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

Jennifer Haase, Paul. H. Hanel & Sebastian Pokutta

MIC Conference 2025, Bolzano, Italy

Where do we start?

Where do we start?

- Core “issue” of creativity measurements

Where do we start?

- Core “issue” of creativity measurements

  → hard to assess fairly, fast, and across languages

Where do we start?

- Core “issue” of creativity measurements

  → hard to assess fairly, fast, and across languages

  → most tools are labor-intensive, culturally biased, and English-centric

Where do we start?

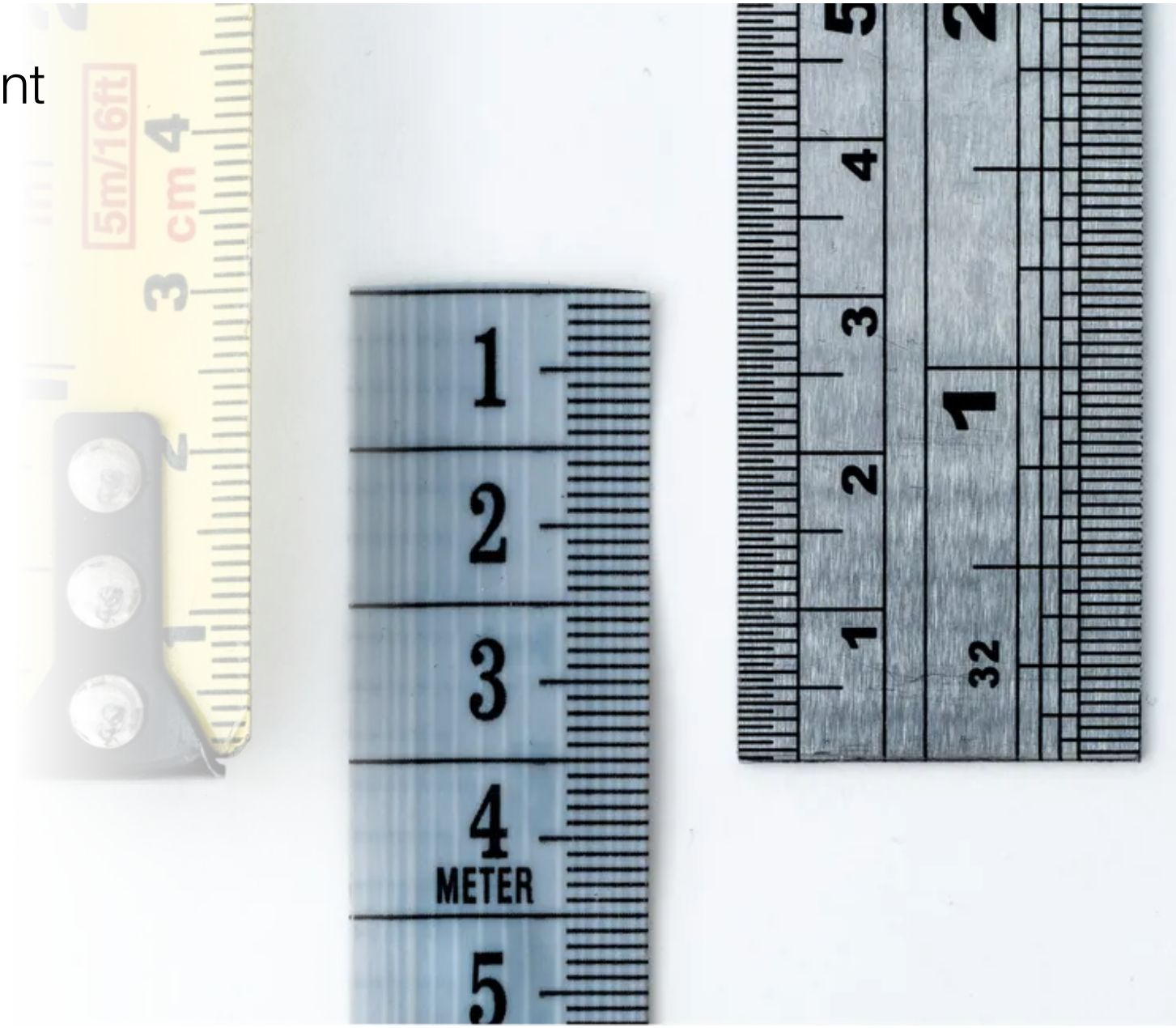
- Core “issue” of creativity measurements

  → hard to assess fairly, fast, and across languages

  → most tools are labor-intensive, culturally biased, and English-centric

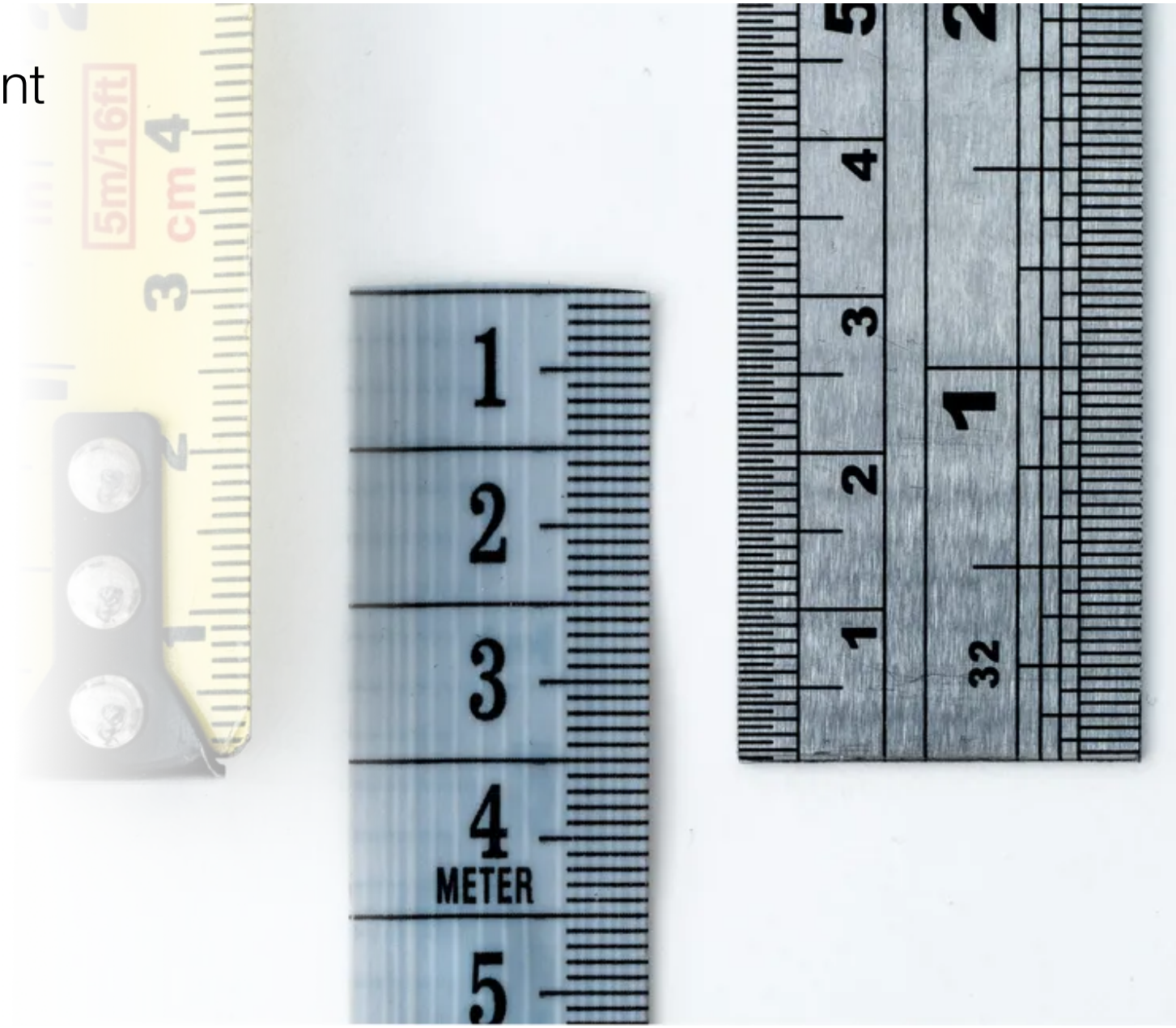
- LLMs have already been used successfully to run and score creativity tests

The Problem with Current Creativity Assessments



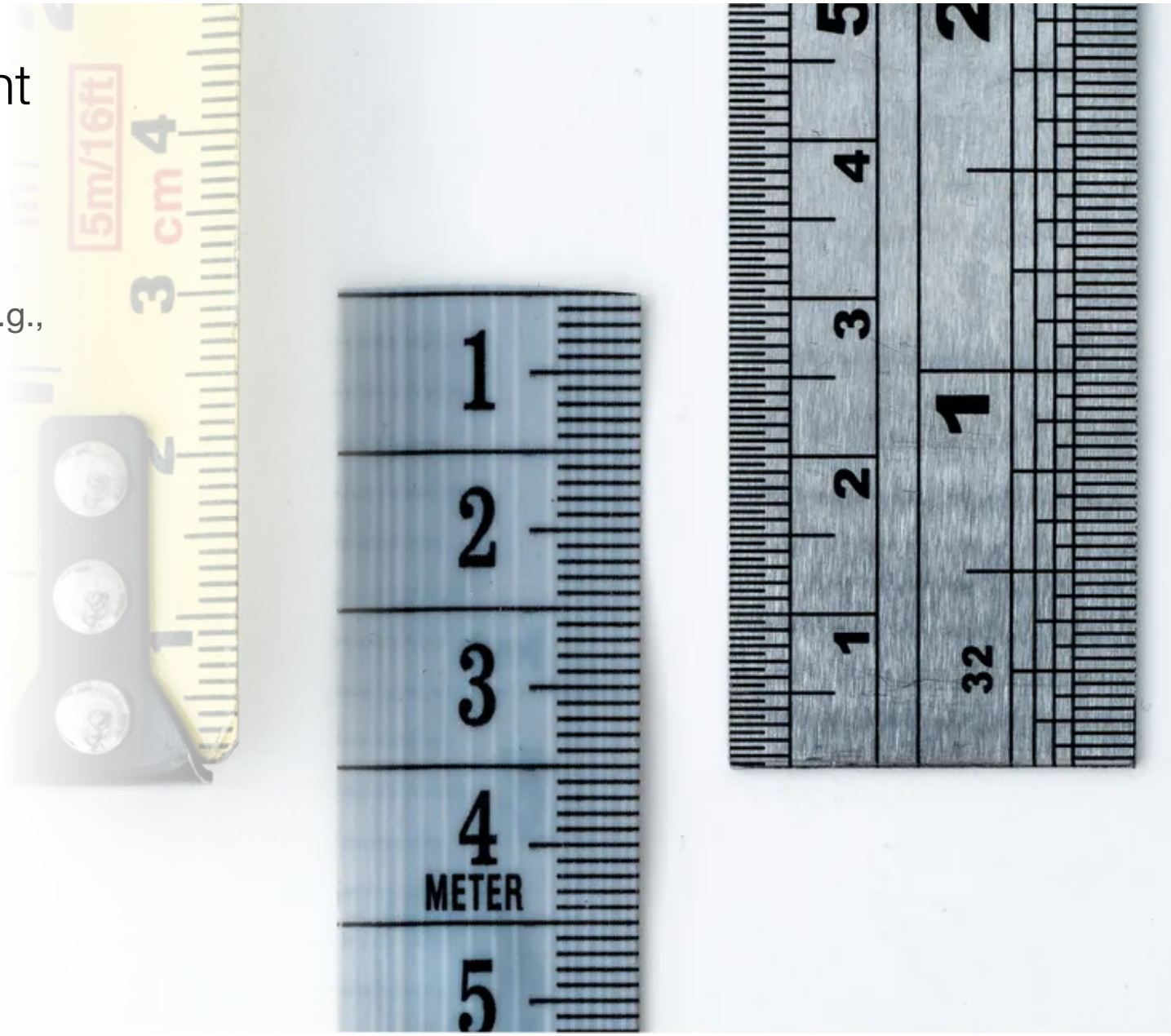
The Problem with Current Creativity Assessments

