

LLQA v1.0

Literary Localization Quality Assessment

TABLE OF CONTENTS

<u>PREFACE</u>	<u>3</u>
<u>Design philosophy</u>	<u>3</u>
<u>Position on localization and adaptation</u>	<u>4</u>
<u>Definition of quality in LLQA</u>	<u>4</u>
<u>Responsibility of the reviewer</u>	<u>5</u>
<u>Relationship between analytical and holistic scoring</u>	<u>5</u>
<u>I. Introduction and scope</u>	<u>7</u>
<u>II. Design Goals</u>	<u>7</u>
<u>III. Conceptual model</u>	<u>8</u>
<u>IV. Sampling protocol</u>	<u>8</u>
<u>V. Localization profiles</u>	<u>9</u>
<u>VI. Phase A: Analytical assessment model</u>	<u>9</u>
<u>6.1 Annotation unit</u>	<u>9</u>
<u>6.2 Error categories</u>	<u>10</u>
<u>6.3 Severity scale</u>	<u>12</u>
<u>6.4 Severity assignment rules</u>	<u>12</u>
<u>6.5 Analytical score computation</u>	<u>13</u>
<u>VII. Phase B: Holistic assessment model</u>	<u>13</u>
<u>7.1 Reading immersion</u>	<u>13</u>
<u>7.2 Character voice preservation</u>	<u>14</u>
<u>7.3 Tone and scene intent preservation</u>	<u>14</u>
<u>7.4 Holistic score calculation</u>	<u>14</u>
<u>VIII. Final score calculation</u>	<u>15</u>
<u>IX. Tier mapping</u>	<u>15</u>
<u>X. Review format</u>	<u>16</u>
<u>XI. Reviewer qualifications and prerequisites</u>	<u>17</u>
<u>XII. Reviewer conduct and constraints</u>	<u>18</u>
<u>XIII. Reviewer disagreement and reuse</u>	<u>19</u>
<u>XIV. Known limitations</u>	<u>19</u>

PREFACE

The following specification has been written to help reviewers assess how effectively (on both a numerical and tier-based scale) an English translation of a literary format recreates the intended reader experience of the target language for a contemporary English-speaking audience.

This specific methodology is part of the larger MNQF (Manga & Novel Quality Framework), which is being developed with the purpose of providing a fast, repeatable, and actionable method to evaluate translation quality from both an analytical and holistic perspective.

The specification is particularly grounded in combining objective empirical quality scoring with subjective scoring. In an ideal world, what is considered a perfect translation would not be a subjective matter; however, we live in an imperfect world where professional translators have different opinions on what constitutes heavy localization, what is missing nuance, etc. This methodology recognizes that difference in opinions and, therefore, is inherently opinionated in of itself. For example, this methodology is written by a reviewer who prefers localization and adaptation over literal translation and direct conveying of nuances, which will unintentionally introduce bias. Please proceed with this in mind.

Design philosophy

This framework is designed specifically for narrative translation (light novels, visual novels, etc.), a medium where translation quality **cannot** be adequately measured through mere sentence-level accuracy. Instead, the LLQA specification proposes an alternative perspective on the evaluation of translation quality, one where the focus lies within the **reconstruction** of the narrative itself. This regards exactly how effectively the translated text is able to recreate:

- 1) emotional impact of the narrative
- 2) character relationships
- 3) the stylistic and authorial intent of the original source text.

At least, from the perspective of LLQA, the ultimate goal of translation is to facilitate an experience where the target language reader processes the narrative identically to the way a source language reader would. Translation, therefore, is about transferring the reading experience of the source text to the appropriate context of the target language and culture. LLQA thereby intrinsically recognizes that the translator becomes a **co-writer** of sorts with the aim of eliciting that emotional and aesthetic response. Thus,

it goes without saying that two translations can *both* be excellent while simultaneously being worlds apart in their approaches to narrative interpretation.

Accordingly, LLQA separates evaluation into the following two stages:

- an **analytical stage**, where the focus lies in the analysis of the overall observable behavior of the translation at the sentence and discourse level, and
- a **holistic stage**, where the focus lies in the overall reading experience of the excerpt and volume.

This separation exists to avoid the common mistake reviewers make where either micro-level linguistic errors dominate their judgment, or macro-level reading experience override the concrete mistranslations. LLQA explicitly treats **both** perspectives as necessary and non-interchangeable.

