

S-DAT: A Multilingual, GenAI-Driven Framework for Automated Divergent Thinking Assessment

Jennifer Haase, Paul H. P. Hanel and Sebastian Pokutta

A multilingual approach to measuring divergent thinking through semantic distance in GenAI embeddings.

Motivation

Problem – Traditional Creativity Assessment:

- Human scoring: Labor-intensive, expensive, limits scalability
- English-centric tools: Barriers for global research
- Cultural biases: Compromise cross-cultural validity

Goal: Develop a multilingual, automated, and free-to-access divergent thinking measure that enables scalable, culturally-sensitive creativity assessment worldwide.

Background: Divergent Thinking & Semantic Distance

Divergent Thinking (DT):

- Cognitive process: Generate multiple, diverse solutions to open-ended problems
- Key creativity indicator: Fluency (quantity), flexibility (variety), originality (uniqueness) [1]

Semantic Distance:

- Computational measure: Quantifies semantic differences in embedding space
- Greater distance → more creative, diverse thinking patterns

Original DAT [3]:

- First automated DT assessment using GloVe embeddings
- Limitation: English-only, context-free word representations
- Cannot capture cross-lingual semantics or cultural variations

Method: The S-DAT Framework

Pipeline Overview:

- Generate 10 words (user input)
- Convert to multilingual embeddings (final choice: IBM Granite 278M)
- Compute pairwise cosine distances
- Calibrate against original DAT distribution
- Compute percentile scores

Semantic Distance Computation:

$$\text{dissimilarity}(\vec{a}, \vec{b}) = 1 - \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|}$$

Model Evaluation:

- 8 models tested:** OpenAI, Cohere, IBM Granite, Sentence-BERT
- 11 languages/scripts:** Latin, Cyrillic, Devanagari, Kanji, Arabic scripts
- Convergent validity:** Alternative Uses Task (AUT)
- Discriminant validity:** Bridge-the-Gap task (convergent thinking measure)

Model Selection – IBM Granite-278M:

- Cross-linguistic stability: Consistent semantic representations
- Calibration consistency: Reliable percentile mapping
- Convergent validity: Strong correlations with human ratings