- **Traditional tools**:



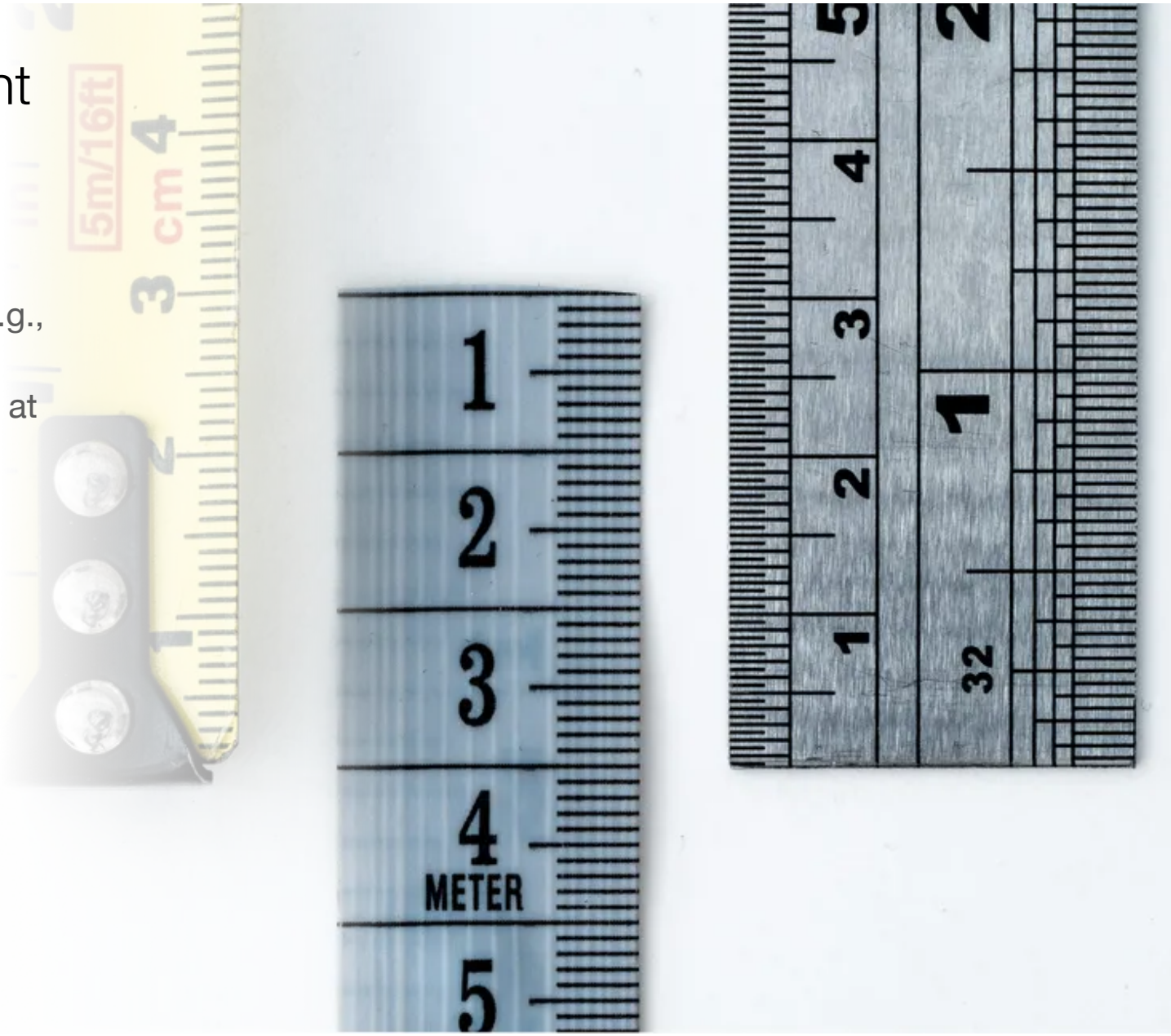
The Problem with Current Creativity Assessments

- **Traditional tools**:
  - Require mostly manual scoring (e.g., CAT method)



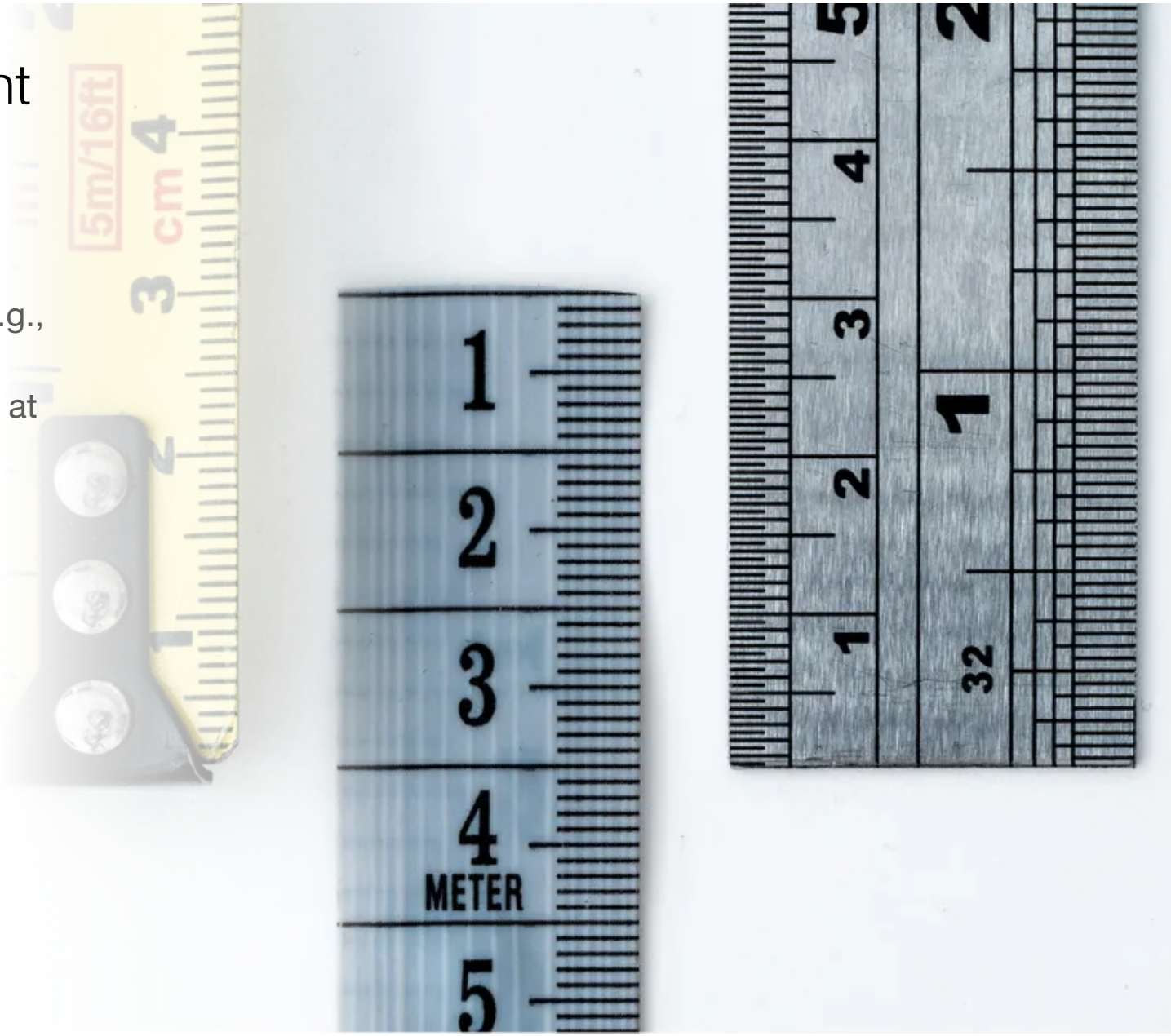
The Problem with Current Creativity Assessments

- **Traditional tools**:
  - Require mostly manual scoring (e.g., CAT method)
  - They are slow and costly to apply at scale (e.g., TTCT)



The Problem with Current Creativity Assessments

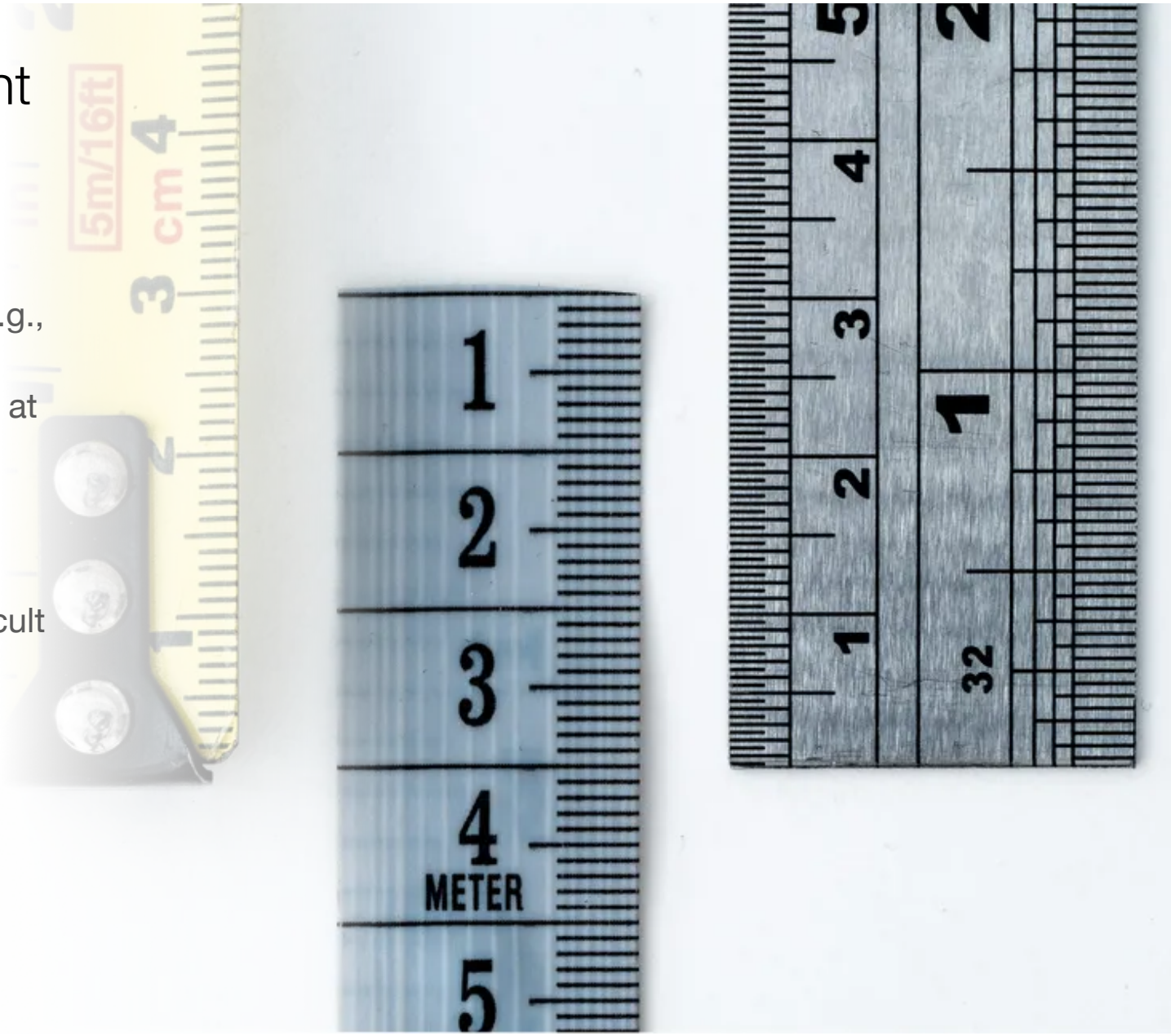
- **Traditional tools**:
  - Require mostly manual scoring (e.g., CAT method)
  - They are slow and costly to apply at scale (e.g., TTCT)
  - Are biased toward English and Western contexts



