# Reflection

# Yunzhao Wu

# Teachers College, Columbia University

# This week’s reading talks about data quality and validity. This first reading *Joint Declaration of Data Citation Principles* provides a standard of what data validity is. A robust, accessible data should be examined by importance, evidence, access, etc. (Data Citation Synthesis Group, & Crosas, M, 2013). In addition, the rest of the readings basically refer to various processes for data quality assessment. The evaluation of data involves the completion of various indicators, so what indicators to use to audit data is still an urgent problem to be solved in the academic community. Some authors also offer advice on how to improve data quality and what features of the data they consider to be meaningful. I think all the articles on data quality are complementary. Since the criteria for judging will be constantly refined in the discussion. If the first article is given alone, the concept is indeed somewhat abstract. But when all the readings were read down, I was able to get a lot of examples.

After reading, what I want to know is whether there are rigorous standards for data validity in different areas of expertise. In the world of physics, for example, data transparency may not be as high because of the achievements of some companies involved. However, in the field of education, is it possible to achieve a higher level of data transparency? Education is a more open environment that focuses on ideas rather than profits.

In my future work, always pay attention to data quality is an important criterion for measuring professionalism. This vast volume of data has advantages for understanding consumer needs, enhancing the quality of service, and risk identification and prevention. The use and analysis of big data, however, must be focused on reliable and high-quality data, which is a necessary requirement for the value of big data to be produced (Cai, L., & Zhu, Y, 2015). In the information age, a large amount of data has brought convenience to people. We can analyze market demand and customer preferences through data. However, what is more noteworthy is the authenticity of the data. Since all predictions are based on existing data, if the quality of the data is low, then all work will be futile and even counterproductive.

If I am engaged in a data job in the education industry in the future. My research topic is the employment situation of international students in the United States. From this research, I need to write a detailed Data Management Plan. DMP is a very important part, but many institutions do not take it seriously. A good DMP can not only help researchers standardize the process, but also facilitate future review. After finishing the DMP, I should strictly implement peer review. A dataset should pass examination by experts in the field, just like a journal article. Datasets are particularly vulnerable to cross-disciplinary use, in which case the user does not have the background to analyze the information itself (Kratz, J, 2014). Through peer supervision and review, data quality will be further improved. In a normal working environment, we cannot find the problem of our own work, so we need peer experts to double check. The next step is to make the research work open to the public. For the individual researcher, scientific community, and society at large, openness and accessibility are practical imperatives that come with many benefits. The way forward is Accessible Education Science (van der Zee, T., & Reich, J, 2018). Opening up data and improving data availability can help advance education policies. Other teams doing related research can also improve and develop based on existing information.

# To sum up, combining this week and previous reading materials, I have a deeper understanding of education data. In fact, regardless of the industry, the quality of the data should be taken seriously. A reliable data must first meet all indicators. Then you need to develop a detailed plan and make it public after passing the review.

# References

Data Citation Synthesis Group, & Crosas, M. (2013). *Joint Declaration of Data Citation Principles*. San Diego CA: FORCE11. https://doi.org/10.25490/a97f-egyk

Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, *14*(0), 2. <https://doi.org/10.5334/dsj-2015-002>

Kratz, J. (2014, May 8). Fifteen ideas about data validation (and peer review). *Data Pub*. <https://datapub.cdlib.org/2014/05/08/fifteen-ideas-about-data-validation-and-peer-review/>

van der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open*, *4*(3), 2332858418787466. <https://doi.org/10.1177/2332858418787466>