The purpose of this memo is to highlight some of the starker challenges in the current data reporting systems at DC DOES.

The data: 10 years spanning 2011 to 2021.

Problems that this memo will address:

This memo primarily addresses 3 variable categories:

The main 2 are education variables. The other is date variables.

**Problem I:** End Date < Begin Date (3 counts)

The first, and most trivial issue is that there are 3 occurrences of end date preceding begin date.

**Problem II:** CASAS Reporting System: Scales

The second is far more pressing. This set of education variables comes from CASAS.

Now, CASAS is a flagship database, used to make a number of key and crucial decisions when enrolling customers in workforce development programs.

Therefore, it stands to reason that remedying this system is of utmost importance especially if we want to withstand an audit of our systems zooming out and going back.

The problem here is that the reading and writing scores are reported on multiple scales!

Far too inconsistent to be of any use when conducting any analysis linking customer scores to their success with obtaining employment.

1. The first scale: Actual score range as per test
2. The coded grade level: this scale has the highest frequency in the data. And while the conversion scale says to convert scores to integers, there are far more decimals at around the grade levels.
3. Lastly, text based grade levels

**Problem III:** Education level categories.

The education level categories (source unknown) that are included in the Maryland custom report, feature data outside the categories listed in the data dictionary [specifically - 89].

Moreover, and this may seem trivial. The data dictionaries lists the data beginning from the 9th grade. But the actual data contains values from 1 to 8 as well, and significant counts of the same. Of course it can be inferred that these values correspond directly to years of schooling (1=> 1st grade, 8=> 8th grade), but this should be made clear in the documentation. And if missing, should be updated to reflect the same.

This data is pulled from what is entered by customers upon first enrolling in DC Networks. Enrolling creates an automatic enrollment in Wagner Peyser.

According to Jon Toye:

The Maryland custom reports, while it does not necessarily manipulate the underlying data when delivering it to us, may be pulling it from the wrong source. Moreover, it may be counting customers who have since moved. It may be pulling their earlier entered initial information. And may not have the most up to date information.