The Problem with Current Creativity Assessments

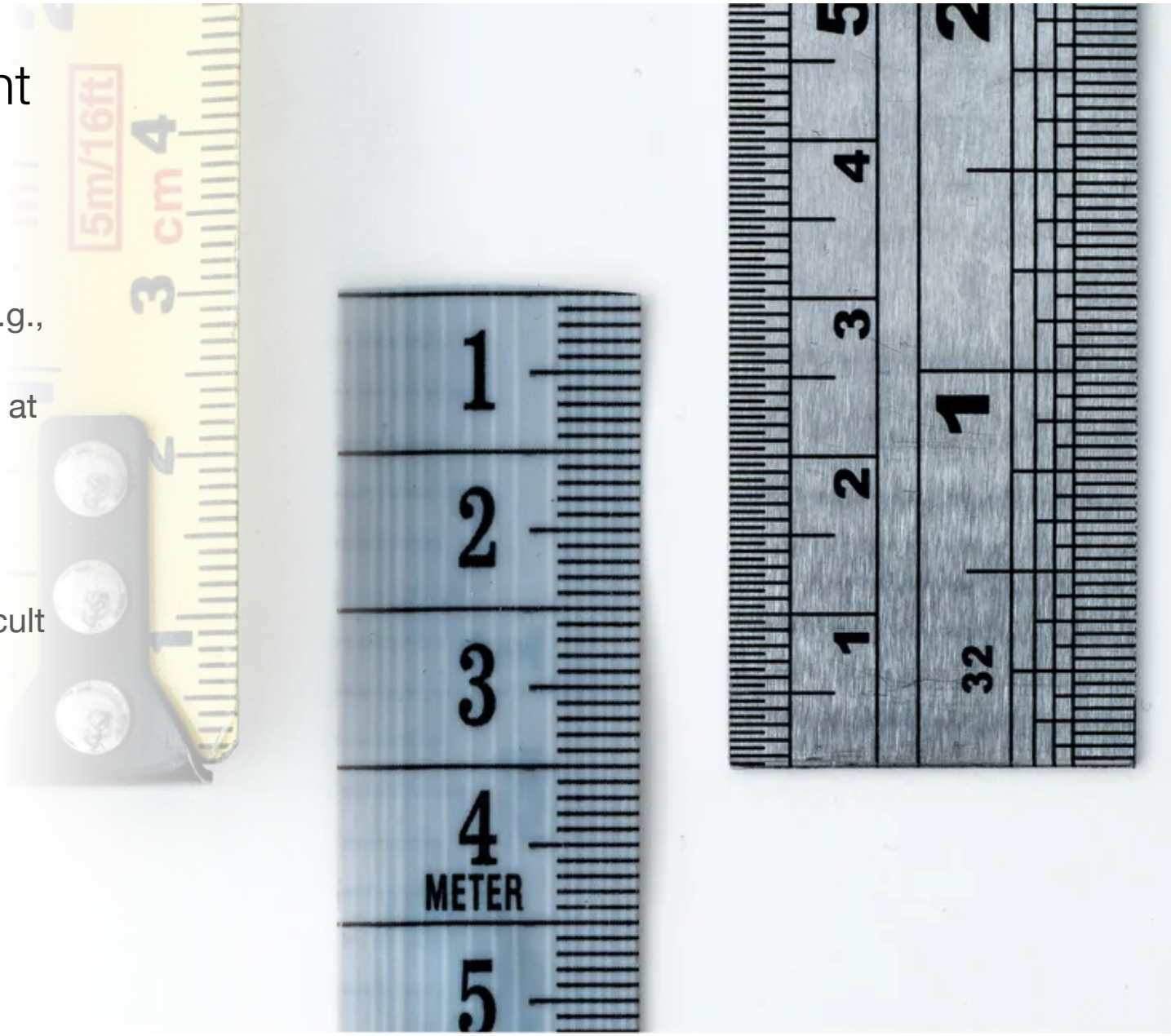
- **Traditional tools**:
  - Require mostly manual scoring (e.g., CAT method)
  - They are slow and costly to apply at scale (e.g., TTCT)
  - Are biased toward English and Western contexts

- **Cross-cultural assessment** is difficult with static or English-only tools

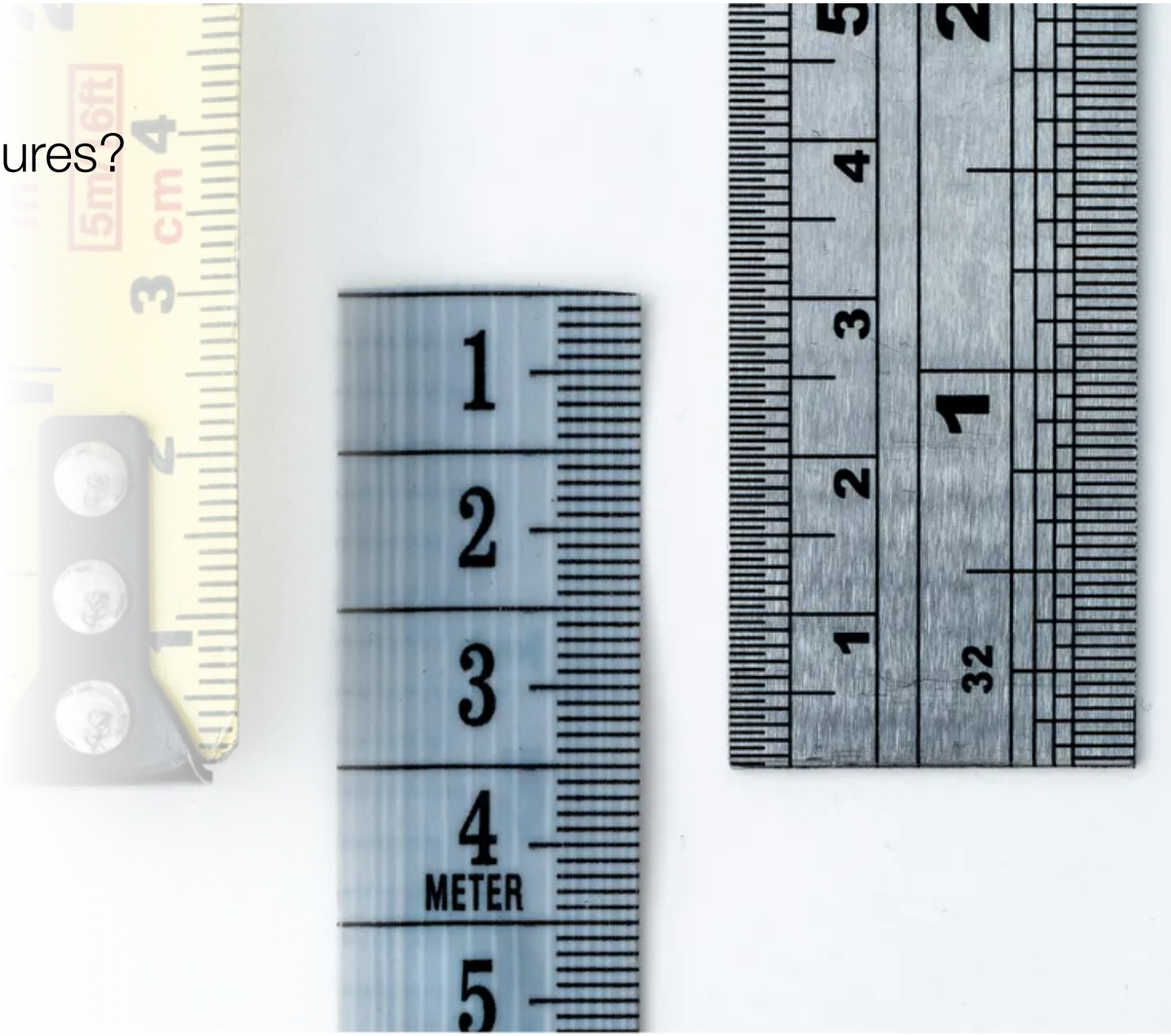


The Problem with Current Creativity Assessments

- **Traditional tools**:
  - Require mostly manual scoring (e.g., CAT method)
  - They are slow and costly to apply at scale (e.g., TTCT)
  - Are biased toward English and Western contexts
- **Cross-cultural assessment** is difficult with static or English-only tools
- We need tools that are:
  - **Fast** (low burden)
  - **Fair** (language-agnostic)
  - **Comparable** (across cultures)

