Appendix: Intervention Protocol

**Phase 1. Design and Interpretation of Experiments**

Here are the applicants’ records on cards. (*Facilitator displays cards and explains how to read a record. Each card shows background information about this applicant with a blank space for the applicant’s performance rating in the simulator to be recorded.*) Studying the records carefully, you will be able to see which factors make a difference to performance and which don’t.

You and your partner will share the records, and one rule is that everything you do, you have to agree to first. (*Facilitator is to ensure that pairs are discussing their actions and decisions with each other and that each student is paying attention to what the partner says.*)

You’ll do best if you investigate one factor at a time. Which factor do you want to start with? You can start with one you think is going to make a lot of difference, but eventually you’ll investigate them all.

(*Pair discusses and responds*.)

(*Facilitator displays phase 1 factors list on a stand and puts an arrow post-it on the factor pair has selected for investigation. As needed, she points to the stand to remind pairs which factor they are investigating*.)

(*For the first factor only*) Great, now choose one or two records you want to look at first. If you want to find a particular case, let me know and I can help you find it. Remember to discuss and agree before you make your choices. Once you have decided on the records to look at, I will go to the database to find the performance outcome for that applicant for you. You can record it on that record card. (*Pair is shown where to write down the outcome on the record card.*)

(*Pair discusses and agrees on choice of record card(s) from the set.)*

A. (***If pair chooses only one record***) What are you going to find out from this record? (Pair responds.)

(Facilitator provides the outcome for the chosen record. Pair records the outcome on the record.)

What did you find out? (*Pair responds.*)

Can you conclude whether [factor being investigated] makes a difference to the performance in the simulator? (Pair responds.)

1. (***If pair says yes, they can tell whether factor makes a difference***) What will happen to the outcome for this applicant if [factor being investigated] goes up? (*Pair responds based on belief.*)

Do you know for sure? Why don’t you test this out to be sure? What cards would you need to test this out?

2. (***If pair says no***) What cards would you need in order to be able to find out whether this factor makes a difference? (*Pair responds.)*

(*If pair does not suggest finding another record to compare*) What will happen to the outcome for this applicant if [factor being investigated] goes up? (*Pair responds based on belief.)*

Do you know for sure? Why don’t you test this out to be sure? What cards would you need to test this out?

B. (***If pair chooses two records***) What are you going to find out by comparing these two records? (*Expected response: whether X makes a difference to the outcome.*)

(*If pair answers anything else, guide pair to only think about finding out whether* [factor being investigated] *makes a difference to the outcome.*)

(*Provide pair the outcome for the chosen record. Pair records the outcome on the record.*)

(*Pair examines the two records side-by-side, with outcomes shown.*)

What did you find out? (*Pair responds.*)

(*Facilitator to repeat pair’s claim so that all parties are clear about the claim the pair is making.*)

1. (***For causal claim, if not controlled***) That may be true. But can we really tell for sure that [factor being investigated] makes a difference to the performance? (*Point to another factor on one of the cards*) Couldn’t someone say that it is because applicant A has [certain level] for [another factor] and that is why applicant A has a better grade than applicant B? Do you really know for sure that [factor being investigated] made the difference?

a. (***If pair says, “no, you cannot be sure”***) How can we be 100% sure? Is there a better record to compare applicant A’s record to? (*Pair chooses a new record to compare with A, and is provided with outcome information.)*

b. (***If pair says, “yes, you can be sure”****,**ask the* ***Fallback Questions*** *at the end of this section.)*

2. (***For non-causal claim, if not controlled***) That may be true. But [factor being investigated] and [another factor] for the two cards are different but the outcome is the same. Can you really know why their outcomes are the same? Maybe one makes the outcome go up and the other makes it go down so they offset each other? Can you really be sure that [factor being investigated] does not make a difference? (*Pair responds.*)

a. (***If pair says, “no, you cannot be sure”***) Is there a better record to compare it to so that we can be 100% sure? What cards would you need to see instead so that you can be sure?

b. (***If pair says, “yes, you can be sure”***, *ask the* ***Fallback Questions*** *at the end of this section.*)

3. (***If records were controlled***) What will outcome be when [factor being investigated] changes from one level to another? (*Pair responds. If necessary facilitator reviews what it means that a factor makes a difference.*)