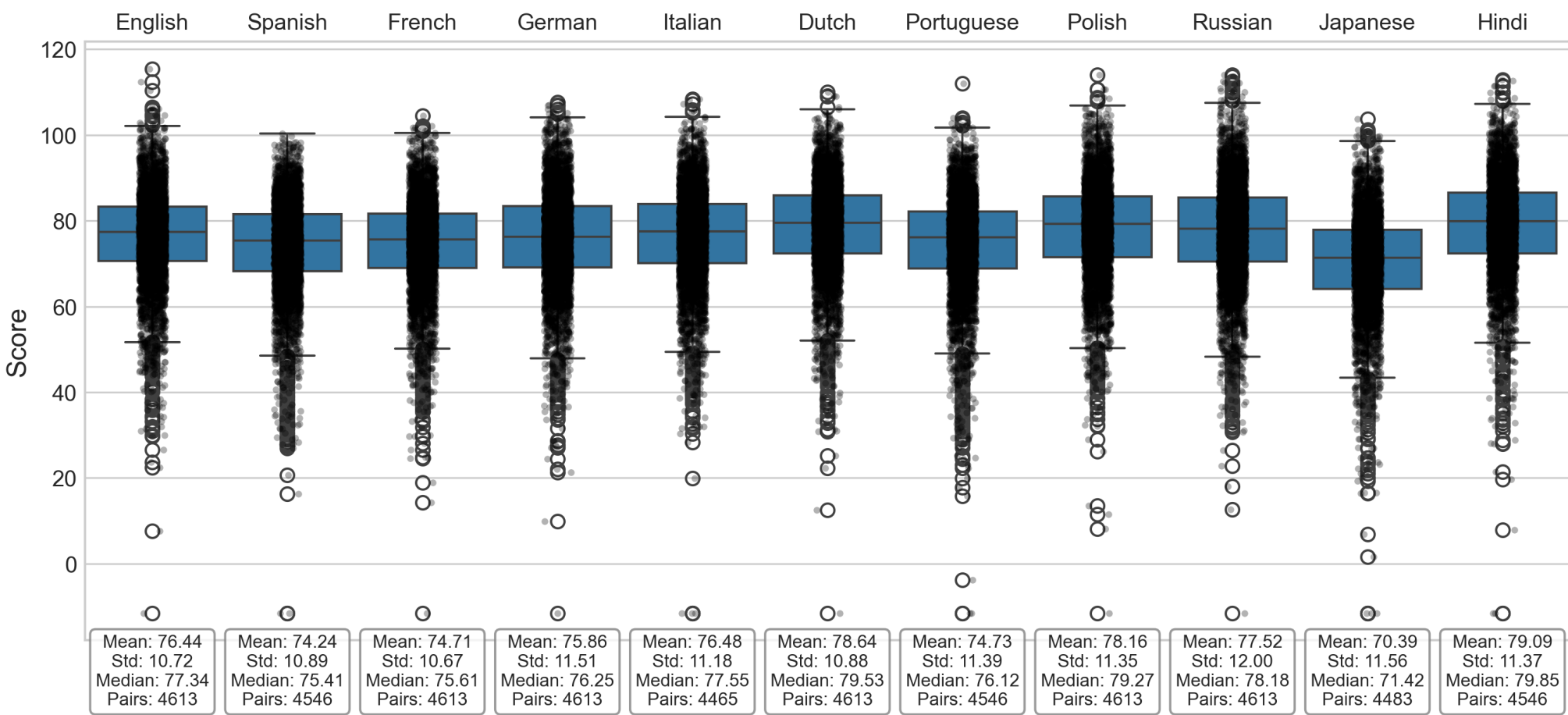


Figure 1. IBM Granite-278M shows most stable cross-linguistic performance, tested with 11 languages/scripts.

Key Results

Model Selection: IBM Granite-embedding-278M-multilingual

- 278 M parameters, 12 layers / 12 heads, embedding size = 768
- Strongest cross-linguistic calibration stability and lowest cross-linguistic variance ($\tau_{\Delta} \leq 0.06$)
- Trained on multilingual corpus with 250 k vocabulary tokens

Validity Results:

- based on $\approx 9,000$ human responses
- Convergent validity:** AUT correlations $r \approx 0.19\text{--}0.27$ (significant across languages/scripts)
- Discriminant validity:** No correlation with Bridge-the-Gap $r \approx 0.08\text{--}0.11$ (convergent thinking)
- Measures divergent thinking specifically, not general cognition

Discussion & Outlook

Strengths:

- Scalability:** Automated assessment enables large-scale studies
- 11 languages/scripts:** Multiple scripts break language barriers
- Validity:** Strong alignment with human-rated creativity measures
- Open, free access:** via website and API