Position on localization and adaptation

LLQA does not assume that a more literal translation produces higher quality. As a matter of fact, LLQA instead argues the opposite—that localization is more often than not the better approach. Therefore, adaptive localization, restructuring, idiomatic substitution, and selective explicitation are allowed and are more likely to be rated favorably by our approach, as long as they maintain the author’s intention and original interpretation of the text.

To avoid bias toward strategies that make use of adaptive localization, LLQA ensures a clear distinction between *translation errors* and *deliberate localization choices*. Reviewers are also required to apply tags (e.g. *Heavy Localization*, *Literal Translation*, *Keeps Honorifics*, *Drops Honorifics*) to classify the intended localization approach the translator employs. Keep in mind, however, that these tags are to be used solely for the interpretation of results and may not be treated as penalties.

Hence, a translation *may* receive high analytical and holistic scores even if it is highly localized, as long as the narrative, tone and character voice and intent are maintained and still remain faithful to the source.

Definition of quality in LLQA

Within LLQA, “quality” is defined as the degree to which the translated text enables an English-language reader to experience:

- the same narrative
- the same emotions
- the same character relationships and personalities

- and at least a somewhat comparable level of style and authorial voice

—as intended by the original source text.

Accuracy is therefore treated as accuracy with respect to the *narrative*, rather than to every individual sentence. That is, however, **not** to say that sentence accuracy is not considered, and improper conveyances of nuance are still erroneous. The evaluation merely prioritizes whether overall meaning, voice, and intent are preserved, even if the individual sentences are reorganized. As such, sentence-level deviations are only tolerated as long as they do not distort, omit, or mistakenly reframe the underlying meaning conveyed by the source.

Responsibility of the reviewer

This methodology is intentionally opinionated in its design. LLQA prioritizes the philosophy that reader immersion and adaptive localization is more important than strict sentence-level accuracy.

Therefore, reviewers using LLQA are expected to: **1)** disclose their own translation preferences where relevant, **2)** apply localization tags accurately, and **3)** avoid penalizing stylistic choices solely because they differ from their personal philosophy.

Where disagreement exists between reviewers, LLQA encourages comparing their analytical annotations rather than directly the outcomes of their final tiers.

Relationship between analytical and holistic scoring

LLQA combines analytical scoring (error-based) and holistic scoring (experience-based) for a reason. They are meant to answer separate questions, neither sufficient in isolation and both required for a realistic evaluation of a translation.

The analytical stage is (broadly) intended to answer:

| *What objectively went right or wrong in the translation when compared to the source?*

And the holistic stage is (broadly) intended to answer:

| *How well does the translated text function as a literary work for the reader?*

A translation with few analytical penalties may still receive a low holistic score if the tone, pacing, or narrative cohesion are not up to par. Likewise, a translation with noticeable analytical issues may still succeed holistically if its narrative intent and emotional impact seem to be preserved.

The final LLQA score is therefore a weighted combination of the two rather than a replacement of either perspective.

I. Introduction and scope

LLQA (Literary Localization Quality Assessment) is a source-based evaluation framework and methodology designed to assess the quality of English translations of Japanese light novels and similar narrative works.

The framework is intended for:

- dialogue-heavy and narrative fiction
- serialized or volume-based light novels
- works where tone, character voice, and pacing are central to the reader's experience

This specification is not designed for subtitles, manga panel translations, marketing material, game UI text, or technical documentation with no narrative to speak of.

The objective of the specification is to provide a method for evaluating both the correctness and faithfulness of the translation relative to the source text and the literary effectiveness of the translation as a stand-alone English novel. LLQA is meant primarily to be used for comparative evaluation between editions, translators, or localization approaches.

Note that the LLQA specification alone does not prescribe a single preferred localization style or translation style. It is designed to accommodate multiple legitimate translation strategies and to evaluate translations within the context of their declared localization profile. However, LLQA recognizes that adaptive localization and idiomatic restructuring are often more effective at preserving authorial intent and reader interpretation. Please be advised that translations that succeed in doing so may therefore be rated more favorably.

II. Design Goals

The framework is designed to:

1. separate translation correctness from literary quality
2. allow multiple localization and translation strategies to exist without bias
3. constrain reviewer subjectivity through explicit procedure and definition
4. enable fast evaluation for reviews comparing different translation work
5. allow both fine-grained and high-level comparisons between a translation

III. Conceptual model

LLQA's evaluation model occurs in two phases.

