

EVALUATING DESIGNED METRICS

1

What is the purpose of this measure?

Examples: evaluating project status or staff performance, a self-assessment or to informing external actors

2

What is the scope of this measure?

Are you collecting data one single time, a single project or is it supposed to be measured continuously in all projects from now on? It may be that metrics that are good on a smaller scale will fail to give the right result, because other factors also affect the outcome.

3

What is the relationship of the attribute to the metric value?

This is the important construct validity problem: How do we know that the metric measures that attribute in a good way?

QUALITY ATTRIBUTE	QUALITY METRIC
What attribute are we trying to measure?	What is the metric (the function that assigns to the attribute)?
What is the natural scale of the attribute we are trying to measure?	What is the natural scale for this metric?
What is the natural variability of the attribute?	What is the natural variability of readings from this instrument?

If you have clear idea of what you actually want to measure, it is easier to establish a good relationship between what you measure and the idea. Some attributes are easy to decide how to measure, such as length. It is more difficult to decide on the correct scales for attributes such as skills, code quality or testing thoroughness. It is natural that there is some variation in an attribute. What is the natural variation of the amount of tests a person can go through each day? And even more important - what causes these natural variations to occur?

4

What measuring instruments do we use to perform the measurement?

Examples of metrics are counting (bugs, lines of code), matching ("Equal complexity of requirements"), comparing (code quality) and timing (time to finish a requirement). This can be done manually or automatically.

5

What are the natural and foreseeable side effects of using this instrument?

Introducing metrics to measure an attribute can be a positive thing. On the other hand it can also give way to a set of unforeseen and negative results, and the less tightly linked a measure is to the underlying attribute, the more side effects can occur. For example, if the quality of a coder is measured by the number of finished requirements per time, this could lead to developers to cherry pick the simple tasks.