# **Class Exercise Part 1: Observing specific behaviors and coding as a class**

In this part of the lab we will be viewing a music video and coding for two specific behaviors. The video is for “Somebody That I Used To Know,” by Gotye. We will be coding for two specific variables/behaviors:

1**. Eye Dryness**. Operational definition = normal blinking

2. **Tiredness**. Operational definition = sustained, longer than normal blink or eyes closed

We are using this video because it is relatively short in length, but keep in mind that most observations take place over longer periods of time, and a big issue that we must address is fatigue for the observer. The different types of observational techniques, the number of observers, and the design of the observation (do observers work in pairs? individually? what materials are needed?) must all be decided carefully and thoughtfully. Today’s exercise will provide you with some experience in the difficult (and often monotonous) process of making observations.

We will be splitting you into two groups for this exercise (via random assignment):

1. **Time Sampling** (**Interval Method**) group

Recall that in Time Sampling, the entire observation period gets broken down into intervals of observation and intervals of recording. So, if you are in this group you will be assessing whether or not each of the two behaviors of interest occurred during each 30 second interval of the video (no matter how many times it occurred).

1. **Event Sampling** (**Frequency Method**) group  
   In Event Sampling, you record each and every instance of the behaviors during the entire observation period (though for convenience of coding, you can break this down into time intervals as well).

We will watch the video TWICE so you have a chance to try to each method of coding.

**Gotye: Time Sampling Record Sheet**

**Your name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Observed behavior(s):** (1) eye dryness (2) tiredness

**Behavior Operational definition(s):** Eye dryness will be defined as any time someone blinks normally. Tiredness will be defined as any time someone blinks for an extended, longer than normal time period, or closes their eyes purposefully.

**Total Observation Session:** 4:04 (4 minutes, 4 seconds) **Interval length:** 30 seconds

**Use the letters Y (yes) and N (no) to indicate whether or not the behaviors occurred during each interval.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Interval Number** | | | | | | | | **Total Number of Intervals Behavior Observed** |
|  | **1**  0:00–0:29 | **2**  0:30–0:59 | **3**  1:00–1:29 | **4**  1:30–1:59 | **5**  2:00–2:29 | **6**  2:30–2:59 | **7**  3:00–3:29 | **8**  3:30–4:03 |
| **Behavior 1**  **Y or N** |  |  |  |  |  |  |  |  |  |
| **Behavior 2**  **Y or N** |  |  |  |  |  |  |  |  |  |

**Gotye: Event Sampling Record Sheet**

**Your name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Observed behavior(s):** (1) eye dryness (2) tiredness

**Behavior Operational definition(s):** Eye dryness will be defined as any time someone blinks normally. Tiredness will be defined as any time someone blinks for an extended, longer than normal time period, or closes their eyes purposefully.

**Total Observation Session:** 4:03 (4 minutes, 3 seconds)

**Use tally marks (e.g. | | | | | ) to note number (frequency) of occurrences below during each interval:**

|  |  |  |
| --- | --- | --- |
| **Time Period** | **Behavior 1** | **Behavior 2** |
| **0:00 – 0:29** |  |  |
| **0:30 – 0:59** |  |  |
| **1:00 – 1:29** |  |  |
| **1:30 – 1:59** |  |  |
| **2:00 – 2:29** |  |  |
| **2:30 – 2:59** |  |  |
| **3:00 – 3:29** |  |  |
| **3:30 – 4:03** |  |  |
| **Total Frequency** |  |  |

# **Class Exercise Part II: Observing and Coding Behaviors In Small Groups**

In this part of the lab you will be split into small groups of coders. Each group will watch and code for two behaviors in two movie clips (links below) using the **Event Sampling** method. We have included a coding sheet below. We want a way to measure someone’s interest in a romantic date partner. Our behaviors of interest are:

A**. Smiles**. Operational definition = ???

