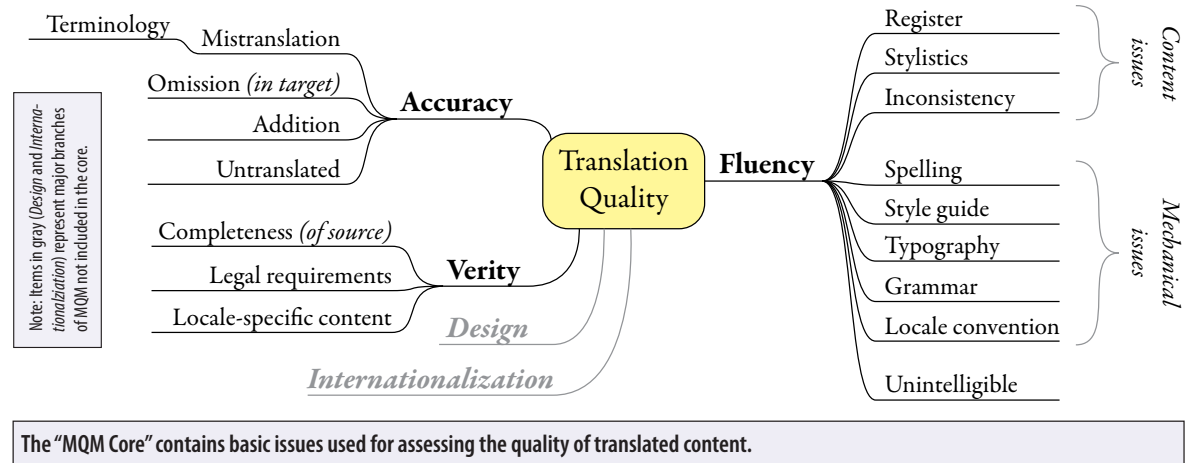


MQM (Multidimensional Quality Metrics)

Multidimensional Quality Metrics (MQM) is a framework for defining metrics and scorecards used to assess the quality of translated texts. The MQM framework defines standard names and meanings of error categories (called *issue types*) used to assess aspects of the quality of a translation and identify problems to be resolved.

MQM metrics are defined with respect to particular *translation specifications* that define expectations for a particular type of translation. The issue types in MQM are organized into a hierarchy such as the **MQM Core** (right).

Translation quality is often assessed using *scorecards*, a type of tool used to identify specific problems found in translated texts so that they can be addressed. Scorecards are typically created independently by individual organizations with little consistency among them. MQM provides this missing consistency, by allowing all stakeholders to know precisely what the issue types mean. (MQM can also apply to *holistic* metrics that evaluate texts as a whole.)



MQM applies to the translation industry, broadly understood to include localization. It can apply to the requirements for various types of translation, ranging from “quick and dirty” to high-end, polished translations. Because of the variety of translation requirements, there is no single metric that is appropriate to all translation projects. There are no absolute requirements for degree of accuracy and fluency independent of the audience, purpose, and other specifications for the project. Thus, rather than talking about absolutes, MQM suggests we talk about quality relative to a metric based on particular specifications. Each issue type can be checked independently, but generally metrics will be a composite of measures for multiple individual issue types.

An assessment of translation quality is based on the choice of a metric. The diagram on the other side of this handout shows what goes into defining a composite metric of quality: the quality dimensions determined by the specifications, a method of assessment, and weights and thresholds. The user of the metric decides on thresholds of acceptance. Various software tools are available to facilitate building and using customized metrics. Ideally, the result of applying a metric is independent of the individual assessor, although some subjectivity is unavoidable for certain issue types, and metrics may not distinguish between high-end translations that meet all requirements. A translation quality assessment workflow might include multiple metrics, such as a “sanity check” followed by a detailed analysis for texts that pass.

If you already have a scorecard in hand (based on a metric derived from a set of specification), then you do not need to consult the next page, which applies only to individuals creating new metrics.