Phase A (Analytical Assessment) is based on the deduction of penalties based on the source text, similar to the MQM model. Reviewers compare each target sentence with its source counterpart and identify sentence-level or clause-level deviations using predefined error categories and severity levels.

Phase B (Holistic Assessment) is based on the quality of the target text, and the experience the reader collects as they read the narrative. Reviewers evaluate the translated text as a narrative piece / literary work without consulting the source during scoring.

The two phases measure combinatory but non-interchangeable properties:

- Phase A measures semantic fidelity, correctness, and technical translation quality
- Phase B measures perceived quality, narrative effectiveness, and literary success

Phase A judgements must always be justified with comparisons of source and target evidence. Phase B judgements are allowed to be based on subjective impression, but must follow standardized scales, which are defined as per §7.1, §7.2, and §7.3.

IV. Sampling protocol

In an ideal world, there would be no need for sampling, and the quality assessment could be conducted on the entire literary work, as that would provide the most accurate measure of the translation's quality. However, reviewers unfortunately cannot realistically assess entire novels at a time, hence why sampling is necessary. Given this, LLQA must ensure that the selected samples accurately represent the novel's content while, simultaneously, remaining feasible enough in length for the reviewers to evaluate.

To accomplish this, three chapters are selected per volume:

- the opening chapter, to capture exposition and the establishment of character voice
- one randomly selected middle chapter, to capture the raw narrative
- the final chapter, to capture climactic and emotionally dense material

From each selected chapter, a contiguous 500-word English segment is extracted. Segments may begin at any narrative point, though the best portrayal of the chapter is a randomly selected segment. Sampling must be performed *before* any review begins and must not be adjusted based on perceived quality.

V. Localization profiles

Prior to error annotation, reviewers are to assign descriptive tags that classify the translation strategy observed per each excerpt. These tags are not evaluative and may not affect the score in any shape or form. Their purpose is to provide interpretive context and to prevent penalizing intentional stylistic or localization choices.

At minimum, one tag must be selected from each required pair:

- **Localization:** [Heavy Localization] or [Semi-localized Translation] or [Literal Translation]
- **Cultural Markers:** [Keeps Honorifics] or [Drops Honorifics]
- **Register Tendency:** [Formal / Polite] or [Casual / Slang-heavy]

Additional tags may include, for example:

- [Westernized idioms]
- [Character-specific slang]
- [Reordered narrative structure]

VI. Phase A: Analytical assessment model

6.1 Annotation unit

Each annotation unit, in essence, is a single sentence or clause. A clause-level annotation may be used when only one part of a sentence exhibits an error.

Each annotation must include:

- an identifier of the target excerpt
- an identifier of the target sentence or clause,
- the exact corresponding source span,
- the exact target span containing the issue,
- one error category
- one severity level
- a brief justification

Justifications should be concise and factual, describing the error and why it qualifies under the selected category. Overlapping annotations are permitted when multiple independent issues occur in the same text span.

6.2 Error categories

Each identified issue must be assigned one primary error category. When a single problem plausibly fits multiple error categories, the reviewer is to select the category that best reflects the main category that impacts the reader.

The categories are defined as follows.

<i>Category</i>	<i>Description</i>	<i>Examples</i>
Accuracy / semantic fidelity	This category covers failures to preserve explicit propositional meaning. It focuses simply on how the content of the source is reflected with literal equivalence in the target, independent of stylistic quality.	<ul style="list-style-type: none"> - Incorrect “do-er” and “under-goer” roles - Altered quantities - Incorrect tense or modality - Incorrect causal relations - Altered conditional structure
Pragmatic Meaning & Implicature	This category covers mismatches in implied meaning rather than literal content.	<ul style="list-style-type: none"> - Failure to preserve indirect refusals - Softened commands - Ironic statements - Sarcasm - Honorific-based politeness - Conversational implicatures - Indirect nuance conveyance
Terminology & Naming Consistency	This category covers inconsistent or incorrect treatment of terminology that function as conceptual anchors across the narrative	<ul style="list-style-type: none"> - Ability names - Clubs - Locations - Organizations - Nicknames - Titles - Recurring metaphors - Catchphrases - ...etc
Register & Tone	This category evaluates social and interpersonal relationships encoded in	<ul style="list-style-type: none"> - Unintended shifts between polite and bluntness - Excessive casualization - Over-formalization - Emotional hardening/softening that alters how that character relates to another
Style & Voice	This category evaluates preservation of distinct narrative or character expression	<ul style="list-style-type: none"> - Flattening idiosyncratic speech habits

