Foundations for Data Literacy

To be a smart consumer and analyst of data requires the knowledge of a few primary concepts. This document will introduce you to some of the terms or concepts that you will encounter as you deepen your understanding of data.

It is important to start with a way to describe each assessment in terms of purpose and cycle length. Two terms should be used to describe each assessment: the purpose and the cycle length. By providing only one term, an incomplete picture of the assessment is given.

Cylce Length – each term describes how often a particular assessment is/should be given.
- Summative – long cycle length; at the end of a semester or year.
- Interim –mid length cycle; roughly every 6-8 weeks.
- Formative – short cycle length; can be daily.

Purpose – each term decribes what the fundamental purpose of an assessment is/should be.
- Instructional – results are used to directly impact classroom instruction.
- Predictive – results are used to predict student performance on a future assessment.
- Evaluative – results are used to evaluate student performance against what is expected.

Benchmark – A benchmark is a standard (or score) that other things (e.g., scores) are judged or compared against.

> On the President's Physical Fitness Test, the qualifying benchmark for Curl-ups (sit ups) for a 10 year-old boy is 45 in one minute.

Confidence Interval – A confidence interval is an indicator of the reliability of a sample or data point. It is often a "range" that indicates the confidence that the "true value" lies within that range. An example: a certain percentage of time, (e.g., 95% of the time if using a 95% confidence interval) the calculated range of values will actually include the true population value. This is very similar to the concept of Standard Error of Measure (SEM).

> A student scored a 336 on a national standardized test. The 95% confidence interval was +/- 4 points. This means that there is a 95% chance that the true value (does the student know this concept/standard?) lies within a range of 332 to 340 points.

Correlation – Correlation is a measure of the strength of a relationship between two variables. It ranges from -1 (negatively correlated, or as one variable increase, the other decreases) to +1 (positively correlated, or as one variable increases so does the other), with a 0 indicating no correlation or relationship between the two variables. Note that with correlation, the directionality of the relationship is unknown and as such, correlation does not mean that one condition causes another. Pearson's correlation coefficient is denoted by *r*.

> At one school, test scores are positively correlated with classroom grades (r = .9), which means that as one increases so does the other; yet the direction of that relationship is unclear and/or unknown (do high test scores cause higher grades or do higher grades cause higher test scores?).

Cut scores – A cut score is a point or score, based on prior data, that differentiates among categories or classifications of a selected measure or test. For example, cut scores can be used to differentiate if a student has demonstrated a basic, proficient, or advanced understanding of the standards being assessed.

> On a unit test, the teacher used a percentile score to differentiate between grades. For this test, a 93% represented the "cut point" between an A and a B. The "93%" represents a cut score.
> In another class, the teacher decided that students needed to earn a 3 out of 4 (using a rubric) in order to be considered to have shown "masteray" of a specific concept. This "3" represents a cut score.

Descriptive statistics – Descriptive statistics are numbers that describe aspects of a dataset that relate to central tendency (e.g., mean or average) and spread (e.g., range).

> Ms Vang had five students in her classroom. Those five students took a quiz with 10 possible points – Student A scored 5 points, Student B scored 8 points, Student C scored 2 points, Student D scored 10 points, and Student E scored 9 points. Descriptive statistics imdicate that the average score in Ms Vang's classroom was 6.8 (measure of central tendency) and that the range is 8 (measure of spread – lowest score was 2, highest score was 10: 10 – 2 = 8)

Distribution – Distributions show all possible values of a dataset and how frequently they occur.

> An common distribution is the normal distribution or "bell curve." Mr. Smith gave his science class a unit test on animal behavior and then displayed all 91 student scores, from all of his sections, on the test (unlabeled, of course). The scores fell along a bell curve, with most of his students scoring near the average with a few F's and A's.

Evaluative -- an evaluative assessment would be used to evaluate a course, program or the entirety of student learning in a content area and grade level.

> The state mandated Wisconsin Forward test is used to collect data on how well students are progressing towards meeting grade level standards.

Formative – A formative assessment is an assessment or activity that monitors student learning and provides feedback to teachers and students about their learning. This may be given, in different forms, several times a week.

> Mrs. Thomas asked her students to submit an English paper for early review. This formative assessment allowed her to provide her students feedback about their work and also allowed her to get a sense of their progress and understanding.

Instructional – Data that can be used to guide instructional practice. This is typically linked to specific standards and allows for specific teaching moves to be made in support of student learning.

> The 4th grade team was able to monitor student growth throughout the year by comparing the fall, winter, and spring RIT scores on the MAP test.

Interim – An interim assessment is an assessment typically given a few times during the academic year. The results typically would not drive daily instruction, and the assessment would not be sensitive enough to change in order to use it for progress monitoring.

