

# Translation Quality Testing Approach Document

## North Star Metric = Total Quality Score

Total Quality Score is a composite metric that includes:

1. Fluency
2. Accuracy
3. Readability
4. Regulatory Tone Match

**TQS** =  $0.25 \times \text{Fluency Score} + 0.30 \times \text{Accuracy} + 0.20 \times \text{Readability} + 0.25 \times \text{Regulatory Tone Score}$

## Fluency

Fluency refers to correct **grammar + natural sounding** translation.

Fluency can be calculated by:

**Fluency** =  $0.55 * \text{GER} + 0.45 * \text{Accuracy}$

where,

GER is calculated using a LanguageTool API

Accuracy is calculated using Perplexity Score

## Accuracy/Faithfulness

Accuracy refers to correctness which means preservation of semantic meaning.

Accuracy can be calculated by:

**Accuracy** = COMET score

## Readability

Readability refers to how easy it is to understand in the target language for that audience.

Readability is calculated specific to target language based on the index available.

$$\text{Readability} = 0.7 * \text{Index} + 0.3 \text{ ASL}$$

Standard Indexes mentioned below:

Language	Index	Scale
English	Flesch Reading Ease	0–100
Spanish	Fernández-Huerta	0–100
Italian	Gulpease	0–100
German	Wiener Sachtextformel	Grade 4–15 (normalize to 0–100 for composites)
Dutch	Flesch-Douma	0–100
French	Kandel-Moles	0–100
Portuguese	Flesch (Portuguese adaptation)	0–100
Swedish	LIX	Unbounded → normalize to 0–100
Danish	LIX	Unbounded → normalize to 0–100
Norwegian	LIX	Unbounded → normalize to 0–100
Finnish	LIX	Unbounded → normalize to 0–100
Fallback	LIX	Unbounded → normalize to 0–100