B. **Disinterest/Boredom**. Operational definition = ???

*In order to code the clips you must first decide, as a group, what concrete behaviors will count as an instance of our behaviors of interest. In other words, you must specify the* ***operational definitions*** *for these two behaviors. Make sure that you determine all of the criteria for counting a behavior, including (if you think it is important) the duration necessary for the behavior to count, as well as what verbal cues may count in addition to facial or body language cues.*

1. **Each member of the group should code individually by themselves WITHOUT discussing it with other members of the group.** Therefore, once you have established your operational definitions, you should split up and work at separate computer stations to code the clips on your own. Note that you may need to watch each video clip a few times in order to code both behaviors of interest.

**Event Sampling Record Sheet**

**Your name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Group member names: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Observed behavior(s):** (1) Smiles, (2) Boredom/Disinterest

**Behavior definition(s): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Total Observation Session:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Use tally marks (e.g. | | | | | ) to note number of occurrences below:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **When Harry Met Sally** | | | | **Hitch** | |
|  | Harry | Jess | Sally | Marie | Sarah | Hitch |
| Smiles |  |  |  |  |  |  |
| Boredom/Disinterest |  |  |  |  |  |  |

**MOVIE CLIPS**

**When Harry Met Sally:** <https://www.youtube.com/watch?v=zGw4fC_Dxo4>

Marie

Jess

Harry

Sally

**Hitch:** <https://www.youtube.com/watch?v=KqfV9eQuyrs>

Hitch

Sarah

1. Once you have all completed coding, the group should create an data file to input the total frequencies for each behavior for each coder in the group.

You should follow the template below, where each column represents one of the group members (coder) and each row represents one of the movie characters. There will be two columns for each coder, one for the smile behaviors and one for the boredom behaviors. You are welcome to use your actual names instead of the word “coder” in the column header, and please note that your group may contain more or less than 4 coders, this template is just an example of how to structure your file but you may need to adapt it for your particular group.

****

1. Once you have entered in the scores (i.e., coded frequency of each behavior) for each member of your group, save the data file in our Google Sheets file, with each group using a separate sheet: <https://docs.google.com/spreadsheets/d/1MS-LoyYyMOOKuhCxNgK9Tm-Ed85TO4kZNwSmW9SgVjI/edit?usp=sharing>

**IMPORTANT: Make sure to replace “Coder 1” etc with the last names of the members in the group!**

# Part III. Calculating Inter-rater Reliability

When we have multiple measurements intended to be measurements of the same variable (in this case smiles or disinterest), we prefer to average them together rather than analyzing every single one separately. This is because any individual measurement will have more measurement error than a group average (just like taking the average of a smaller samples has more error than taking the average of a larger sample).

However, we cannot average things together if they do not seem to actually be measuring the same thing (otherwise, what does the final average even represent?!) Therefore, prior to averaging them together, we need a way to assess whether these items appear to be measuring the same thing. If they do, then we would expect the measurements to relate to each other and be relatively similar. By conducting a reliability analysis, we can get one type of assessment of the degree to which these items (i.e., the different measurements) relate to each other and therefore are likely to be capturing the same underlying thing.

If the reliability is high, then the items show a strong relationship with each other and we may conclude that they are measuring roughly the same thing. If the reliability is low, then the items do not seem to relate and we instead assume they are measuring different things and cannot be reasonably combined together. Typically we would consider anything from .00 to .70 low reliability and anything 0.71 -0.99 (it should not ever be above 1.0) an acceptable level of reliability.

1. Let’s calculate the inter-rater reliability for your group in regards to your measurements of smiles. To do this you first need to download the file, make sure it is a .csv and uses the correct sheet for your group.

2. Open the file in JASP. Under descriptives, we are running a Reliability Analysis, and checking off the boxes for “Chronbach’s Alpha” and “Chronbach’s Alpha (if item dropped)

3. [You will need to wait what seems like FOREVER for this analysis to finish running. I have no idea why it is glitch like this but it does eventually finish running and give you the result] Note your group’s reliability: α =

4. Repeat this process for your disinterest measurements. α =

5. For which measurement did your group show better reliability? Look back to your operational definitions. Why do you think you were more in agreement about one than the other?

