**Issue 1:** In 2018 data, facility location/ camp name was devasted. Zack and I tried to clean the camp name as much as possible (slightly error is possible). There are many data where it is not possible to identify the camp name in such case we dropped the data. On contrary in 2019s data, camp names are straight and cleaned. Can we ignore 2018 camp name. I was wondering, if we can ignore 2018 camp name (instade using camp name from 2019) we got ***9593 matched (after removeing demotions,…. he number is 7174)* data (# of matched increased because we are not dropping data from 2018 where we don’t have the camp name) where as we have *8385****matched data.*

As we are pulling data from 2018 and 2019 based on facility and student name I think we can use the camp data from 2019s data which is clean and precise.

**Issue 2:** Number of unique facility in matched data is 1234. However, in ASER data we have only 110 facility where host teachers’ education level data is available and 143 facility mynmr teacher education level data is available.   
In term of overall data, 1037(14.45%) data has host teachers’ education level and 1110 (15.47%) data has Mynmar teachers’ education level.

Age group <- 3-5

Education level

There are 7 surveys where number of of children in 0 resulting infinity to I.SCHOOL.ratio\_teacher\_students and I.HB\_LEARNING.ratio\_teacher\_students