	patterns	<ul style="list-style-type: none"> - Loss of recurring rhetorical patterns - Altered narrative distance - Using generic prose where there is distinct diction
Fluency & Literary Naturalness	This category evaluates the target language quality independent of the source.	<ul style="list-style-type: none"> - Incorrect or broken syntax - Unnatural word or clause order - Awkward collocations - Repetitive sentence patterns - Constructions that appear mechanically translated
Localization & Cultural Adaptation	This category evaluates whether adaptations preserve cultural and narrative integrity.	<ul style="list-style-type: none"> - Inappropriate cultural substitutions - Misleading equivalents - Removal of cultural-specific references - Inconsistent treatment of setting-specific conventions
Narrative Function & Pacing	This category evaluates the functional role of the sentence in the construction of the overarching scene	<ul style="list-style-type: none"> - Over-expansion of brief beats - Compression of suspenseful buildup - Premature clarification of reveals - Reordering that weakens narrative timing
Emotional & Aesthetic Impact	This category evaluates preservation of affective and sensory impact.	<ul style="list-style-type: none"> - Neutralizing emotionally charged wording - Exaggerating mild reactions - Loss of imagery - Disruption of poetic or rhythmic effects
Cohesion & Coherence	This category evaluates local discourse connectivity.	<ul style="list-style-type: none"> - Ambiguous pronouns - Unclear referents - Broken topic continuity - Missing discourse markers - Altered information structure that weakens logical flow
Linguistic Conventions	This category evaluates the formal correctness of written English.	<ul style="list-style-type: none"> - Grammar mistakes - Incorrect punctuation - Capitalization errors - Misspellings

		- Inconsistent typographic conventions that interfere with professional presentation
--	--	--

6.3 Severity scale

Severity reflects impact on reader interpretation and understanding, not personal stylistic preference. Reviewers are to keep this in mind when classifying the severity of an error.

Neutral (0)

| Reviewer preferences or stylistic notes where the translation is valid and accurate

Minor (-1)

| Meaning is clear and intact. Issue affects style, tone, consistency, or technical quality, but does not disrupt understanding or require rereading

Moderate (-3)

| Meaning is partially altered or clarity/flow is disrupted enough to affect immersion or require rereading to understand.

Severe (-6)

| A strong misrepresentation that changes how the reader interprets a character's intent, emotional stance, or situation, but does not constitute a structural narrative failure.

Critical (-10)

| Core meaning, plot, character intent, or speaker attribution is wrong or missing (e.g., hallucinations, role reversal, fabricated content).

6.4 Severity assignment rules

Severity is determined by the most severe plausible reader impact within the local context. If multiple small problems jointly cause a single major misunderstanding, they may be grouped as one Moderate or Severe issue.

Conversely, multiple independent clauses must be annotated separately. Severity must also not be increased solely to compensate for reviewer dissatisfaction with stylistic choices.

6.5 Analytical score computation

All analytical penalties are aggregated across the evaluated excerpts. The total is then normalized per 1000 words to allow comparison between samples of different length.

Let:

- `TOTAL_PENALTIES` be the sum of all penalty values assigned during Phase A,
- `TOTAL_WORDS` be the total number of words evaluated across all excerpts.

The normalized penalty rate is computed as:

$$\text{penalties_per_1000_words} = \text{TOTAL_PENALTIES} / (\text{TOTAL_WORDS} / 1000)$$

The analytical score is then computed as:

$$\text{analytical} = \max(0, (60 - \text{penalties_per_1000_words}))$$

This normalization ensures that scores are comparable across reviews even when the exact excerpt length varies. Analytical scores are lower-bounded at zero and are not capped above 60.

Please note that the score reflects relative translation reliability and functional correctness rather than literary excellence.

VII. Phase B: Holistic assessment model

Holistic assessment evaluates each translated excerpt solely as an English literary narrative without consulting the source text. Reviewers must only rely on the target while assigning holistic scores for proper interpretation of this phase. The reviewer is to read, or at least skim, the entire excerpt continuously before assigning scores.