6. Now, go back to Excel and create a variable that represents the AVERAGE of your group members’ smile ratings.

7. Repeat this process to create a variable that represents the AVERAGE of your group members’ disinterest ratings.

# PART IV. Reverse-Coding Items for Averaging

The process we just did to calculate the reliability of your group member’s ratings is similar to what we do with survey questionnaires. We rarely ask participants about a construct we care about just one (for example, we rarely would have participants respond to just one statement, “I am depressed” because that would have too much measurement error. Instead, we create questionnaires in which we ask about the same construct we want to measure (like your depression levels) multiple times and in slightly different ways so that we can average across all your answers. Before we can average them together, however, we need to demonstrate that the questionnaire has adequate reliability by conducting and reporting our reliability analysis (usually using Chronbach’s alpha, like before).

In many questions we purposefully phrase some items in the affirmative, so that higher numbers means the participant is reporting MORE of the thing we want to measure.

So, for example, on a scale from 1 (Strongly Disagree) to 7 (Strongly Agree), a person who responds “7” to the item “I am depressed” is scoring a high number because they are showing *more* depression.

HOWEVER, we often also have items scaled in the negative. If we have an item stated negatively, such as “I feel good most of the time,” then a person who responds a “1- Strongly Disagree” is reporting *more* depression than someone who selects the bigger number “7- Strongly Agree.” So we can’t just average this response score with the earlier one because one number represents more depression and the other number represents the opposite. [*Another way to think about this is how you can’t just add 5 and -5 to get 10 while ignoring the negative sign]*. Before we can estimate the reliability OR average our questionnaire, we need to **reverse-score** items that are worded negatively (in other words, they are worded in the reverse way as the rest of the items).

a. For example, here are three items. Which is worded the opposite as the rest and needs reverse-scoring?

1. I like cats.

2. I hate cats.

3. Cats are the best

How do we reverse-score? We can do this in Excel and all we need to know is how many points were on the Likert-type response scale that participants used to respond. Whatever the response scale was out of (for example a 5-point scale), you add 1 and then subtract each person’s actual response from that to get their reverse-scored score.

b. Let’s try it. Imagine I responded a 4 on a 5-point scale and we need to reverse-score it. What is my new, reverse-scored score?

Now let’s do it on this dataset. Open up the file QuestionnaireReliability.csv from Moodle.

c. Imagine that participants responded to a 5-item questionnaire about their confidence in their math abilities. They responded to each statement on a 7-point Likert-type scale from 1 (Strongly Disagree) to 7 (Strongly Agree). The items are (in the same order as they appear in the data):

1. I feel confident in my math skills.

2. I feel able to do most math problems.

3. I worry about my understanding of math.

4. I can follow along with the teacher in my math class.

5. I get lost when the instructor is working through math problems.

d. Which items are phrased in the opposite way as the rest and need reverse-coding?

e. In order to reverse-code them, we will need to use the following formula:

( 1 + # of scale-points available ) – participant’s response

What value should we be subtracting participants’ responses from? \_\_\_\_\_\_\_\_

f. Now make a new variable in a new column. You can call it the same as the variable you plan to recode with \_REV at the end. In other words, if the variable you were reverse-coding was called Test1, you would call your new variable Test1\_REV so that it’s clear it’s a reverse-coded version of Test1.

g. Now you need to input a formula in the first row of data in your new variable. The formula is

= [Value you determined above in step f) – [select corresponding cell from the original variable]

h. Once you have the formula, you can copy it, paste it and drag it down so that it reverse-codes the score for each participant.

i. Repeat this process for all items that need reverse-scoring.

j. Once you have all the items reverse-coded that need it, we will make a new variable that represents people’s average confidence. You can call it MathConf.

k. In this column for the first row of data, you want to insert a formula that will take the average of your five items. BUT, for anything reverse-coded, you need to use your reverse-coded variable NOT the original. So you will have a formula averaging five cells together. For this average, greater numbers mean more math confidence while lower average scores would mean lower math confidence.

l. You may want to save your data separately as an Excel file if you want to keep your formulas saved. Then save it as a .csv and open it in JASP.

m. Run a reliability analysis to determine whether our five statements reliably measured math confidence. Keep in mind that you should be using five items and always using Reverse-scored items when available!

α =

Now paste the reliability output below:

n. Now run descriptives on your MathConf variable to see if you created your average correctly and paste your output below.