

# Data Cleaning Report S23

Kennedy Zapalac

2023-05-24

**Check that the data was cleaned properly by reading the report below**

## Pre-Course Survey Data Processing

Raw data source (there should be 2 if you pulled the “Recorded” and “In Progress” responses from Qualtrics):

- S23Pre-test\_completed\_041223 copy.csv
- 

These are the variables you computed composites for:

- ERQ, StressM, MWQ, RS, GrowM, SelfEf, Motiv

These are the corresponding numbers of questions for the variables you computed composites for:

- 6, 4, 5, 6, 3, 3, 3

Here is a list of all the variables contained in the dataset:

- StartDate, EndDate, Status, IPAddress, Progress, Duration (in seconds), Finished, RecordedDate, ResponseId, RecipientLastName, RecipientFirstName, RecipientEmail, ExternalReference, LocationLatitude, LocationLongitude, DistributionChannel, UserLanguage, student\_ID, student\_ID2, Q4\_Browser, Q4\_Version, Q4\_Operating System, Q4\_Resolution, Q202\_First Click, Q202\_Last Click, Q202\_Page Submit, Q202\_Click Count, ERQ1\_pre, ERQ2\_pre, ERQ3\_pre, ERQ4\_pre, ERQ5\_pre, ATT\_CHECK\_pre, ERQ6\_pre, StressM1\_pre, StressM2\_pre, StressM3\_pre, StressM4\_pre, MWQ1\_pre, MWQ2\_pre, MWQ3\_pre, MWQ4\_pre, MWQ5\_pre, MW\_class\_pre, MW\_hw\_pre, MT\_class\_pre, MT\_hw\_pre, RS1\_pre, RS2\_pre, RS3\_pre, RS4\_pre, RS5\_pre, RS6\_pre, GrowM1\_pre, GrowM2\_pre, GrowM3\_pre, SelfEf1\_pre, SelfEf2\_pre, SelfEf3\_pre, Motiv1\_pre, Motiv2\_pre, Motiv3\_pre, undivided\_actual\_pre\_1, undivided\_ideal\_pre\_1, demands\_pre, Age\_pre\_1, birthday\_1, birthday\_2, birthday\_3, Grade\_pre, Gender\_pre, EthnoRace\_pre, Q199\_First Click, Q199\_Last Click, Q199\_Page Submit, Q199\_Click Count, teacher, ERQ\_pre\_avg, StressM\_pre\_avg, MWQ\_pre\_avg, RS\_pre\_avg, GrowM\_pre\_avg, SelfEf\_pre\_avg, Motiv\_pre\_avg, room\_to\_grow

There was originally `r_pre_length_raw` rows of data from the pre-course survey. We removed 8 previews from our team members, 0 blank responses, and 12 duplicates. In total, we removed 20 rows from the pre-course survey data. This left us with 484 pre-course survey responses.

Of the 484 total respondents, 347 (71.7%) passed the attention check, whereas 102 (21.1%) failed the attention check, and 35 (7.2%) did not answer the attention check question.

Here is how many respondents there were by teacher:

```
##
## Teacher1 Teacher2 Teacher3 Teacher4 Teacher5 Teacher6
##      122      66      50      111      49      86
```

You saved the cleaned dataset as: S23Pre-test\_cleaned.csv

## Post-Course Survey Data Processing

Raw data source (there should be 2 if you pulled the “Recorded” and “In Progress” responses from Qualtrics):

- S23Post-test\_completed\_042123 copy.csv
- S23Post-test\_inProgress\_042123 copy.csv

These are the variables you computed composites for:

- ERQ, StressM, MWQ, RS, GrowM, SelfEf, Motiv

These are the corresponding numbers of questions for the variables you computed composites for:

- 6, 4, 5, 6, 3, 3, 3

Here is a list of all the variables contained in the dataset:

- StartDate, EndDate, Progress, Duration(inseconds), Finished, student\_ID, ERQ1\_post, ERQ2\_post, ERQ3\_post, ERQ4\_post, ERQ5\_post, ATT\_CHECK\_post, ERQ6\_post, StressM1\_post, StressM2\_post, StressM3\_post, StressM4\_post, MWQ1\_post, MWQ2\_post, MWQ3\_post, MWQ4\_post, MWQ5\_post, MW\_class\_post, MW\_hw\_post, MT\_class\_post, MT\_hw\_post, RS1\_post, RS2\_post, RS3\_post, RS4\_post, RS5\_post, RS6\_post, GrowM1\_post, GrowM2\_post, GrowM3\_post, SelfEf1\_post, SelfEf2\_post, SelfEf3\_post, Motiv1\_post, Motiv2\_post, Motiv3\_post, undivided\_actual\_pst\_1, undivided\_ideal\_post\_1, demands\_post, CF1\_post, NPS\_post, CF1\_no\_post, CF1\_no\_post\_10\_TEXT, CF2\_post, CF3\_post, CF4\_post, CQ1\_post, CQ2\_post, CQ3\_post, Age\_post\_1, birthday\_1, birthday\_2, birthday\_3, Grade\_post, Gender\_post, EthnoRace\_post, teacher, ERQ\_post\_avg, StressM\_post\_avg, MWQ\_post\_avg, RS\_post\_avg, GrowM\_post\_avg, SelfEf\_post\_avg, Motiv\_post\_avg, CQ\_score

There was originally 472 rows of data from the post-course survey. We removed 1 previews from our team members, 39 blank responses, and 4 duplicates. In total, we removed 44 rows from the post-course survey data. This left us with 428 post-course survey responses.