For the formal definition of MQM, please visit the following website: <http://www.qt21.eu/mqm-definition>

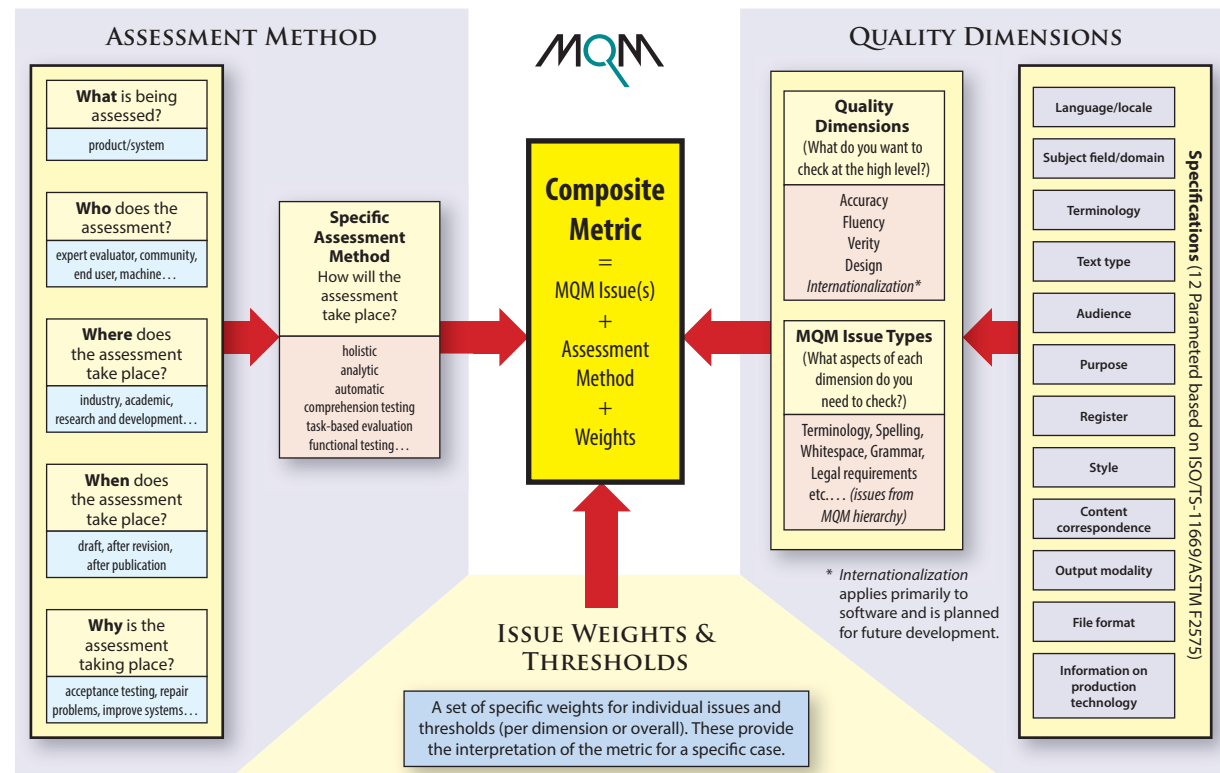
Creating new metrics

In many cases an appropriate metric will already have been chosen for a task. *Individuals evaluating translations will generally not need to consult this process since metrics will be defined by others.* For cases where a metric does not already exist, the process outlined below describes how an appropriate MQM metric is created.

MQM is suited for assessing a **product** (i.e., target text and other deliverables) or **system** (e.g., diagnosis of problems in an MT system). MQM could be expanded through the use of additional parameters to assess a **project** (e.g., whether the product was delivered on time) or a **process** (e.g., whether all agreed-upon tasks were actually performed).

The following inputs are combined to create an MQM metric:

- **Assessment Method.** On the left, the answers to *what, who, where, when, and why* determine the specific assessment method (e.g., detailed error analysis or a holistic set of criteria that apply to the entire text).
- **Quality Dimensions.** On the right, the *specifications* (values of the 12 parameters) determine the quality dimensions and the associated **MQM issue types** needed to assess them. (Note that even though dimensions are broad categories, not all dimensions apply to all tasks, e.g., if Design is not a concern for a particular task, it will not be assessed.)
- In the middle, **issue weights** and overall **thresholds** provide the interpretation of the results of the assessment task. (For example, a metric might specify that compliance with *Legal requirements* is more important than *Stylistics* and also require that the overall quality score exceed a specific threshold).



A composite MQM metric therefore consists of a method, a set of one or more issue types that correspond to each of the dimensions that need to be assessed, and an accompanying set of weights for each issue type. Together these components enable a relevant score (or other indication of quality level) to be calculated and compared against agreed-upon thresholds.

For more information on MQM or to provide feedback on MQM, please send an email to info@qt21.eu.