This is my log for cleaning this dataset, G&T Results 2017-18 (Responses). This is the original.

Here's the cleaned version (copy).

STEPS:
Added Headers

- The blog where I found this dataset also had the dataset from the previous year, which included the headers, so I used it as a template for the purpose of this exercise. In real life I would be calling the school/facilitator to verify what the headers should be instead of guessing they were the same as a different year's spreadsheet.
- Case in point: This dataset has a mysterious column M, which is an additional column not included in the previous year's format. Based on the responses, I'm guessing it asked parents if they did test prep with their kids beforehand, where it was and with whom. This is something I would confirm when I called the school/facilitator with the other questions I had when cleaning this data.

Column A fixes:

- Row 1 date is 2017. Changed it to 2018.
- Tried to format dates in A39, A44, to match DD/MM/YYYY format. It did not like that. Instead I highlighted the whole column and formatted it by going to Format/Number/Date in Google Sheets.
- Missing date values in A45, A50, A51 and A52.
- Changed year in A98 from 2021 to 2018
- Here's an issue: The title of the dataset is "G&T Results 2018-19 Responses." Where are all the responses from 2019????? This is something I'd ask when I call.

Column B fixes:

- Centered Ks in B44 and B45.
- Made all lowercase Ks uppercase (highlighted the column) and unbolded the K in B57

Column C fixes:

- Fixed lowercase months (9 instances of september vs. September, etc)
- Changed a Feb to February

Column D, E fixes:

- Highlighted some missing values.

Column F fixes:

- none

Column G fixes:

- Highlighted some missing values and highlighted in yellow a suspiciously low score that is probably a typo in G33

Column H and I fixes:

- None

Column J fixes:

- I'm going to be honest. I don't know what to do with this column. The data have been entered by different parents, leading to lots of different spellings and tons of variation between each entry. I'm also not sure what's a more helpful formatting for this. Columns for each school? And how would I show that there were folks who did not enter any

preferences? Is that an accidental omission or does that mean they do not have preferences?

- Attempted to standardize spellings for schools so the column is at least searchable. Changed Tag in J94 and J53 to TAG; Nest in J9, J45, 47 48 and 53 with NEST. Changed instances of PS (number) to PS(number)
- I see references to both LL and Learning Lab, which I am guessing are the same thing but I don't want to assume. Another question for whoever answers the phone when I call for data clarifications! :D

Column K fixes

- Same problem as with column J. Find+Replace to standardize a few school names.
- Deleted backslash in K96

Column L fixes:

- Changed "yep" in L16 to Yes
- Standardized caps and spelling of Yes using match case and match cell
-

Column M fixes:

- Moved entry from N11 to M11 (seemed erroneously entered in wrong column)
- I broke Column M into three sections. The original column M had information such as whether the student had studied beforehand, how/where they studied, and with whom. I broke each of those details into separate categories. One worry I have with doing that is making assumptions about what the data infers rather than what it actually says-- for example, one person
- M81 had duplicate data from Column L, so I deleted
- Changed instances of y, Y and yes into Yes and instances of N to No