Of the 428 total respondents, 321 (75%) passed the attention check, whereas 100 (23.4%) failed the attention check, and 7 (1.6%) did not answer the attention check question.

Here is how many respondents there were by teacher:

```
##
## Teacher1 Teacher2 Teacher3 Teacher4 Teacher5 Teacher6
##      85      61      40      107      46      89
```

You saved the cleaned dataset as: S23Post-test\_cleaned.csv

## Combing Pre and Post Data

The pre-course survey had 484 respondents, and the post-course survey had 428 respondents. If we were able to match all of our pre and post surveys, we would have had 428 matches. However, we had 405 matched pre and post-course surveys. We were able to match 387 by their student ID and teacher, and we matched 18 more students based on their birthday and teacher.

Here is a list of all the variables contained in the dataset:

- student\_ID, teacher, StartDate.x, EndDate.x, Status, IPAddress, Progress.x, Duration (in seconds), Finished.x, RecordedDate, ResponseId, RecipientLastName, RecipientFirstName, RecipientEmail, ExternalReference, LocationLatitude, LocationLongitude, DistributionChannel, UserLanguage, student\_ID2, Q4\_Browser, Q4\_Version, Q4\_Operating System, Q4\_Resolution, Q202\_First Click, Q202\_Last Click, Q202\_Page Submit, Q202\_Click Count, ERQ1\_pre, ERQ2\_pre, ERQ3\_pre, ERQ4\_pre, ERQ5\_pre, ATT\_CHECK\_pre, ERQ6\_pre, StressM1\_pre, StressM2\_pre, StressM3\_pre, StressM4\_pre, MWQ1\_pre, MWQ2\_pre, MWQ3\_pre, MWQ4\_pre, MWQ5\_pre, MW\_class\_pre, MW\_hw\_pre, MT\_class\_pre, MT\_hw\_pre, RS1\_pre, RS2\_pre, RS3\_pre, RS4\_pre, RS5\_pre,

RS6\_pre, GrowM1\_pre, GrowM2\_pre, GrowM3\_pre, SelfEf1\_pre, SelfEf2\_pre, SelfEf3\_pre, Motiv1\_pre, Motiv2\_pre, Motiv3\_pre, undivided\_actual\_pre\_1, undivided\_ideal\_pre\_1, demands\_pre, Age\_pre\_1, birthday\_1, birthday\_2, birthday\_3, Grade\_pre, Gender\_pre, EthnoRace\_pre, Q199\_First Click, Q199\_Last Click, Q199\_Page Submit, Q199\_Click Count, ERQ\_pre\_avg, StressM\_pre\_avg, MWQ\_pre\_avg, RS\_pre\_avg, GrowM\_pre\_avg, SelfEf\_pre\_avg, Motiv\_pre\_avg, room\_to\_grow, StartDate.y, EndDate.y, Progress.y, Duration(inseconds), Finished.y, ERQ1\_post, ERQ2\_post, ERQ3\_post, ERQ4\_post, ERQ5\_post, ATT\_CHECK\_post, ERQ6\_post, StressM1\_post, StressM2\_post, StressM3\_post, StressM4\_post, MWQ1\_post, MWQ2\_post, MWQ3\_post, MWQ4\_post, MWQ5\_post, MW\_class\_post, MW\_hw\_post, MT\_class\_post, MT\_hw\_post, RS1\_post, RS2\_post, RS3\_post, RS4\_post, RS5\_post, RS6\_post, GrowM1\_post, GrowM2\_post, GrowM3\_post, SelfEf1\_post, SelfEf2\_post, SelfEf3\_post, Motiv1\_post, Motiv2\_post, Motiv3\_post, undivided\_actual\_pst\_1, undivided\_ideal\_post\_1, demands\_post, CF1\_post, NPS\_post, CF1\_no\_post, CF1\_no\_post\_10\_TEXT, CF2\_post, CF3\_post, CF4\_post, CQ1\_post, CQ2\_post, CQ3\_post, Age\_post\_1, Grade\_post, Gender\_post, EthnoRace\_post, ERQ\_post\_avg, StressM\_post\_avg, MWQ\_post\_avg, RS\_post\_avg, GrowM\_post\_avg, SelfEf\_post\_avg, Motiv\_post\_avg, CQ\_score

Here is how many respondents there were by teacher:

```
##
## Teacher1 Teacher2 Teacher3 Teacher4 Teacher5 Teacher6
##      82      59      41      100      45      78
```

You saved the matched pre and post dataset as: S23matched\_data\_cleaned.csv