Make an argument to the foundation of how you’re sure [factor being investigated] is/isn’t important. (*Pair responds.*)

But couldn’t I say that applicant A has [certain level] for [another factor] and that is why applicant A has a better grade than applicant B? (*Expected response: no, because applicant B also has the same level for [the other factor] as applicant A. So it cannot be [the other factor] that is making the difference in performance. If student does not provide the expected answer, guide the partner to challenge the student’s response. If both students fail to note that [the other factor] was the same for both records, facilitator asks this question again when investigating the next factor and points specifically to the records.*)

Suppose someone disagrees with you and doesn’t think that [factor being investigated] does/does not make a difference; what would you say to them to convince them? (*Pair responds. Eventually, expected response: show them the cards.)*

Let’s write down what you found out here as a memo to the foundation. (*Gives pair a memo form*)

(*Once pair has successfully controlled for one or two factors*) What if we change your comparison so that [another factor] also differs. Can you still use this comparison to show that [factor being investigated] makes a difference to the outcome? (*Pair responds.*)

Why is this comparison not convincing? (*Pair responds.*)

Now choose another factor to investigate.

4. (***Fallback Questions for if, after a few tries, pairs still do not choose controlled records***) What do you think would happen to applicant A’s performance, if, for [factor being investigated], she has a different level? (*Pair makes a guess of performance.*)

Should we find out what the records show? Which record do we need? (*Pair responds*.)

a. (***If pair still does not choose a controlled record***) What would applicant A’s record look like if she has a different level for [factor being investigated]? (*Pair responds.*)

Let’s look for that card so that we know what happens to applicant A’s performance if she has a different level for [factor being investigated].

**Phase 2. Multivariable Coordination**

So far our conclusions are all based on just two cases. We would be more sure of our conclusions if we looked at more cases. We have a way to do that. (*Facilitator introduces InspireData chart. She discusses how to read a chart, showing how a record is represented by one diamond by showing the details of a diamond and highlighting the fact that the information is the same as one record card in Phase 1.*)

Here, we collected a lot of records. Look just at the “average” level for fitness. Why is there a range of different outcomes, even though everyone is average on fitness? (*Expected response: Because other factors also matter as well. Facilitator shows the details of some records with different performance levels and points out that, except for fitness, the records have different levels for other factors.*)

So, all these other factors may also make the outcome change. That’s why we collected lots of cases and we have the same number of records for all different levels. For example, we have the same number of records of people who have no college, some college and college education. Let’s see if you come to the same conclusion as you did earlier now that you are looking at more results.

Which factor do you want to look at first using the charts? (*Pair responds*.)

(*Shows chart for* [factor being investigated]*.)*

So, remember, other things may be contributing as well. But can we say OVERALL that [factor being investigated] makes a difference?

A. (***If pair refers to beliefs or Phase I comparisons***) Remember that the organization wants to be really sure before we make any decisions. So we don’t want to rely on opinions/just two cases.

What does the chart say about whether the factor matters or not to the performance?

B. (***If pair is unsure how to tell if [factor being investigated]*** ***makes a difference***) Let’s look at the chart together. What do you notice about the performance levels when the applicants have [a certain level, e.g. average] and what about when they have [a different level, e.g., excellent] for [factor being investigated]? (*Pair responds.*)

Do the applicants have different or similar kinds of performance? What does that tell you about whether [factor being investigated] makes a difference to the performance?

C. (***If pair correctly compares the distributions***) Make an argument to the foundation of how you are sure [factor being investigated] is/isn’t important. (*Correct response: by comparing the 2 distributions*)

Suppose someone disagrees with you, what would you say to them to convince them? (*Pair responds. Expected response: show them the chart.*)

How would the graph look different if [factor being investigated] makes no difference to how well people do?

Let’s write down what you found out here as a memo to the foundation. (*Give pair a memo form.)*

Now choose another factor to investigate.

(*Once all factors have been examined*) Now, let’s make a summary for what you have figured out. (*Facilitator assists as needed as pair completes summary memo.*)

*Now, what if the director told you that students from another school wrote a memo that says they found that the applicants’ age makes a difference to the performance in space and asks you what you think about that. What would you say to her?*

*(****If pair says something about whether age does or does not make a difference****) How do you know? (Pairs give some opinions.) Do you know for sure? [The purpose here is to help them realize that they actually have no basis for evaluating that claim, which leads to the realization that the other person can possibly also be right.]*

*(****If pair says that but we found that the other things make a difference****) Can only one thing make a difference? Can someone else discover that other things also make a difference?*

*​*

**Phase 3. Prediction**

You’ve worked hard to figure out what factors mattered to the applicants’ performance. Now, you can use your knowledge to predict how well they will do. I have some new applicants here. Can you predict how well they will do? Then you can choose a best team of five.