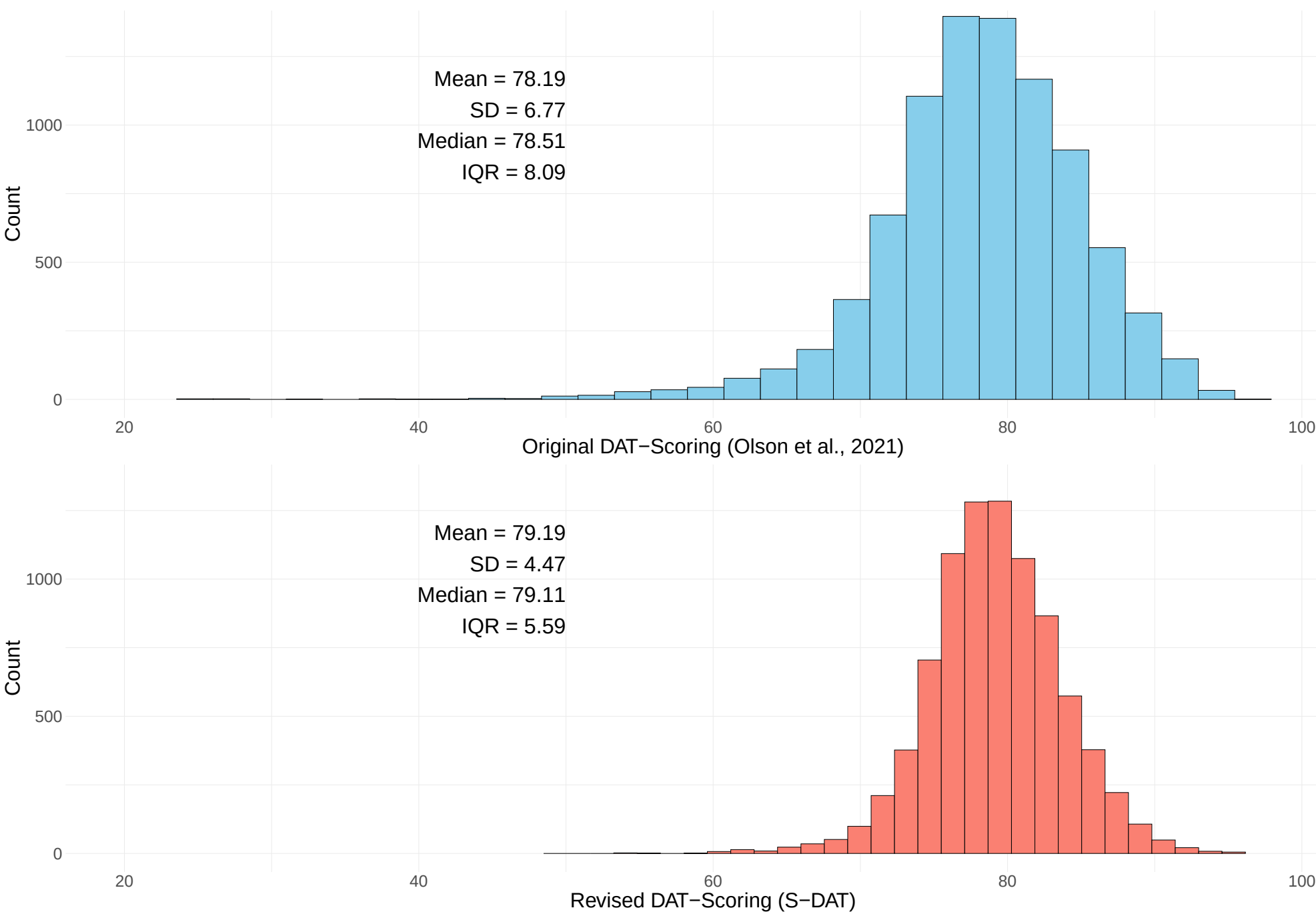


Figure 2. S-DAT produces more stable, less skewed distributions

Limitations:

- Context sensitivity:** Single-word focus may miss contextual expressions
- Language-specific calibration:** Some languages may need fine-tuning
- Cultural biases:** Embedding models may reflect training data biases

Future Directions:

- Validate S-DAT with data from different languages
- Validate S-DAT against other creativity measures
- Test fairness & bias across cultural semantics
- Norm percentile ranges for each language

