ensure that the respondent understands our intended meaning of “extent” in the context of this question. However, we have streamlined the example text to reduce the length of time needed to read it. (This question is TL38 in the current version of the instrument.)

TL37: Save example for respondent follow-up

We think it is important to provide an example to ensure that the respondent understands the question and what we mean by a “spread” of teacher rating categories. However, we have streamlined the wording of the example to reduce the length of time needed to read it. (This question is TL47 in the current version of the instrument.)

TL38: Recall issues

Please see response to comment on state interview question SC3. (This question is TL48 in the current version of the instrument.)

TL45: It seems like all you really care about here is the standardized test subjects. Might as well just phrase the question that way.

In response to feedback during the pilot testing of the instrument, we streamlined the wording on this question, but retained the reference to “tested” (grades and/or) subjects (and “non-tested” grades and/or subjects in question TL56) because that is the standard wording used to refer to these different types of teachers. (This question is TL55 in the current version of the instrument.)

TA7, 41, CH2, 7, 8, 17, 19, 20: Let them know you’ll be asking for this specific info

Please see response to question 5 above.

State-District Survey Instrument

SA1: I’m not sure why this is being asked of the district representatives. If it’s to see whether or not the standards have been implemented due to the funds, why not just start with SA2 for all districts and then use a skip pattern?

We agree with this comment. We have deleted this question.

SA3, TA2, 6, 30, 31, 32, 33, 36, 39: Make sure to let them know in advance that you’ll be asking for this information.

Please see response to comment on state interview question SA3. Note that we assume TA35 (now TA31) should be highlighted in advance, not TA36 (now TA32).

D: Have you considered asking all of the questions that relate directly to the list of schools sent beforehand all at once? There might be some cog. issues switching from one frame to another.

During pilot tests of the instrument, we learned that these questions were in fact problematic for respondents to answer (since practices and policies can vary across schools) and have subsequently dropped them from the instrument. The revised district instrument now only asks about district policies that are likely to be uniform across the district. Information that may vary across schools within the same district (e.g., the types of support received from the district) is now only collected in the school survey.

DA1: There's a recall issue here if you list them and then ask them to say which ones they use. Why aren't you asking this as a series of yes/no questions?

We have changed the format of the question to code one response per row (Yes/No/Don't Know/Refused/Not Applicable).

DA13: Recall issues

Please see response to comment on state interview question SC3. (This question is DA9 in the current version of the instrument.)

TL4: Reserve example for respondent follow-up

We think it is important to provide an example to ensure that the respondent understands our intended meaning of "extent" in the context of this question. However, we have streamlined the text to reduce the length of time needed to read it.

School Survey

Why do they need to enter a start/end time?

Respondents do not need to enter a start/end time. That was only for the purposes of the pilot testing of the instrument to gauge the length of the instrument. The start/end time question has now been deleted.

DA1: For longer questions like this one (12 items), consider breaking them up on different pages to reduce respondent fatigue.

We will investigate the feasibility of breaking this question up across multiple screens in the web survey and implement this suggestion if feasible.

DA3: Is this the format the respondent will see? Can the frequency pop up if the code is 1?

The respondent will not see the whole table on the web survey. The frequency questions will pop up only if the respondent answers "yes" for whether the activity occurred in the school.

DA4: Are the respondents going to be responsible for the skips, or will they be automatic?

The skips will be automatic.