Response to OMB comments on the MTSS-B Study
September 21, 2015

We appreciate the feedback on the OMB package. In this memo, the MTSS-B study team lists the comments made by OMB and our response. We also conducted an additional review of the package, partially prompted by the OMB comments. We note other very minor changes that we have made to the supporting statements in response to this review.

Document: OMB package Part A		
Page# / Paragraph#	**OMB Comment**	**Study Team Response**
Pg. 6, third bullet	Is there a plan for choosing other subgroups besides at-risk students that the study will examine? Have the districts and schools been chosen with other particular subgroups of interest in mind? *(Repeated in email)*	MTSS-B is a practice that is implemented widely across the country. As a result, finding districts that were not currently implementing a systematic MTSS-B program was the priority during site recruitment since service contrast is essential for a fair test of the program. Because of the policy interest in improving behavior and academic outcomes for disadvantaged students, priority was given to districts serving large proportions of students who qualify for Free or Reduced priced lunch. Because student behavior is such a central outcome of the study, we powered the study to investigate a high-risk sub-group based on baseline teacher ratings. We also plan to conduct exploratory sub-group analyses based on the baseline academic achievement of the upper grade students (because we will be able to collect state testing data for these students). We also plan to conduct subgroup analyses based on gender, race and ethnicity and family income (based on FRPL status, if possible to consistently collect). Supporting statement B provides more detail on the guidelines used to recruit the districts and schools in the study.
Pg. 7, paragraph 5	Does this timeline need to be updated?	The timeline does not need to be updated.
Pg. 11, paragraph 1	How were the sample sizes of districts, schools, and students determined? Are sample sizes appropriate for the hypothesized impact on school staff practices, school climate, and student outcomes? *(Repeated in email)*	We worked with IES to set the target minimum detectable effect size (MDES) based on reasonable and policy relevant expected impacts on various types of outcomes for this kind of intervention. We then calculated the required sample size for the target MDES, using parameter assumptions based on a review of existing literature. For academic achievement, the estimated MDES for the year 2 estimates are 0.189 for reading and 0.195 for math. The estimated MDES for behavior ratings in year two is 0.089 for the high-risk subgroup and 0.069 for the random sample of all students. The estimated MDES for teacher practice outcomes range from 0.118 to 0.201 depending on the parameter value assumptions and number of classrooms observed. The estimated MDES for school climate measures from the teacher survey in year 2 is between 0.186 and 0.410 depending on the parameter value assumptions. The estimated MDES for the school climate measures from the student

		survey (grades 4&5) is between 0.149 and 0.285 depending on the parameter value assumptions. We have added this language into Supporting Statement A (see text on page 11 and footnote #7).
Pg. 14	This sounds like a lot of burden is being placed on schools and teachers to participate in the study. The teacher ratings, in particular, seem lengthy, especially when added to the surveys.	The proposed activities are essential for the measurement of the key outcomes of interest—teacher practice, school climate, student behavior and academic achievement. Each data source contributes in a unique way to the measurement of these outcomes. For example, the student survey is the only measure of student's perception of classroom climate and the staff survey is the only measure of staff's perception of school climate. Student behavior is the key outcome of this study. Administrative records regarding student behavior are limited to suspensions and expulsions. Elementary schools have relatively low rates of these behaviors. While office referrals are more common, they are not systematically collected in, nor are they available from, all districts. As a result, the study team must engage in original data collection to measure this outcome. The teacher rating instrument we are using has been slightly adapted from an instrument used in prior studies of MTSS-B where it was found to be sensitive to the intervention (e.g. Bradshaw et al., 2012). The instrument measures key domains of student behavior closely tied to the theory of change of MTSS-B —prosocial behavior, concentration problems, disruptive behavior, internalization, emotional regulation and bullying. Additionally, the ratings will allow us to assess the intervention's impact on schools' use of disciplinary and behavior support practices. The teacher ratings instrument is the most reliable and sensitive way to collect useful student behavior information. The most burdensome measure is the teacher ratings of student behavior. If a teacher has twenty consenting students in her class, she may need to spend up to 100 minutes (or 1 hour and 40 minutes) completing this rating form in Fall of 2015. In the Spring of 2016 and 2017, we are only asking teachers to rate a sample of their students and so the teachers will only need to spend up to 40 minutes on ratings each spring. If district rules allow, we are compensating teachers for this time with an incentive based upon the number of ratings they complete (up to $50). Teachers will also be asked to participate in a web-based staff survey that will take no more than 35 minutes to complete. To reduce burden, we significantly shortened the survey by cutting scales not as closely tied to the MTSS-B theory of action and by shortening existing scales with guidance from experts in the field. Additionally, when we ask teachers to complete this survey in the spring of 2017, we are only asking them to rate a sample of their students on the teacher rating of student behavior.