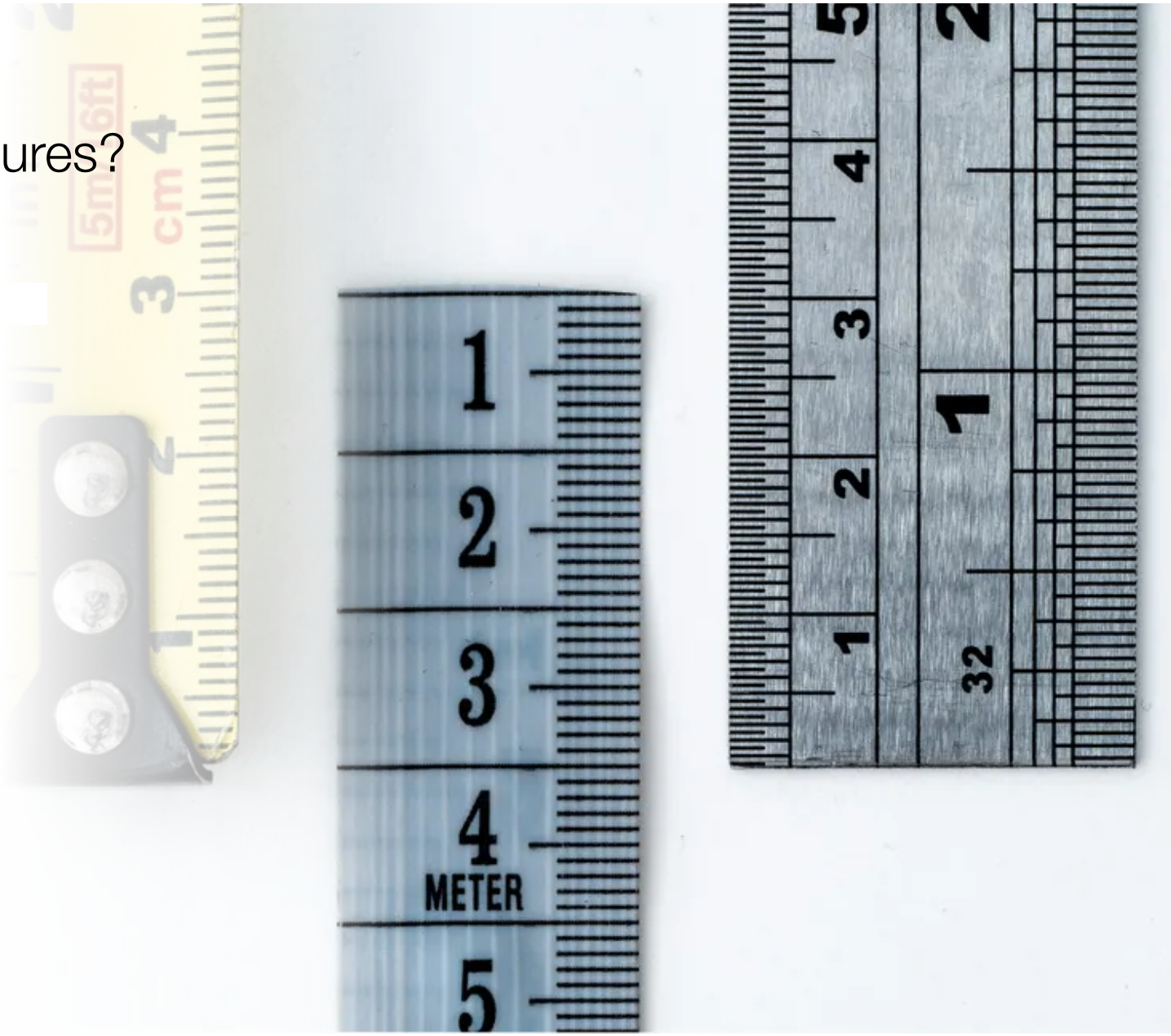


Why Leverage LLMs and GenAI for Creativity Measures?



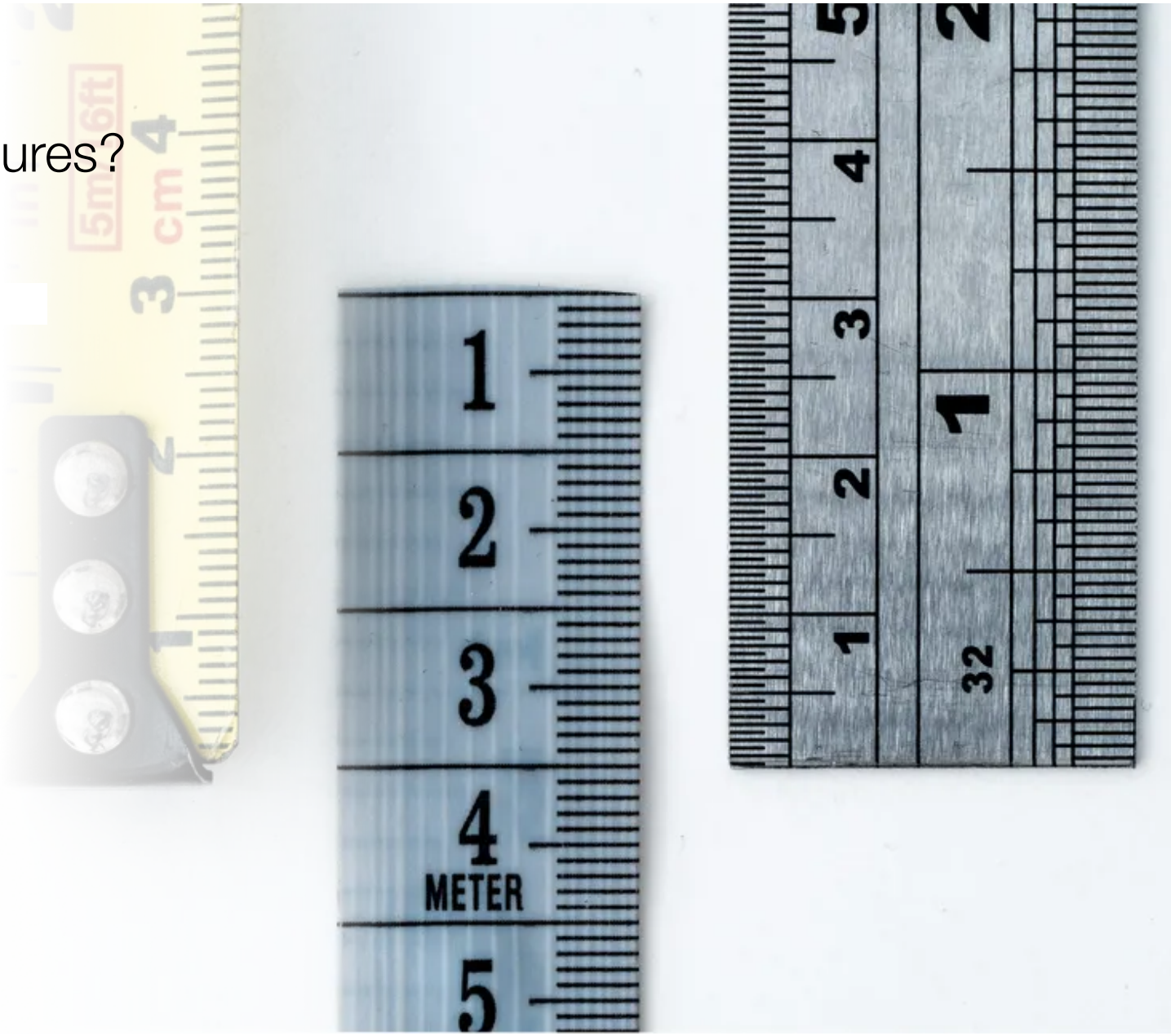
Why Leverage LLMs and GenAI for Creativity Measures?

- For **AUT,** LLM evaluation models correlate with human judges up to r = .813[1]
- LLMs yield **high inter-model agreement** on creativity ratings



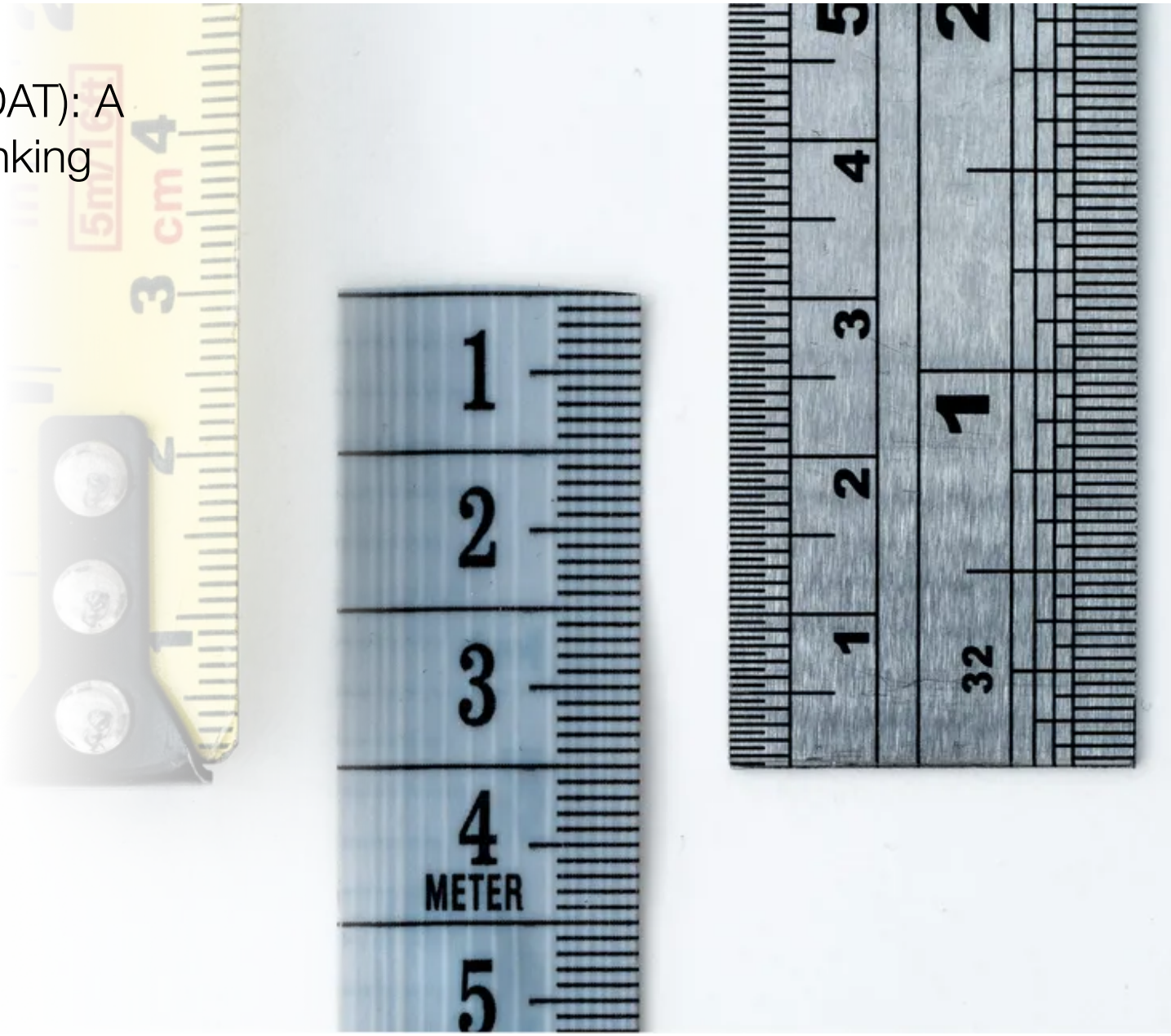
Why Leverage LLMs and GenAI for Creativity Measures?

