

---

Professional Certificate in Data Quality Assurance using AI in Education

## Monitoring Data Quality Metrics

---

### Accuracy

: The degree to which data correctly describes the real-world object or event it represents. High accuracy indicates that the data closely matches the actual value.

Related terms: Data quality, Precision

Challenge: Ensure data accuracy in large datasets with various sources and formats.

Example: A student's address data with 95% accuracy would have five incorrectly recorded addresses out of a hundred.

### Completeness

: The extent to which data is populated in a dataset, meaning there are no missing or null values.

Related terms: Data quality, Validity

Challenge: Maintaining completeness in data entry processes and ensuring consistency across data sources.

Example: A dataset of student test scores with 80% completeness would have twenty scores missing out of a hundred.

### Consistency

: The degree to which data is presented in a uniform manner and follows a specific format or pattern.

Related terms: Data quality, Standardization

Challenge: Ensuring consistency when multiple users or sources contribute data.

Example: A dataset of student ages should consistently use the format "XX years" (e.g., "15 years," not "15 yrs" or "15Y").

### Data governance

: The overall management and control of data and its related resources, including policies, procedures, and standards.

Related terms: Data management, Data quality

Challenge: Implementing effective data governance to ensure high-quality data in a rapidly changing technological landscape.

Example: Data governance policies may include rules for data access, data sharing, and data security.

#### Data lineage

: The life cycle of data, including its origin, any transformations, and its current location.

Related terms: Data provenance, Data quality

Challenge: Tracking data lineage in complex systems with multiple data sources and transformations.

Example: Data lineage tools can help trace data from its original source through various transformations to its final destination.

#### Data quality

: The overall condition of data, encompassing its accuracy, completeness, consistency, timeliness, and relevance.

Related terms: Data governance, Data management

Challenge: Ensuring high data quality in large, dynamic datasets.

Example: A data quality dashboard can provide real-time insights into data quality metrics to help identify and address issues.

#### Data quality metrics

: Measurable attributes of data that indicate its overall quality, such as accuracy, completeness, consistency, and timeliness.

Related terms: Data quality, Data governance

Challenge: Defining and tracking appropriate data quality metrics for specific use cases.

Example: Data quality metrics for student performance data might include accuracy, completeness, consistency, and timeliness.

#### Data quality report

: A document or visualization that summarizes data quality metrics, often including trends, issues, and recommendations for improvement.

Related terms: Data quality, Data governance

Challenge: Creating meaningful and actionable data quality reports that effectively communicate data quality issues and solutions.

Example: A data quality report for a school district might include a dashboard with data quality metrics, trends, and comparisons to benchmarks.

### Data validation

: The process of checking data for errors, inconsistencies, or missing values.

Related terms: Data quality, Data governance

Challenge: Implementing automated data validation processes to ensure high data quality and efficiency.

Example: Data validation rules might include checks for data type, format, range, and completeness.

### Precision

: The degree to which data is free from random errors and closely matches the actual value.

Related terms: Accuracy, Data quality

Challenge: Ensuring high precision in data collection and analysis processes.

Example: Precision in student test scores might be measured as the standard deviation of scores around the mean.

### Relevance

: The degree to which data is applicable, useful, and meaningful for a specific purpose or context.

Related terms: Data quality, Data governance

Challenge: Ensuring data relevance in a rapidly changing environment with evolving data needs.

Example: A dataset of student demographics might be relevant for analyzing achievement gaps, but not for predicting student success.

### Standardization

: The process of defining and applying consistent data formats, structures, and conventions.

Related terms: Consistency, Data governance

Challenge: Implementing standardization across multiple data sources and stakeholders.

Example: Standardization might include defining consistent data formats for dates, addresses, and phone numbers.

### Timeliness

: The degree to which data is available and up-to-date for its intended use.

Related terms: Data quality, Data governance

Challenge: Ensuring timely data availability in large, complex systems.

---

Example: Timeliness in student performance data might be measured as the time between data collection and data availability for analysis.

#### Validity

: The degree to which data conforms to established rules, conventions, or standards.

Related terms: Completeness, Data quality

Challenge: Ensuring data validity in complex data environments with multiple sources and formats.

Example: Validity in student data might include checks for data type, format, and completeness.