		The description of the data collection instruments and the teacher rating form has been edited in the package to provide a more clear justification (see page 11 and 16).
Pg. 16, paragraph 1	Do you have an estimate for how long this will take teachers, in total?	The ratings are estimated to take 5 minutes per student to complete. This estimate is based on the team's experience fielding similar measures in other studies as well as prior use of this survey in Catherine Bradshaw's studies of MTSS-B. If a teacher has twenty consenting students in her class, she may need to spend up to 100 minutes completing this rating form in Fall of 2015. In the Spring of 2016 and 2017, we are only asking teachers to rate a sample of their students and so the teachers will only need to spend up to 40 minutes on ratings each Spring.
Pg. 27, paragraph 2	What is the randomization procedure? It seems like districts are responsible for randomly assigning their schools into treatment and control groups, is that right? Do the researchers plan to assist or supervise districts in the randomization? *(Repeated in email)*	Districts were recruited for this study and worked with the study team to recruit schools with the appropriate service contrast. After the district and schools had agreed to participate, the study team conducted the random assignment procedures. The multi-stage RA process randomly assigned schools within same school district or random assignment blocks within district to the treatment condition and the control condition with roughly the same probability. After randomization, the district and schools were informed of the results. We have revised footnote #35 to clarify.

Additional study team changes: We have revised the language regarding payments to teachers and schools in the section about teacher ratings to indicate that teachers will only be compensated if they complete all the ratings for their students. We are concerned that sharing with teachers that they will be compensated if they rate 85% will create an incentive to only complete 85% of ratings.

Document: Supporting Statement B
Introduction revised to reflect changes made to Supporting Statement A

Document: Appendix A, Site Visit Principal Interview Procedure

No OMB Comments

Additional study team changes: Minor edits to have been made to the interview protocol as we prepare the protocol for fielding. MDRC's IRB and the districts have confirmed that we do not need to collect written consent for this interview because respondents are merely describing the behavior support practices at their school. Additionally, to preserve resources the team has decided not to audio record the interviews since coding happens during the site visit. We have revised the assent procedures accordingly and deleted the consent form.

Document: Appendix B, Site Visit Behavior Team Leader Interview Procedure		
Page# / Question#	OMB Comment	Study Team Response
Pg. 5, Q 8	Using the generic term "interventions" here and above is confusing to me. Can you use more specific language in this prompt and/or the one above?	The site visitors will be trained to provide the respondent with a definition of the tern "intervention" (e.g program or strategy) during the interview if the respondent is confused by the term.

Additional study team changes: Minor edits to have been made to the interview protocol as we prepare the protocol for fielding. Also, MDRC's IRB and the districts have confirmed that we do not need to collect written consent for this interview because respondents are merely describing the behavior support practices at their school. Additionally, to preserve resources the team has decided not to audio record the interviews since coding happens during the site visit. We have revised the assent procedures accordingly and deleted the consent form.

Document: Appendix C, Site Visit Student Interview Protocol

No OMB Comments

Additional study team changes: Minor edits were made to the wording of the second question

