Some notations:

= time taken by participant *i* to read question *j*, where *j* = 1 to 15 (or 5 or 10 depending on the case)

= median reading time for question *j,* considering all participants

= first quartile reading time for question *j,* considering all participants

= third quartile reading time for question *j,* considering all participants

= median reading time for question *j,* considering only participants who answered correctly

= first quartile reading time for question *j,* considering only participants who answered correctly

= third quartile reading time for question *j,* considering only participants who answered correctly

= answer of participant *i* to question *j* (1 = correct, 0 = wrong)

= label of participant *i* relating to question *j* (UP = has effort, NP = no effort)

Tester/Question

**Only on valid testers** apply as follow:

Then the label is:

What we need to do:

1. Find invalid in term of data (zero value)
2. Find invalid in term of time (ex. Answer too fast)
3. Define label (case: timelimit/no timelimit)
4. Create column for areas (Question/Choices/Time/Submit…)
5. Analysis 2 Phases: reading question phase and start reading answer phase