

# Juntos Project Initial Report

Baseline Assessment Cleaning and  
Recommendations

Alejandra Garcia Isaza

March 2021



# Juntos Project Description

## Study and intervention details

The *Juntos* Project was a three-year study led by the University of Oregon's Center for Equity Promotion [CEQP](#). The project developed a culturally specific family–school partnership intervention, *Conexiones: Families and Schools United for Equity* (hereafter referred to as *Conexiones*), designed to enhance Latino parents' and educators' capacities to effectively support Latino student success.

The *Conexiones* curricula was built on Latino cultural assets, addressed common challenges confronting immigrant students and families in terms of school success, and utilized effective strategies for increasing educators' awareness of Latino cultures and the barriers that exist for Latino immigrant students and families in schools. It also focused on building effective family-school communication and partnerships with the aim of improving Latino students' academic success.

The six participating schools belonged to three different school districts in the state of Oregon and were randomly assigned to either a control group or a intervention group that received the *Conexiones* intervention program. Study participants completed assessments at three different time points (baseline, immediately post-intervention, and 12-month post-intervention). The complete dataset in the project is made of three waves of data with separate assessments for each participant type (parents, students, and educators).

## Report details

This report will be focusing only on the baseline assessment and is intended to describe the data cleaning process with the aim of helping CEQP staff replicate these procedures in subsequent waves of data and future projects. The report will also include a brief description of the sociodemographic characteristics of the study participants, the scale creation process, the average scores of participants' responses in regards to major study constructs, and recommendations for more advanced statistical analyses that link the different types of participants in the study.



# Data Cleaning procedures

The following section describes the data cleaning procedures performed in each of the participant's type datasets.

## Educator's dataset

The raw dataset had 43 observations and 202 variables of which 17 were metadata variables created by Qualtrics, the software used to create the assessment surveys. In the following code, I removed all but one of the metadata variables, `response_id`, that is an unique identifier assigned by Qualtrics that resulted handy in dealing with duplicated ids. Other data cleaning procedures are described in the comments marked with a `#` sign.

```
elt_w1_clean <- w1_raw_elt %>%  
  janitor::clean_names() %>% # function that formats variables names  
  select(-1:-8, -10:-17, -202) %>% # selecting out columns with metadata  
  rename(c("id" = "pj")) %>% # renaming id variable  
  arrange(id) # ordering participants ids in descending order
```

When evaluating if the dataset had duplicated ids, I found that `id 257` was duplicated and there was no `id 254`.

response_id	id	school	q1	q2	q3
R_1NsKbbg0xSNm9DI	251	2	3	3	2
R_Xvok02kOfilkKV3	252	2	3	3	4
R_294kWxlg2imaph1	253	2	4	3	3
R_3NEywl5hBzdP9Kt	255	2	3	2	3

R_3McjQ3QdB3iSnbT	256	2	4	3	4
R_6EELe7Uuwi9W7zX	257	2	2	2	3
R_3IRUos8weYHpWB1	257	2	4	3	3

After checking with the CEQP data manager, I corroborated that one of the duplicated cases of `id` 257 in fact was `id` 254. I fixed this mistake with the code below using the `response_id` variable and the [mutate](#) and [case\\_when](#) functions.

```
elt_w1_clean <- elt_w1_clean %>%
  mutate(id = case_when(response_id == "R_6EELe7Uuwi9W7zX" ~ "254",
    TRUE ~ as.character(id))) %>%
  arrange(id)
```

The `id` protocol followed in CEQP projects is very simple. They usually use three digits for each individual participant `id` and use the first of these three digits to indicate the school. In this system, `ids` in the 100's would belong to school 1, `ids` in the 200's to school 2, and so on.