There are three holistic dimensions that are to be measured by the reviewer, all on a five-point scale.

7.1 Reading immersion

This dimension evaluates how seamlessly the translation functions as natural English prose. It concerns overall readability, rhythm, narrative flow, and absence of translation friction. Scores reflect how easily a general reader could consume the passage as a narrative and fictional piece without being distracted by awkward phrasing, unnatural structure, or stylistic stiffness that makes the text feel translated.

Scoring

- 5 – Reads like a professionally published English LN.
- 4 – Minor stiffness, but mostly natural and flowing.
- 3 – Frequently feels translated; rhythm and phrasing are stiff.
- 2 – Consistently feels translated at the prose level.
- 1 – Actively awkward or difficult to read.

7.2 Character voice preservation

This dimension evaluates the extent to which individual characters’ speaking and narrative voices remain distinct, stable, and recognizable in translation. Scores reflect how well the translation preserves differences in personality, attitude, social positioning, and expressive habits, such that characters can be reliably distinguished through dialogue and narration alone.

Scoring

- 5 – Clear, distinct voices; dialogue alone identifies characters.
- 4 – Mostly distinct, slightly flattened.
- 3 – Voices blur; personality is muted.
- 2 – Characters sound largely similar.
- 1 – Voice is broken or misleading.

7.3 Tone and scene intent preservation

This dimension evaluates how accurately the translation reproduces the intended emotional tone, atmosphere, and narrative function of a scene as conveyed in its source text. Scores reflect whether translation causes the reader to interpret the scene’s purpose (e.g., comedy, tension, romance, conflict, or reflection) in the same way as the original.

Scoring

- 5 – tone and scene purpose fully preserved.
- 4 – mostly preserved with small shifts.
- 3 – intent clear but tone muted or off.
- 2 – tone or intent largely incorrect.
- 1 – tone or intent broken or misleading.

7.4 Holistic score calculation

Holistic evaluation is performed independently for each excerpt using these individual dimensions. Holistic score per dimension is a summation of these individual dimensions, weighted to forty percent of the overall LLQA score.

Per each excerpt, the computation occurs as follows:

$$\text{holistic} = (\text{reading immersion} + \text{character voice} + \text{tone}) / 15 * 40$$

The final holistic score for the volume is the arithmetic mean of each excerpt-level holistic score.

VIII. Final score calculation

The final LLQA score represents the combined evaluation of source-based translation correctness and reader-based literary quality.

The score is computed using the results obtained in:

- §6.5 (Analytical score calculation), and
- §7.4 (Holistic score calculation).

The final score is computed as follows:

$$\text{Final score} = \text{Analytical score (0–60)} + \text{Holistic score (0–40)}$$

The maximum possible score is 100.

The final score is intended for **high-level comparison and tier assignment only**. Interpretation of translation quality must always consider the underlying analytical annotations and holistic sub-scores.

It is **not intended** to support fine-grained ranking between translations with very small score differences. Differences of only a few points may fall within normal reviewer variance and sampling variance.

IX. Tier mapping

The tiers of translation are defined from a ranking of S–F, excluding E. There are six tiers in total. The LLQA score ranges they fall within are as follows:

Tier	Score
<i>S</i>	95–100
<i>A</i>	85–94
<i>B</i>	70–84

<i>C</i>	55–69
<i>D</i>	40–54
<i>F</i>	< 40

X. Review format

Each LLQA review must include, at minimum, the following components:

1. Evaluation metadata

- Work title and volume
- Evaluated language pair
- Reviewer identifier
- Date of review
- LLQA specification version used

2. Sampling record

- Identification of the three sampled chapters
- Start and end location of each 500-word excerpt
- Total word count evaluated

This information must be sufficient for another reviewer to reproduce the same sampling window.

3. Localization profile

- All required translation strategy tags assigned per excerpt
- Any optional reviewer-defined tags used during evaluation

4. Analytical annotation units (as per §6.1)

For each annotated issue:

- excerpt identifier
- sentence or clause identifier
- source span
- target span
- error category
- severity level
- brief justification

5. Score summary

- total penalties
- total words evaluated
- penalties per 1000 words
- analytical score (0–60)
- holistic sub-scores per excerpt
- averaged holistic score (0–40)
- final score (0–100)
- tier assignment

6. Reviewer notes (optional)

A short qualitative summary highlighting **1)** notable strengths of the translation **2)** recurring or systematic issues and **3)** any uncertainty encountered during the review. The notes must not override or reinterpret the numerical results.