Document: Appendix D, Site Visit Staff Interview Protocol

No OMB Comments

Additional study team changes: Minor edits were made to the protocol

Document: Appendix E, Phone Interview MTSS-Coach

OMB suggested some copy edits, which have been accepted

Document: Appendix F, Phone Interview Principal and Behavior Team Leader

No changes made

Document: Appendix G, Staff Teacher Survey		
Page# / Question#	OMB Comment	Study Team Response
Pg. 3, Q A2	What are we getting at here? Is this the best way to provide a third option?	We have deleted this option.
Pg. 6, Q B1	Is there a reason why B1, B2 and B3 are broken out separately? Each section uses the same prompt. Are there too many sub questions to fit on one page of the survey if you combine this into one table?	This will be a web-based survey and so each table indicates the questions we anticipate including on each screen. In our experience, it is best to limit the number of items on each screen.
Pg. 7, Q B4	Is there any concern that for B4 and B5, most of the statements are negative? Will teachers read that to	B4 and B5 measure distinct conceptsB4 measures emotional exhaustion (demoralization and disaffection whereas B5 measures work stress. The burnout scale was adapted/shortened by her from the Maslach Burnout Inventory. The work stress scale was adapted

	mean that we expect them to feel badly about their stress levels?	from the NIOSH Generic Job Stress Questionnaire. The scales have been or are being used in prior studies of educational interventions including the IES funded randomized control trial of training in the Good Behavior Game. No concerns regarding the tone/negativity of the items have been noted in their use. However, in response to the concerns about the number of items, the team has deleted two items from the burnout scale (B4) and items that could potentially be the most likely to provoke the most negative response due to their wording.
	These seem really similar to me, why do we need all of them?	
Pg. 9, Q B7	Are we asking for school-wide, consistent systems across all classrooms, or things in the particular teacher's classroom?	We agree that the wording is potentially confusing. To clarify, we have adjusted the introduction to note that we are referring to a teacher's classroom and adjusted the skip patterns to make clear that non-classroom teachers should not answer the question.
Pg. 12, Q C5	Aren't we calling this the positive behavior game? Will this be confusing to the teachers?	We wanted to include "Good Behavior Game" as a possible response in case it is being used in BAU classrooms. However, we agree the teachers in program schools might be confused. As a result, we will give teachers the opportunity to select both and list Positive Behavior Game as a separate response category.
Document: Appendix H, Student Survey		
Page# / Question#	OMB Comment	Study Team Response
Pg. 1, Q A1	For teachers, we provide a third option, "not sure." I don't know if a third option is appropriate here, just flagging.	We agree a third option is not appropriate here. It has been deleted
Pg. 2, Q E1	Do you intend for this to be ambiguous as to whether this means physical or other types of hurting?	These items were taken directly from the student survey fielded by Catherine Bradshaw and colleagues for the Maryland Safe and Supportive School Initiative. The high school version of this survey has been validated in prior studies and we are inclined to keep this specific language as it appears in the original (Bradshaw et al., 2014). Also, we are okay with the respondent interpreting the item as reflecting both physical and other types of hurting.
Pg. 2, Q G1	All of the other items in this section are positives, things we would want to see, while this is something we would not want to see. Will it confuse children to answer these questions side-by-side?	The team has fielded scales with a mixture of positive and negative items in prior studies including the Maryland Safe and Supportive Schools Initiative. Students have not reported confusion from these items. We also think that including a mixture of positive and negatively stated items in the same scale will help reduce the problem of "response sets."
Pg. 3, Q I1	Is this language too sophisticated for elementary school children?	The language was simplified to better reflect the age of the respondents. This definition of bullying is consistent with the definition offered by the Center for Disease Control and was not problematic in the Maryland study that was fielded among 4th and 5th graders.
Document: Appendix I, Parent Consent form		
Additional study team changes: The parent informed consent form has been edited for clarity. MDRC's IRB has approved this template.		

Document: Appendix J, Teacher Ratings		
Page# / Question#	OMB Comment	Study Team Response
Pg. 3, Q A2	Why are these numbers (letters e and f) backwards? I understand that these are positive statements, but so is letter a, which is not backwards.	Thank you for highlighting this. We have changed the numbering so it is consistent throughout.
Pg. 3, Q A3	Same as above.	See above response.
Pg. 4, Q A4	Same as above.	See above response.
Pg. 4, Q A5	Same as above.	See above response.
Pg. 5, Q A6	Same as above.	See above response.
Additional study team changes: We have adjusted some of the wording and formatting to make questions more clear (e.g. changed items to past tense).		
Document: Appendix K, District Data Request		
Copy edits offered by OMB have been accepted. The variables list and letter have been edited		
Document: Appendix L, IDEA Excerpt		
No changes made		