The 4th grade team was able to monitor student growth throughout the year by comparing the fall, winter, and spring RIT scores on the MAP test.

Margin of Error – The margin of error is a measure of the error found in the sampling(assessment). It shows the difference in a specific data point from that of the "true" value. Margin of error is expressed as a range -- +/- a number. In other words, when taking a measure or score from an entire population, the sample being measured may or may not be representative of the population value as a whole. The margin of error denotes how different the sample's value could be from the population's.

Of all the students who scored a 216 RIT on the fall MAP Reading test, there was a margin of error of +/- 3. This indicates that a student who scored a 216 is likely to fall within the range of 213 - 219

Mean – The mean is a measure of central tendency; the "average" or the sum of the observations divided by the number of observations.

Back to the example of Ms Vang and her five students who took the quiz out of 10 points. As a recap, Student A scored 5 points, Student B scored 8 points, Student C scored 2 points, Student D scored 10 points, and Student E scored 9 points. To calculate the mean of these scores, first sum the scores together: 5 + 8 + 2 + 10 + 9 = 34; then divide by the number of observations: 34 ÷ 5, which equals 6.8.

Outlier – An outlier is a data point that is far away from the other data points.

Mr. Chiu gave his students a pop quiz on American History. Most of the students scored an 80% or higher, but one student scored a 30% - that student's score was an outlier relative to the other scores.

Percentile – A percentile can be thought of as a measure of relative standing comparing how one item or person relates to the comparison group as a whole.

Samantha just took the ACT. Her composite score was in the 98th percentile, meaning that her composite score was higher than 98% of people who took that administration of the ACT.

Predictive – a predictive assessment will provide an estimated performance on a future exam. Often, the vendor of a predictive exam has completed a student that links student scores to a future exam.

By scoring a 395 on the ACT Aspire Reading test, the student was predicted to earn a 20-22 score on the up-coming ACT exam.

Quartile – Quartiles are the values that separate or divide your dataset into four equal groups.

On a 40 point test, the 3rd quartile contains scores of 21 to 30. Each quartile represents a range 10 points.

Quintile – Quintiles are the values that separate or divide your dataset into five equal groups.

> The MAP RIT Scores are sometimes divided into 5 sections base on percentiles: Low, Low Average, Average, High Average and High. Each quintile represents a range of 20% of the total RIT score range.

Random Error – Random error is error that is unpredictable.

> Sometimes student mood can affect how well or how poorly they test. The fluctuation of a student's mood is random error that can have an effect on how accurately their abilities are measured.

Range – The range is a measure of spread and represents the difference between the largest and smallest values.

> Ms. Smith gave a test in her math class and the lowest score was a 46 out of 100 and the highest score was a 99 out of 100, so the range of scores in his classroom was 99 – 46, or 53.

Sample – A sample is a subsection of a population of interest.

> A single classroom may represent a sample of a specific grade population across a school.

Standard Deviation – Standard deviation is a measure of how tightly data points are clustered; a small standard deviation means that data points are closely clustered to the average, while a larger standard deviation means that the data points are more dispersed from the average. (~68% of data will fall within a single standard deviation of the mean; ~95% will fall within two standard deviations of the mean; ~99.7% of data will fall within three standard deviations).

> Mrs. Bailey gave a spelling test in her class. She calculated that the average score on the test was 80% and then she calculated the standard deviation. The standard deviation was very small (+/- 2%), indicating that most of her students scored around the average and were at a similar place in their learning (95% of her students scored between 76% and 84%). Mr. Lively gave the same test to his class, the average score was also 80% but the standard deviation was larger (+/- 6%), indicating a high amount of spread (95% of his students scored between 68% and 92%). This indicated to Mr. Lively that his students were at very different places in their learning.

Standard Error of Measurement (SEM) – The standard error of measurement is a measure of how precise an instrument or assessment is, and it represents the distribution of scores around a "true score" with repeated administrations of the instrument or assessment. If the standard measure of error is small, that means that the instrument or assessment is more precise and more sensitive.

> Makenna took the MAP assessment and her RIT score was reported as such:
>
> Student Score Range: 187-190-193
>
> This means that the best guess for Makenna's "true" score is 190, but there is an SEM of +/- 3 RIT, meaning that it is likely that her score falls somewhere between 187 and 193.

Stanine – Stanine is a method of scaling scores along nine points so all test scores become a single digit score between one and nine.

> Students who scored in the 74th percentile on a test can be said to bein the 7th stanine.

Summative – Summative assessments are assessments that check student leraning over a longer period of time (quarter, semester, etc) and compare it against a benchmark expectation or standard.

The ACT is an example of a summative assessment. In the ACT example with Gunnar, his score was compared to others who took the test during his same administration.