Appendix G

Description of Factor Analysis and Principal Components Analysis

PRR used Factor Analysis and Principle Components Analysis to determine how to reduce the number of items for the TRB survey. Both of these analyses explore how similar each item is based on the way people responded. Many of the analysis steps are the same for factor analysis and principle components analysis; the key difference is the interpretation of the correlation between items and patterns for how people respond. As a result, each analysis offered nuanced insight into how to shorten the instrument.

Terms Guide

* A **survey** is the questionnaire used to collect data. It includes non-numeric data (e.g., project name).
* An **item** is an individual question on the survey. For example, “I understood the benefits of the project”.
* An **indicator** is a group of items (e.g., aggregate three items about timeliness for the “Timing” indicator).
* An **index** is a group of indicators. An index aggregates numeric indicator scores, only.

Factor Analysis

PRR conducted a Factor Analysis to describe underlying patterns of how people answered survey items so that we could simplify the overall survey.

Any time we conduct a survey, we can expect certain patterns to the way people answer. For example, we know that when people skip items it is usually the last few items. People are not likely to skip the first few items and then finish the rest of the survey. These patterns show up when we analyze the survey statistically. If an item is difficult or unclear, this can also show up as a pattern. At times, things we’re not directly asking about can also affect how people respond (e.g., people who are depressed may answer a survey in a different way than people who aren’t).

Factor Analysis is one approach to understanding hard-to-see patterns like these in survey data. We used Factor Analysis to see if we could use correlations to describe the different patterns in how people answered the items on the survey and then determine whether items are unnecessary based on these correlations.

First, we looked at the strongest correlations between all of the items and a specific pattern. Each pattern had strong correlations with 3-8 individual items. Each time a pattern had strongly correlated items, we considered them a “factor”. We repeated the process over and over again until every item belonged to a factor. For example, many people answered the items about benefits the same way (pattern) for five questions. After accounting for these questions, we noticed that many people answered items about timing/timeliness the same way for six items. We kept going until all items belonged to a related pattern.

Because the factors were related to each other (e.g., how satisfied people are with timeliness affects how satisfied they are with their ability to influence the project), we made two types of graphs to visually explore the factors. One was a graph we called “orthogonal” because it did not account for how correlated the factors were with each other. As expected, this graph was inadequate at giving clear direction on which items to keep. The other graph we called “oblique” because it did account for the correlations between factors. As expected, the graph that accounted for how similar these factors were produced easier-to-understand results. We used these graphs for subsequent analyses.

Next, we counted the number of factors and explored how many groups of items (i.e., indicators) are reasonable to expect for this survey. We chose five factors or patterns of responses based on the graphs and correlations between items. To determine which items to cut, we considered the unique contribution of each item to the overall survey. For that purpose, we explored how correlated each item was with all the other items in the survey. Last, we made a list of all the items that did not appear in any group. We made a note of these items and compared the results of the factor analysis to determine if the subsequent Principle Components Analysis recommended cutting these items.

Principle Components Analysis

PRR conducted a Principle Components Analysis (PCA) to determine which items were so similar to others that they could be cut from the survey without losing too much information, as well as which items could be grouped into indicators for the scoring tool.

First, we used Principle Components Analysis to see how correlated each item was with other groups of items. It is similar to Factor Analysis, but instead of exploring the correlation between items and patterns of responses, PCA gives us evidence for which groups of items belong together. The result of the PCA can be used to determine how many groups of items (i.e., components) an index might have. The components themselves are evidence for which items belong to which indicators.

First, we looked at the strongest correlations between all of the items. Typically, each item was correlated with 3-6 other items. Each time items are strongly correlated with each other, we consider them a “component”. After we had our first component, we repeated the process over and over again until every item belonged to a component. For example, responses about if the process lasted long enough, if materials were given far enough in advance, and if the public had enough time to understand the information were all correlated with each other. As a result, the PCA considered these three items a component, namely Timing.

Because the components were related to each other, we made two types of graphs to visually explore the components. One was a graph we called “orthogonal” because it did not account for how correlated the components were with each other. As expected, this graph was inadequate at giving clear direction on which items to keep. The other graph we called “oblique” because this one did account for the correlations between factors. This allows us to consider how scores for *Timing* are related to scores for *Transparency and Clarity*. As expected, the graph that accounted for how similar these components are produced easier-to-understand results. We used these graphs for subsequent analyses.

Next, we counted the number of components and explored how many groups of items are reasonable to expect for this survey. We chose 6 components or indicators based on the graphs and correlations between items. To determine which items to cut, we considered the unique contribution of each item to the overall survey and explored how correlated each item was with all the other items in the survey.

Using the Results of the Factor Analysis and Principle Components Analysis

In making final decisions of what to cut, we considered the Factor Analysis and the results of the PCA. We dropped two kinds of items— items that were outliers and items that were redundant. Outlier items were not highly correlated with other groups of items. They added little value to the final survey. Last, we dropped any item that was highly correlated with other items but did not individually add much to the survey. These items were too similar, so removing one could shorten the survey without changing the results.

The full list of items from both the Factor Analysis and the Principle Components Analysis were used to provide insight into which items could be removed from the survey without losing important information for measuring the effectiveness of public involvement. For both analyses, the final survey was compared to the initial (longer) survey and the results were similar enough to feel confident that the shorter version would work well.