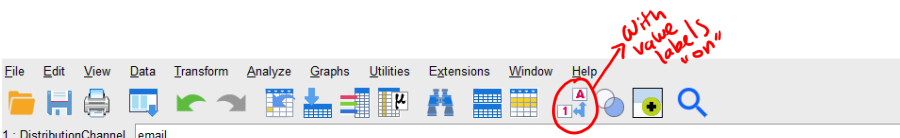
By visual inspection I identified that the first digit of the individual `ids` in the `id` variable did not correspond to the `ids` in the school `id` variable `school` for schools 3, 4, 5, and 6. In the table below, I selected four variables and only the first row of data of each of the six schools to illustrate this point.

As can be seen in the `school` variable, `ids` in the 300's are coded to belong to school 4 and `ids` in the 400's are coded to belong to school 3. I am calling this flip-flopped school `ids`. Schools 5 and 6 are also flip-flopped. This seemed an easy enough fix to make, but later I found an additional error that was coded in the Qualtrics survey that made it challenging to understand what was happening.

id	school	q1	q2	q3
150	1	4	3	3
250	2	4	4	3
350	4	4	3	3

450	3	3	3	3
550	6	3	3	3
650	5	4	3	3

In the images below from the raw data SPSS file it can be seen that when value labels are “on”, that is, instead of showing the values it shows the labels of the values, it appears as if there was no flip-flop because the names of the schools coincided with the numbers that were assigned to them. Indeed, “Kelly” was school 3 and its participants were identified with ids in the 300’s and “ATA” was school 4 and its participants were identified with ids in the 400’s.



1: DistributionChannel\_email

	PJ__	School	Participant_role_5_T EXT	Participant_role_6_T EXT	Q1	Q2	Q3	Q4	Q5
1	150	Cascade Admin...	-99	-99	Strong...	Agree	Agree	Strong...	Strong...
2	151	Cascade Other ...	Registrar	-99	Strong...	Agree	Agree	Strong...	Agree
3	152	Cascade Other ...	-99	Special Education Te...	Strong...	Agree	Agree	Agree	Strong...
4	153	Cascade Other ...	Educational Assistant	-99	Strong...	Agree	Agree	Agree	Agree
5	154	Cascade Aduca...	-99	-99	Agree	Strong...	Agree	Strong...	Agree
6	155	Cascade Teacher	-99	-99	Strong...	Agree	Agree	Strong...	Strong...
7	250	Prairie Mountain Admin...	-99	-99	Strong...	Strong...	Agree	Strong...	Agree
8	251	Prairie Mountain Teacher	-99	-99	Agree	Agree	Disagr...	Disagr...	Disagr...
9	252	Prairie Mountain Teacher	-99	-99	Agree	Agree	Strong...	No Re...	Agree
10	253	Prairie Mountain Teacher	-99	-99	Strong...	Agree	Agree	Strong...	Agree
11	255	Prairie Mountain Other ...	Educational Assistant	-99	Agree	Disagr...	Agree	Strong...	Agree
12	256	Prairie Mountain Couns...	-99	-99	Strong...	Agree	Strong...	Strong...	Agree
13	257	Prairie Mountain Aduca...	-99	-99	Disagr...	Disagr...	Agree	Agree	Agree
14	257	Prairie Mountain Other ...	Media Coordinator	-99	Strong...	Agree	Agree	Agree	Agree
15	350	Kelly Admin...	-99	-99	Strong...	Agree	Agree	Agree	Agree
16	351	Kelly Aduca...	-99	-99	Agree	Disagr...	Disagr...	Agree	No Re...
17	352	Kelly Teacher	-99	-99	Agree	Strong...	Disagr...	Disagr...	Disagr...
18	353	Kelly Teacher	-99	-99	Strong...	Strong...	Agree	Agree	Agree
19	354	Kelly Teacher	-99	-99	Strong...	Disagr...	Agree	Strong...	Agree
20	355	Kelly Teacher	-99	-99	Strong...	Agree	Agree	Agree	Agree
21	450	ATA Admin...	-99	-99	Agree	Agree	Agree	Agree	Agree
22	451	ATA Admin...	-99	-99	Strong...	Agree	Agree	Strong...	Disagr...
23	452	ATA Other ...	Registrar	-99	Agree	No Re...	Disagr...	No Re...	No Re...
24	453	ATA Teacher	-99	-99	Agree	Agree	Agree	Agree	Agree
25	454	ATA Teacher	-99	-99	Strong...	Agree	Agree	Agree	Agree
26	455	ATA Teacher	-99	-99	Strong...	Strong...	Agree	Agree	Agree
27	456	ATA Couns...	-99	-99	Strong...	Agree	Agree	Agree	Disagr...
28	457	ATA Other ...	Attendance Clerk	-99	Agree	Agree	Agree	Agree	Agree

Figure 1: A nice image.

