Defining Data Quality

Damien

- Testing refers to the verification of processes and technology, typically as part of new development works, new builds, new data models and technical changes.
- Data Quality is about ensuring that data meets the requirements for data accuracy, consistency and reliability to enable trusted, actionable insights.
- Observability refers to the ability to understand, monitor, and gain insights into the behaviour, performance, and health of a data system or infrastructure. This includes data processing pipelines, storage systems, databases, and any other components involved in handling and managing data. (Courtesy ChatGPT)

TLDR:

- Testing - New system/works verification
- Data Quality - Health of the data
- Observability - Monitoring and health of system

Rich:

In it's basest form, quality is just referring to some desired *property* of the system. For example: accuracy, latency, performance, access & auth, security, cost.

"Quality is a multi-dimensional **measure that describes how a system or application satisfies the stated or implied requirements**. Stated requirements are equivalent to functions or features and implied requirements are equivalent to performance, useability, security, maintainability or other non-functional requirements." Therefore, DQ goes beyond health of data and should also include metadata. Observability is more about instrumentation - can I see what is going on? In data systems as opposed to applications we have an extra issue in that testing done at "design/build/deploy" time is only half of the story. We also need to be concerned with those same qualities at "run" time because unlike in an app that we own, and implicitly own the (almost) the entire validation process for, we generally do not have that luxury and our biggest operational risk is around issues with the data coming in - as in it deviating from what we expect. So we need observability of both our systems and data/metadata all the time.
I guess summing up is quality is what you want the thing to be like and observability is your ability to verify that it is

DAMA-DMBOK: *The term data quality refers both to the characteristics associated with high quality data and to the processes used to measure or improve the quality of data. These dual usages can be confusing, so it helps to separate them and clarify what constitutes high quality data.*