- For **AUT,** LLM evaluation models correlate with human judges up to r = .813[1]
- LLMs yield **high inter-model agreement** on creativity ratings
- LLMs **automate evaluation** in real time—ideal for large-scale, multilingual creativity studies



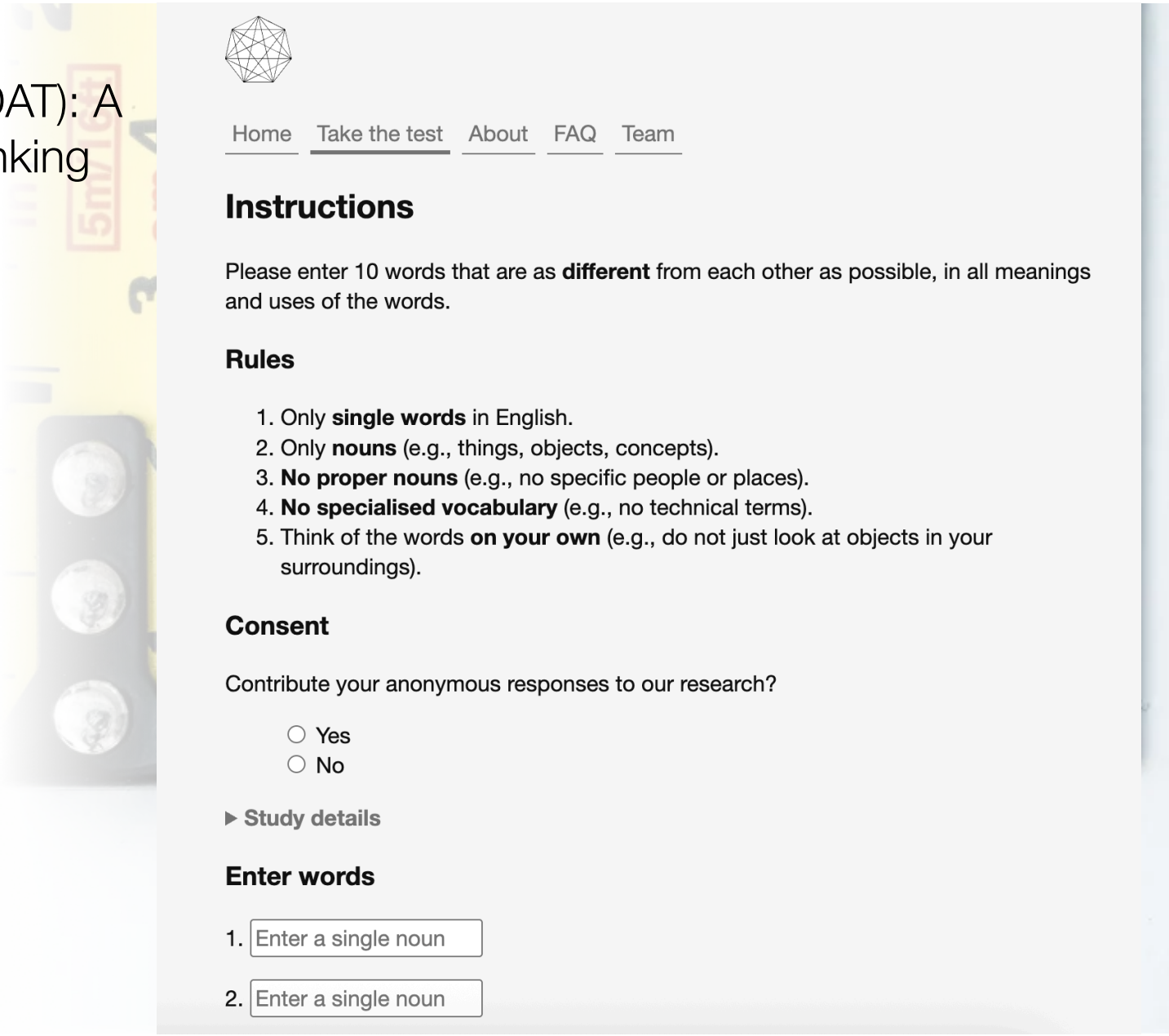
[1] OCSAI tool for AUT; Organisciak, Acar, Dumas, & Berthiaume (2023)

Divergent Association Task (DAT): A Simple Test for Divergent Thinking



Divergent Association Task (DAT): A Simple Test for Divergent Thinking

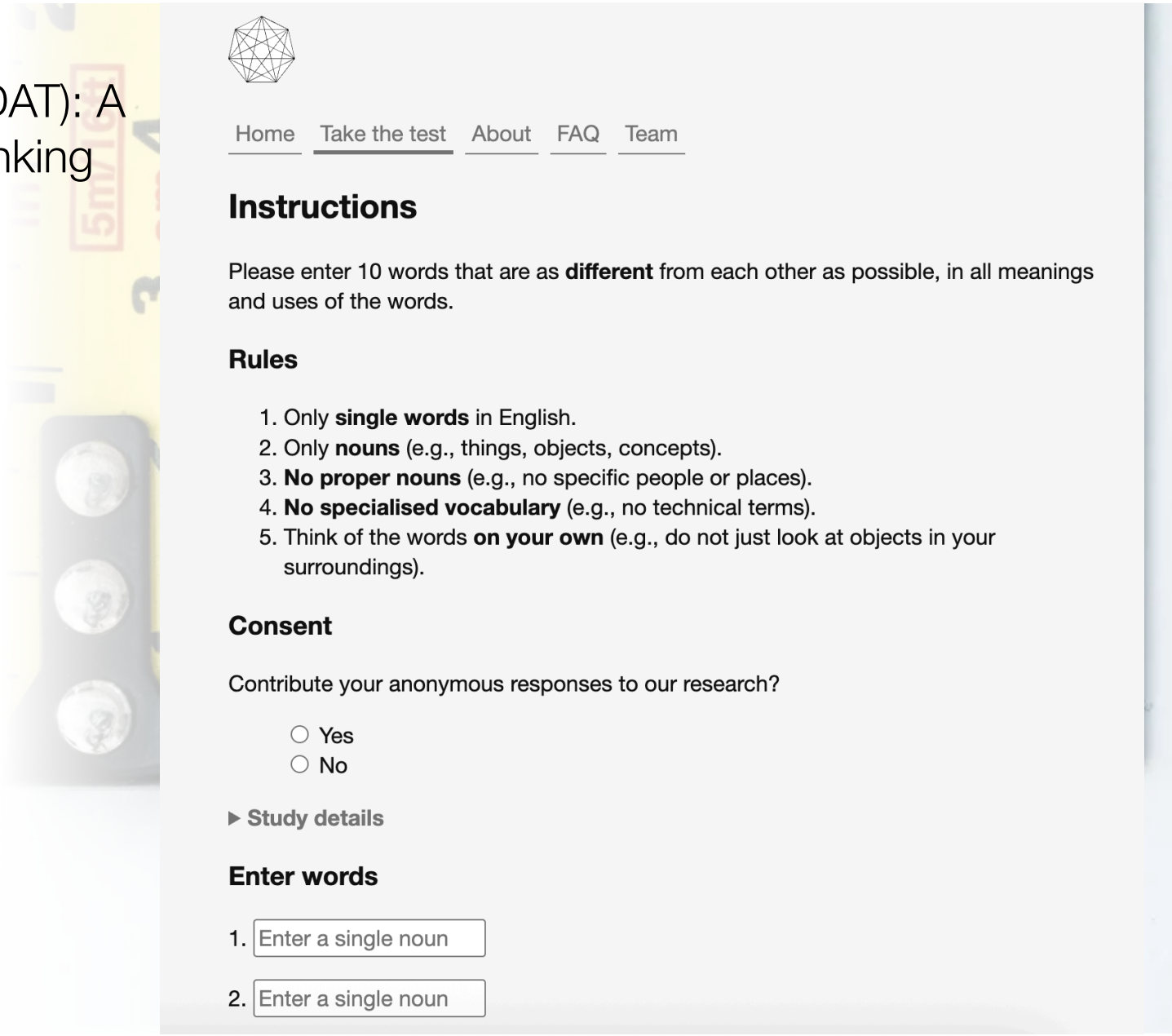
- Developed by Olson et al. (2021)



[1] https://www.datcreativity.com/

Divergent Association Task (DAT): A Simple Test for Divergent Thinking

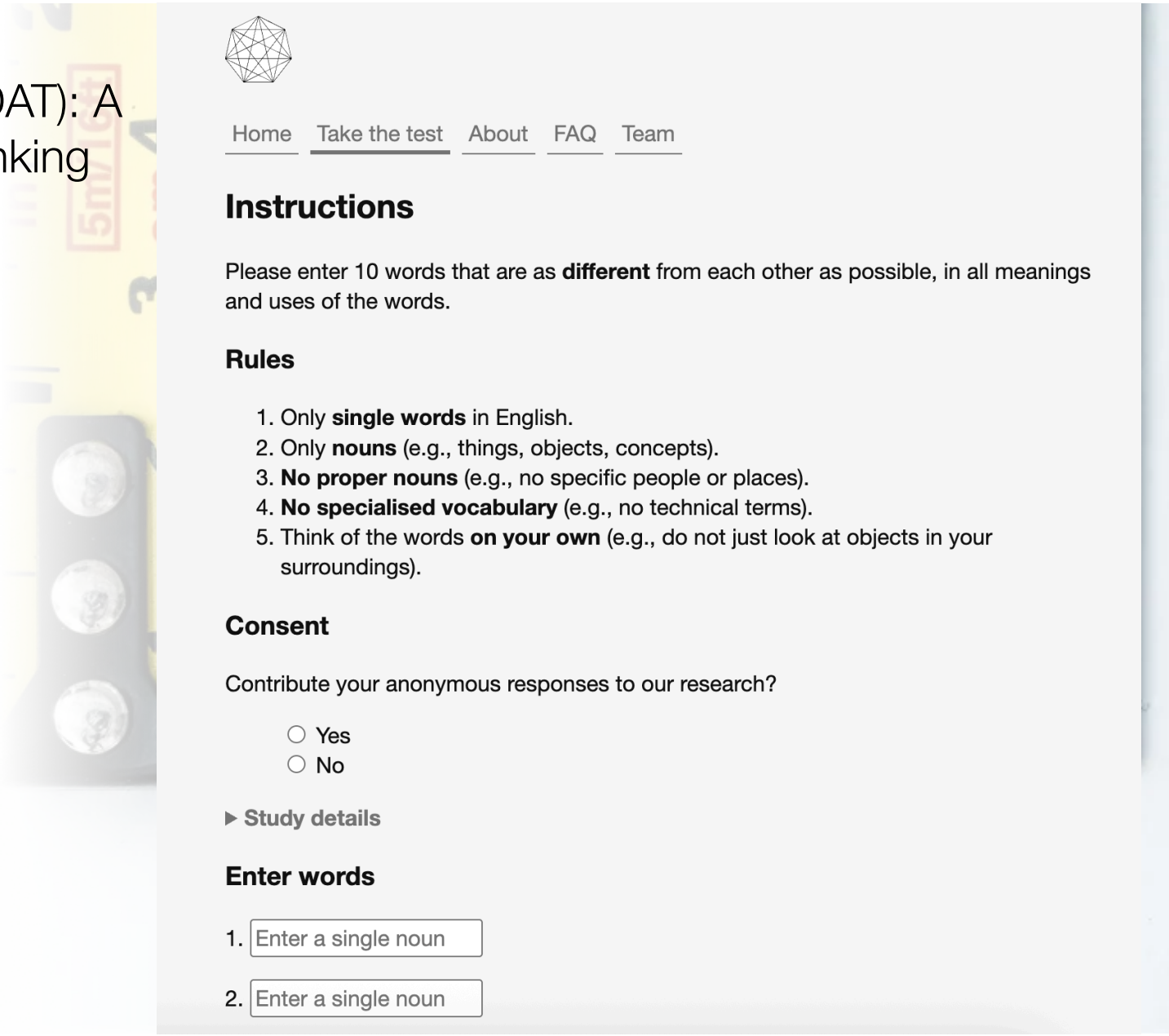
- Developed by Olson et al. (2021)
- Score = average semantic distance between word pairs