This changed when value labels were “off”. In the image below, the flip-flop is evident again:

12 : Q12

	PJ__	School	Participant_role_5_T EXT	Participant_role_6_T EXT	Q1	Q2	Q3	Q4	Q5
1	150	1	1 -99	-99	4	3	3	4	4
2	151	1	5 Registrar	-99	4	3	3	4	3
3	152	1	6 -99	Special Education Te...	4	3	3	3	4
4	153	1	5 Educational Assistant	-99	4	3	3	3	3
5	154	1	4 -99	-99	3	4	3	4	3
6	155	1	2 -99	-99	4	3	3	4	4
7	250	2	1 -99	-99	4	4	3	4	3
8	251	2	2 -99	-99	3	3	2	2	2
9	252	2	2 -99	-99	3	3	4	99	3
10	253	2	2 -99	-99	4	3	3	4	3
11	255	2	5 Educational Assistant	-99	3	2	3	4	3
12	256	2	3 -99	-99	4	3	4	4	3
13	257	2	4 -99	-99	2	2	3	3	3
14	257	2	5 Media Coordinator	-99	4	3	3	3	3
15	350	4	1 -99	-99	4	3	3	3	3
16	351	4	4 -99	-99	3	2	2	3	99
17	352	4	2 -99	-99	3	1	2	2	2
18	353	4	2 -99	-99	4	4	3	3	3
19	354	4	2 -99	-99	4	2	3	4	3
20	355	4	2 -99	-99	4	3	3	3	3
21	450	3	1 -99	-99	3	3	3	3	3
22	451	3	1 -99	-99	4	3	3	4	2
23	452	3	5 Registrar	-99	3	99	2	99	99
24	453	3	2 -99	-99	3	3	3	3	3
25	454	3	2 -99	-99	4	3	3	3	3
26	455	3	2 -99	-99	4	4	3	3	3
27	456	3	3 -99	-99	4	3	3	3	2
28	457	3	5 Attendance Clerk	-99	3	3	3	3	3

Figure 2: Another nice image.

This survey coding error meant that both values and values labels needed to be changed. I decided instead to create a new variable called `school_id` and delete the flawed original variable `school`. In the code below, the individual participant id variable `id` was used as the reference for the new `school_id` variable.

I also created a new variable called `condition` to indicate which schools were randomly assigned to the control group (coded as 1) or to the intervention group (coded as 2). Finally, I also created a `wave` variable to indicate the wave of the data.

```
elt_w1_clean_2 <- elt_w1_clean %>%
  mutate(school_id = str_sub(id, 1, 1), # new school id variable
         condition = case_when(
           school_id == "1" | school_id == "3" | school_id == "5" ~ "1",
           school_id == "2" | school_id == "4" | school_id == "6" ~ "2")) %>% # new
  select(school_id, condition, everything()) %>%
  add_column(wave = 1, .before = 10) %>% # new wave variable
  select(- school) # deleting school variable
```

## Parent dataset

...

## Youth dataset

...



# Participant descriptives

Say something about participant characteristics

## Educator's characteristics

...

## Parent characteristics

...

## Youth characteristics

...

# Scale creation and Testing

Say something about scales

## Educator's scales

...

## Parent scales

...

## Youth scales

...

# Average Scores of Major Study Constructs

Say something about the average scores...

Educator's average scores

...

Parent average scores

...

Youth average scores

...

note: include plots with average scores

# Recommendations

I recommend...

- id protocol
  - ...