XI. Reviewer qualifications and prerequisites

LLQA reviewers must be performed by reviewers who meet the following minimum qualifications.

Reviewers must possess a sufficient degree of proficiency in both the source language and target language to independently interpret sentence-level meaning, nuance, and narrative function without relying on intermediary translations.

Reviewers must be able to read and analyze the source text directly and must not base analytical judgements on existing translations, other fan translations, subtitles, or secondary summaries.

Reviewers must have familiarity with the media, including common narrative structure, dialogue norms, character archetypes, and genre-specific stylistic patterns.

Reviewers must have sufficient proficiency in written English narrative prose to evaluate fluency, stylistic naturalness, and readability at a literal level rather than solely at a grammatical level.

Reviewers must be capable of distinguishing between:

- acceptable alternative renderings,
- localization strategy choices, and
- genuine translation errors as defined by the LLQA error categories.

LLQA does not require formal academic or professional certification. However, the framework assumes that reviewers possess the linguistic and literary competence necessary to apply the defined categories and severity levels consistently and in good faith.

XII. Reviewer conduct and constraints

1. All analytical penalties must reference a specific sentence or clause identifier and corresponding source and target spans.
2. Translation strategy tags must be assigned before any annotation or scoring begins.
3. Localization and adaptation choices must not be penalized unless they cause:
 - loss or distortion of meaning,
 - damage to narrative logic,
 - an error qualifying as Moderate, Severe, or Critical under the defined categories.
4. Holistic scores must be based only on the sampled excerpt.
5. Reviewers must evaluate each excerpt independently before forming any overall impression of the translation
6. Reviewers must not adjust analytical severity levels in order to compensate for strong or weak holistic impression, or vice versa
7. Analytical annotations must be based strictly on direct comparison between source and target text, and must not rely on assumptions derived from other chapters, volumes, or adaptations
8. Holistic judgements must reflect the perceived reading experience of the translated excerpt alone and must not incorporate knowledge of the source text beyond what is necessary to judge tone and scene intent.
9. When uncertainty exists between two severity levels, reviewers are to select the lower severity unless clear evidence supports a higher classification
10. Reviewers must not introduce new error categories or redefine existing categories during a review
11. Any uncertainty regarding name readings, terminology conventions, or established localization choices must be documented as reviewer notes and must not be

penalized unless the inconsistency is clearly internalized within the evaluated text.

XIII. Reviewer disagreement and reuse

LLQA permits and expects reasonable disagreement between qualified reviewers and provides guidance for comparison.

When multiple reviewers evaluate the same work:

- a) each reviewer must perform independent sampling, tagging, annotation, and scoring
- b) analytical annotations and holistic scores must be produced without consulting other reviewers' results in advance

Disagreement between reviewers should be reported using side-by-side analytical scores, side-by-side holistic scores, and a short comparison of moderate or severe discrepancies. Reviewers are also encouraged to identify whether these disagreements arise from. Common disagreements are between different interpretations of source meaning, different severity judgements, or different holistic impressions of literary quality.

When LLQA results are used for aggregation, benchmarking, or public reports:

- individual reviewer scores must remain accessible
- average scores may be reported, but only alongside the full set of reviewer results

LLQA does not define mandatory reconciliation or arbitration procedure.

XIV. Known limitations

LLQA relies on partial sampling rather than full-text evaluation.

As a result, the final score may not capture issues or strengths that occur outside the sampled excerpts, including long-term consistency problems, delayed narrative effect, or stylistic shifts that occur across the volume itself.

LLQA assumes reviewer competence in both the source and target languages.

The framework constraints subjectivity through procedure and definitions, but it cannot compensate for insufficient linguistic or cultural knowledge.

LLQA results are sensitive to excerpt selection and may vary when different random samples are used

Certain dimensions of translation quality (e.g. stylistic elegance, emotional resonance, and perceived naturalness) remain inherently subjective even when structured scoring rubrics are used

LLQA evaluates translations relative to the provided source text only
It does not assess faithfulness to author interviews, external canon, adaptations, or later revisions.

LLQA scores are designed for comparative evaluation and quality signaling, not for determining contractual compliance, professional certification, or translator performance appraisal.