[1] https://www.datcreativity.com/

Divergent Association Task (DAT): A Simple Test for Divergent Thinking

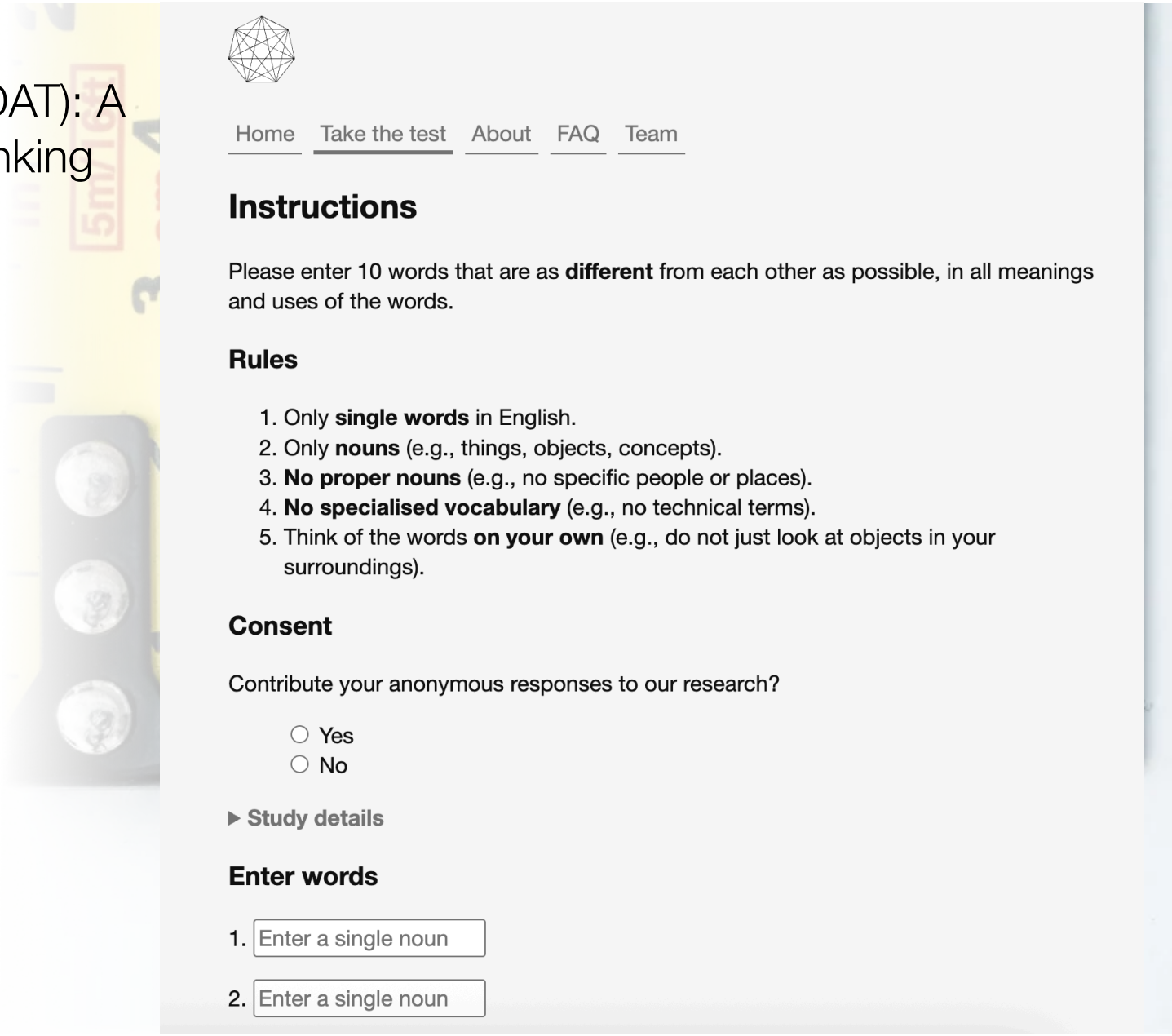
- Developed by Olson et al. (2021)
- Score = average semantic distance between word pairs
- Uses **GloVe embeddings** to calculate distance



[1] https://www.datcreativity.com/

Divergent Association Task (DAT): A Simple Test for Divergent Thinking

- Developed by Olson et al. (2021)
- Score = average semantic distance between word pairs
- Uses **GloVe embeddings** to calculate distance
- Strengths:
  - Fast
  - Low writing demand

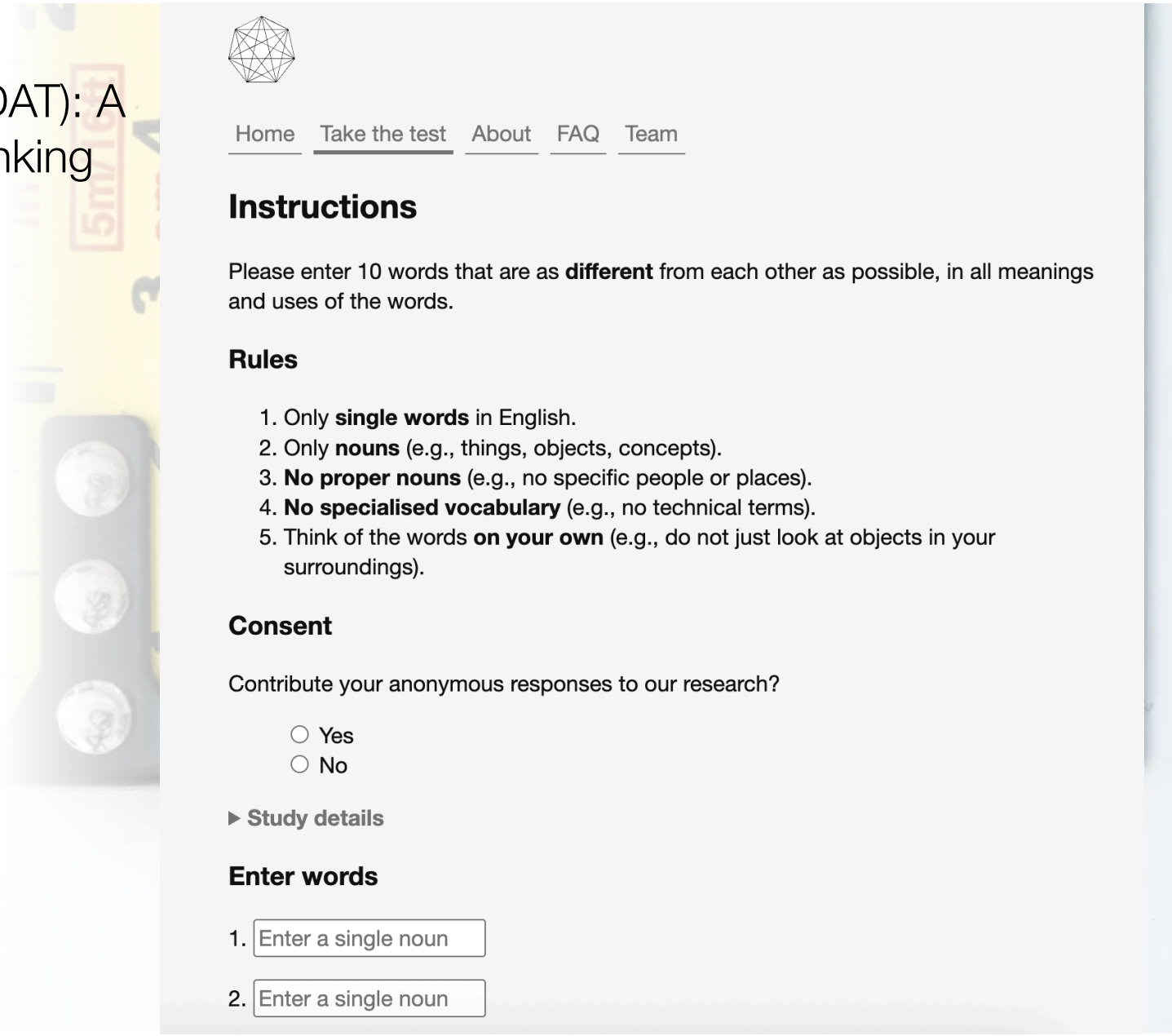


[1] https://www.datcreativity.com/

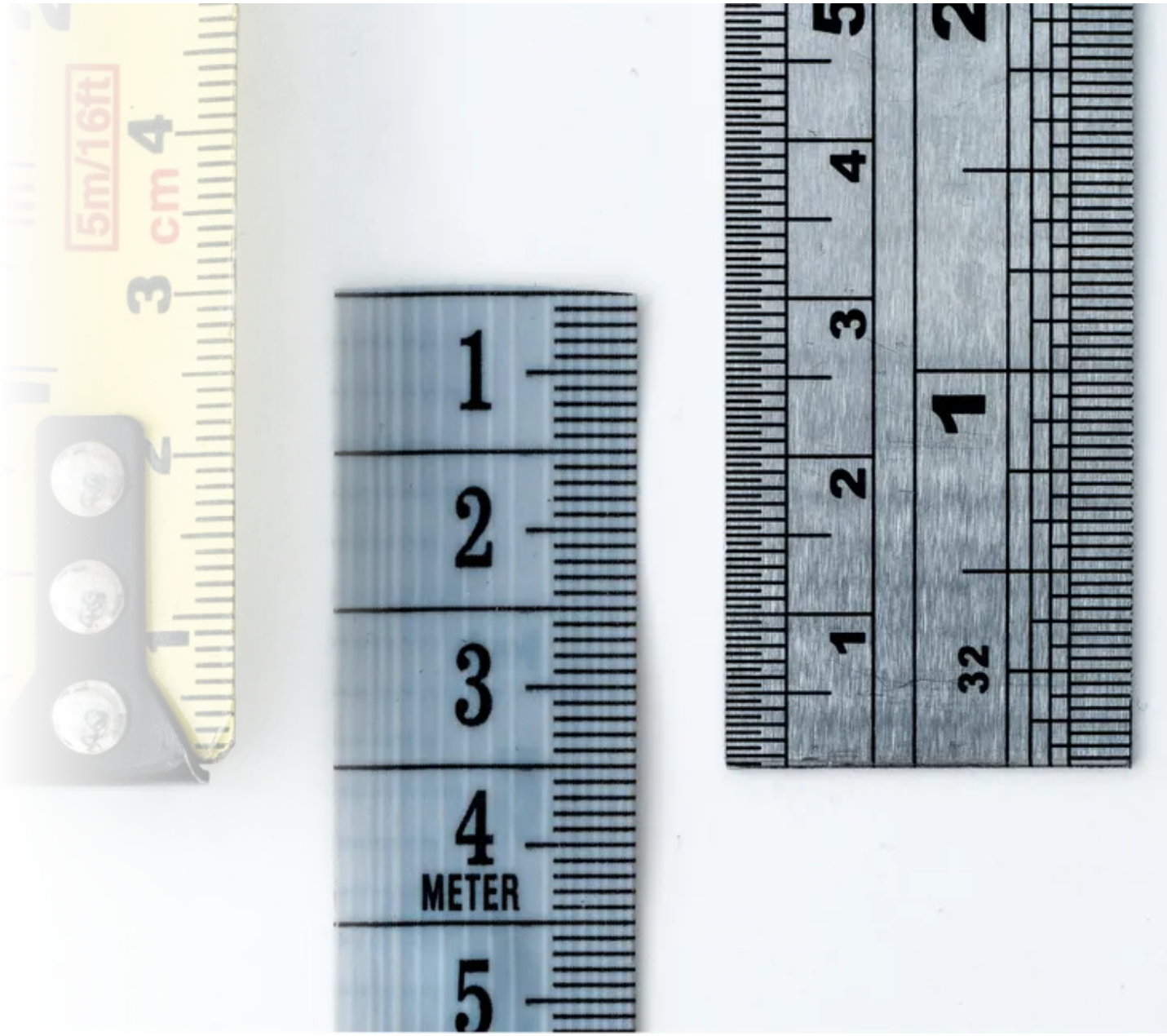
Divergent Association Task (DAT): A Simple Test for Divergent Thinking

