Juntos Project Initial Report

Baseline Assessment Cleaning and Recommendations

Alejandra Garcia Isaza

March 2021

Center for Equity Promotion
Centro para la Promoción de la Equidad

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.

This image is a placeholder:

Data Cleaning procedures

I spent a loooot of time cleaning!

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_294kWxIg2imaph1	253	2	4	3	3
R_3NEywI5hBzdP9Kt	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 the `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 dentified that the first digit of the indvidual 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 ilustrate this point.

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

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

...

Recommendations

I recommend...

- id protocol
  - ...


Center for
Equity Promotion
Centro para la Promoción de la Equidad