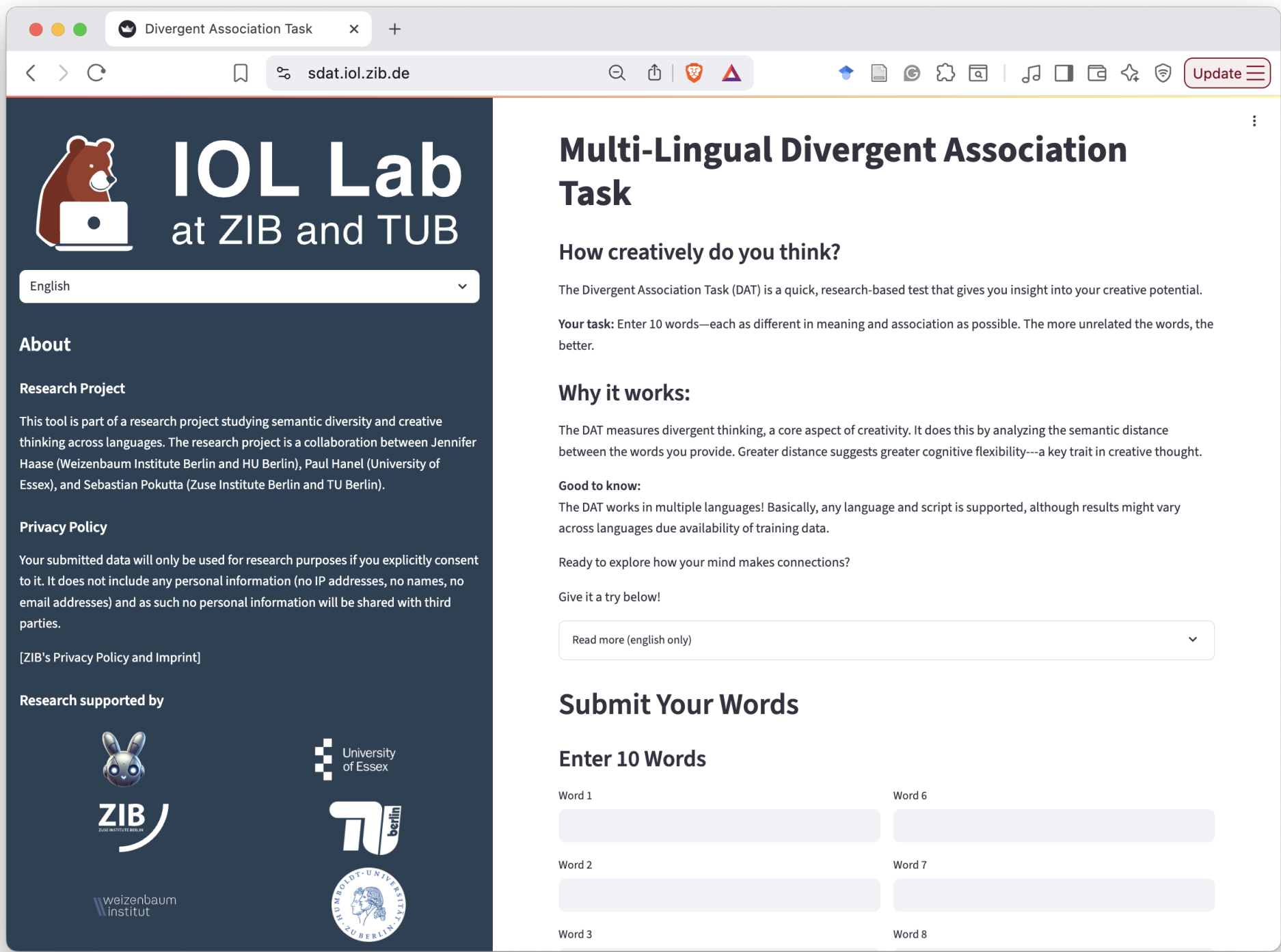


Figure 3. S-DAT webpage (sdat.iol.zib.de)



Key References:

- [1] J. P. Guilford. Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1):3–14, 1967.
- [2] J. Haase, P. H. P. Hanel, and S. Pokutta. S-DAT: A Multilingual, GenAI-Driven Framework for Automated Divergent Thinking Assessment, May 2025. arXiv:2505.09068, doi: 10.48550/arXiv.2505.09068.
- [3] J. A. Olson, J. Nahas, D. Chmoulevitch, S. J. Cropper, and M. E. Webb. Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, 118(25), 2021.