- Developed by Olson et al. (2021)
- Score = average semantic distance between word pairs
- Uses **GloVe embeddings** to calculate distance
- Strengths:
  - Fast
  - Low writing demand
- Limitations:
  - English-only embeddings
  - Static, context-insensitive vectors
  - No cultural adaptability

[1] https://www.datcreativity.com/

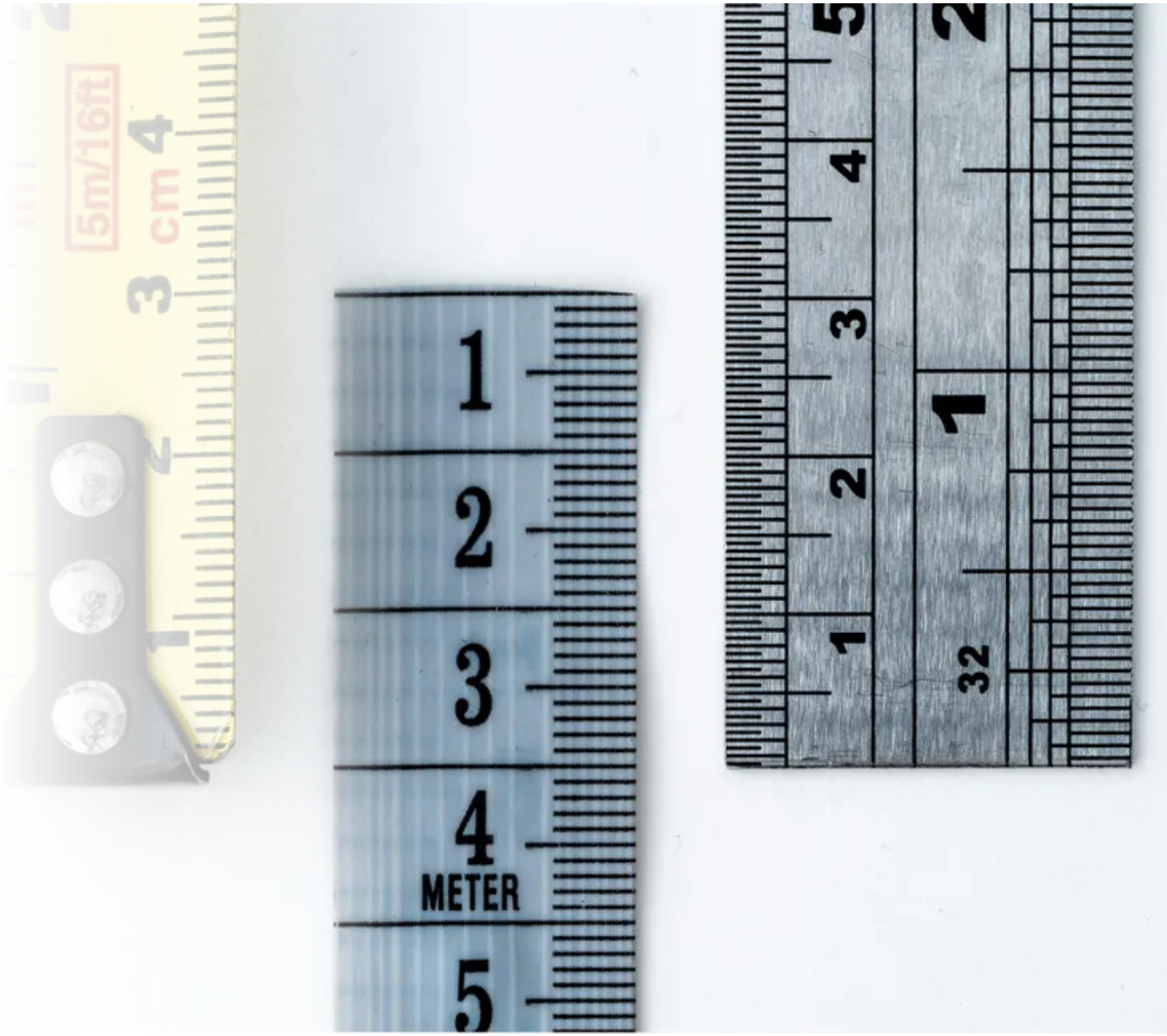


S-DAT: DAT-inspired Multilingual Scoring



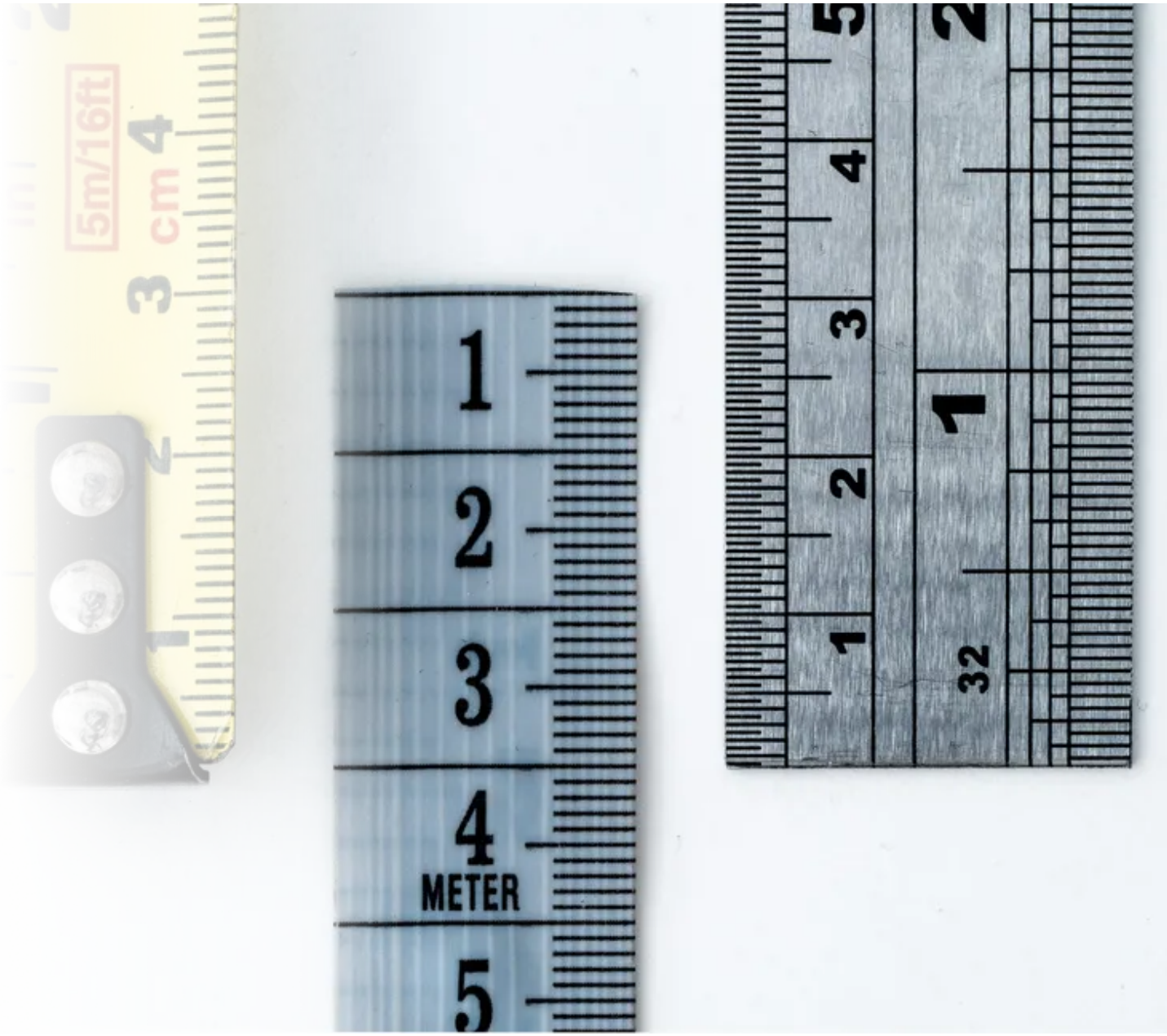
S-DAT: DAT-inspired Multilingual Scoring

- **Same core task as DAT**: “Name 10 unrelated words”