What information about the applicant would you need to predict his/her performance? I can only tell you up to four things about the applicant. Which factors do you want to know about? Look at the list of factors here; what information about them would you like to have? You can make use of the summary sheet or refer to any of the charts anytime. (*Expected response: all causal factors. If not all causal factors are requested, urge pairs to review each chart and if necessary, repeat phase 2 protocols.*)

Here is the information we have on this applicant. (*Facilitator provides data for all requested causal factors and, unless pair chose a non-causal factor, also include Home Climate*, *a non-causal factor, for the applicant*)

Now you can predict. Predict how each one will do, and explain why you made that prediction. Be sure to discuss with your partner before making a final decision.

(*Applicant description sheets are presented, one at a time, with charts handy so that pair can refer to them when needed.)*

*(For the first three predictions, facilitator asks, “Which factors mattered to you prediction?”)*

A. (***If pair selects a non-causal factor as influencing the prediction***) What did you find out about [the non-effective factor]? Did it make a difference to the outcome?

1. (***If pair answers yes***) How did you know? (*Pair responds based on belief*)

Do you know for sure? Let’s look back at the chart?

2. (***If pair answers no***) When you predict how well someone will do, will it help you to know whether they have [a certain level] or [another level] on this factor?

a. (***If pair answers yes***) Do you know for sure? Why don’t you check out the chart?

(*Show chart of the non-causal factor*)

What do you think will happen to the applicant’s performance if for [the non-effective factor], they go from having [a certain level] to [another level]? (*Pair responds.*)

i. (***If pair answers no change***) Do you still need to know about this factor to make your prediction?

ii. (***Otherwise***) Let’s find out. (*Points to chart*) Does it matter to the performance whether [the non-effective factor] is [a certain level] or [another level]? (*Pair responds.*)

Do you still want to know about this factor to make your prediction?

B. (***If pair does not select one or more effective factors as influencing the prediction***) What made you predict this performance level? (*Pair response: because [a causal factor] is at a certain level*)

So you are saying that because [a causal factor] is at [a certain level] that’s why you think this will be the performance level? What about factor [a causal factor not selected]? Did you find out whether [the causal factor not selected] makes a difference? (*Pair responds.*)

1. (***If pair says yes***) Does the applicant’s performance go up when [the causal factor not selected] change from one level to another? (*Pair responds.*)

a. (***If pair says yes***) Then, when someone has a hi level of [the causal factor selected] and [the causal factor not selected], what happens to the applicant’s performance? (*Expected response: the performance goes up even more.*)

(***If needed***) Will the hi level on [the causal factor not selected] make it go up even more than if it was just [the causal factor selected] affecting it?

b. (***If pair says no***) Let’s look at the chart for [the causal factor not selected].

(*Once all predictions have been made*)

Now, choose from all these applicants. Which five should be chosen for the final team? Discuss until everyone agrees.

Now, let’s look at the ones you have chosen.

(*Review each of the chosen applicants and compare their predicted performances. For the ones with different performances, ask pairs to review their predictions side by side.*)

You gave these applicants lower grade, can you explain why?

(*Repeat until all of the predictions of the chosen five applicants have the appropriate*

Here is another question. Let’s say two other students are making predictions. Jack thinks an applicant will get an A because of excellent parent health, Jill thinks that same applicant will get an E because the applicant’s education is no college. What do you think about their predictions?

1. ***(If pair says Jack (or Jill) is right)*** What makes Jack right? *(Pair response: because parent health (or education) makes a difference)* What about Jill (or Jack)? Does education (or parent health) make a difference? (*Pair response: Jill is also right that education makes a difference.*)

2. (***If pair says both Jack and Jill are right and that parent health and education both make a difference)*** Why do they have different predictions?What do they need to think about to come to the best prediction?

3. ***(If pair says both of them are wrong in their predictions)*** What do they need to think about to come to the best prediction? (*Pair responds)*