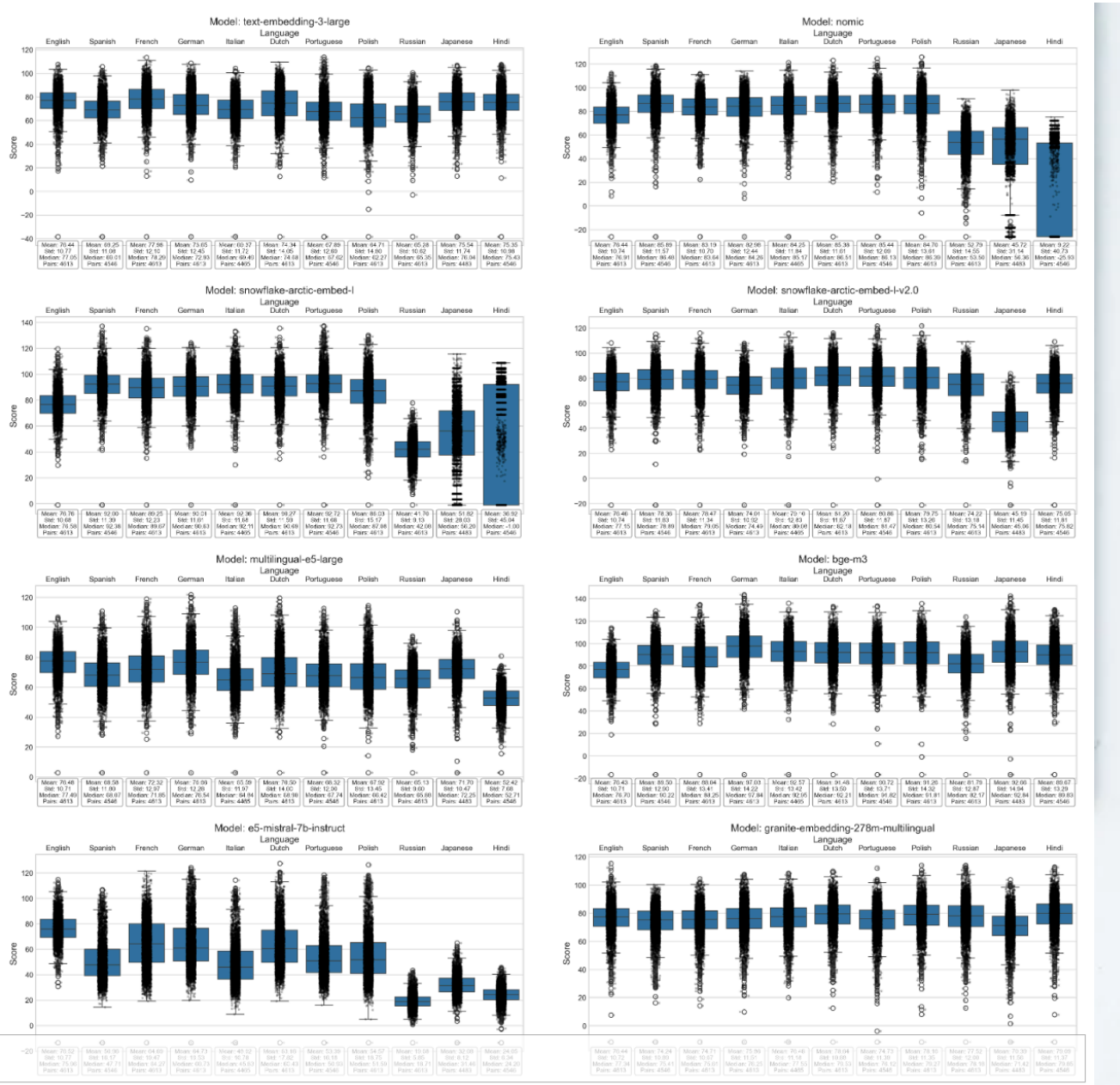
S-DAT: DAT-inspired Multilingual Scoring

- **Same core task as DAT**: “Name 10 unrelated words”
- Semantic distance calculated by multilingual, transformer-based embeddings (currently we use IBMs granite-embedding-278m-multilingual model)



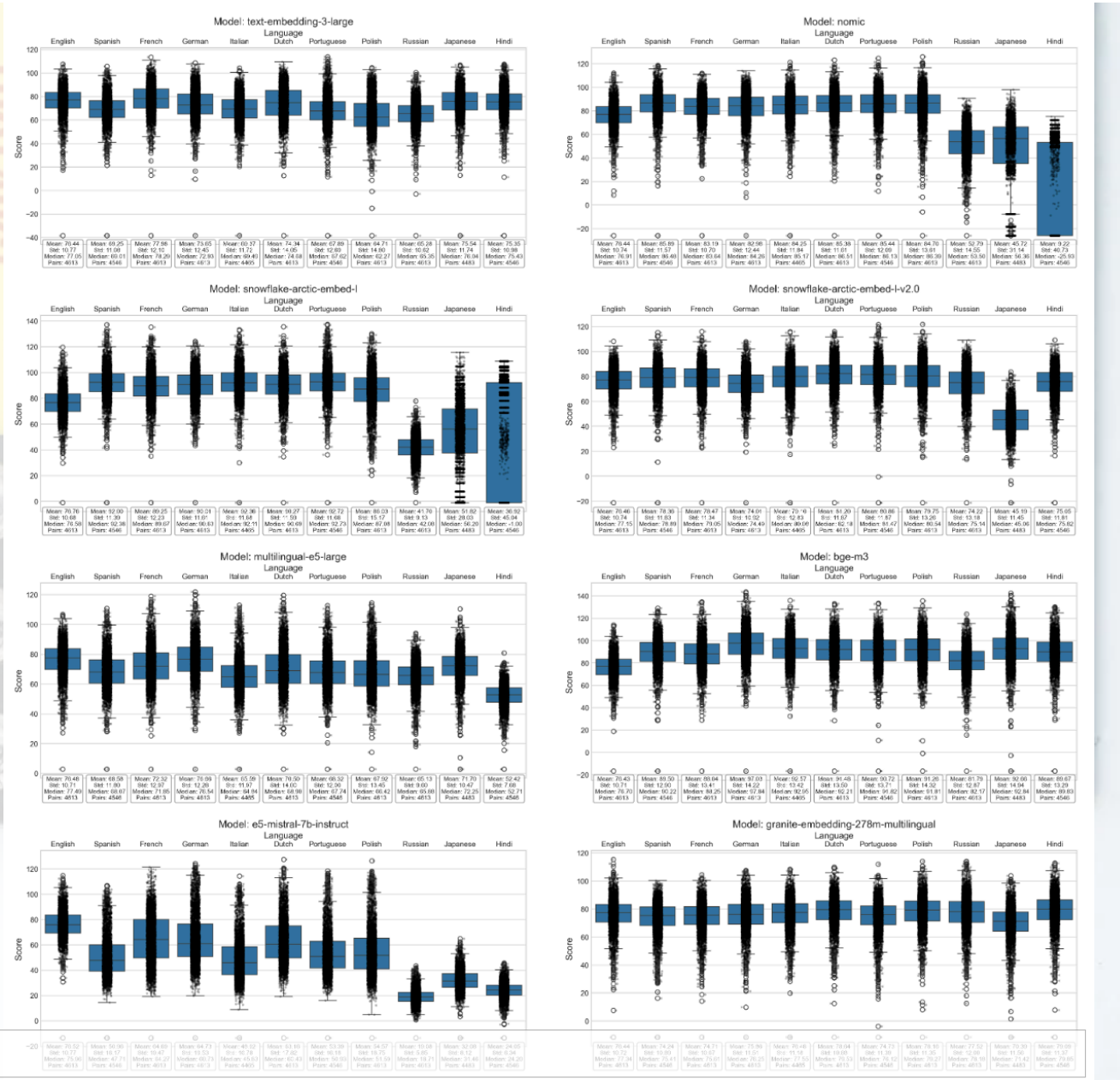
S-DAT: DAT-inspired Multilingual Scoring

- **Same core task as DAT**: “Name 10 unrelated words”
- Semantic distance calculated by multilingual, transformer-based embeddings (currently we use IBMs granite-embedding-278m-multilingual model)
- **Works natively in 12 fine-tuned languages**— including English, Hindi, Japanese, Korean, Arabic, Chinese, and Russian

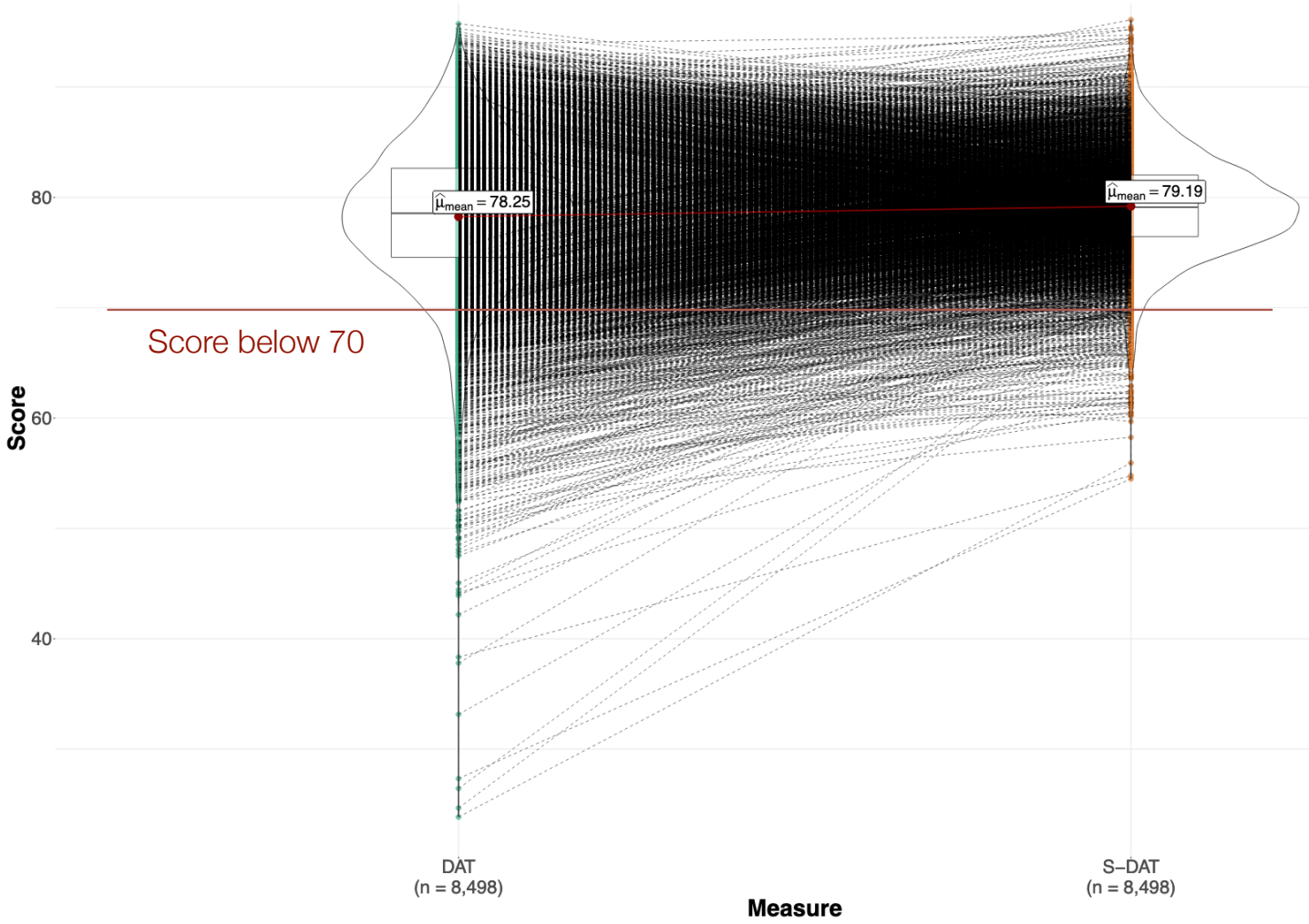


S-DAT: DAT-inspired Multilingual Scoring

- **Same core task as DAT**: “Name 10 unrelated words”
- Semantic distance calculated by multilingual, transformer-based embeddings (currently we use IBMs granite-embedding-278m-multilingual model)
- **Works natively in 12 fine-tuned languages**— including English, Hindi, Japanese, Korean, Arabic, Chinese, and Russian
- Works even across languages in one entry

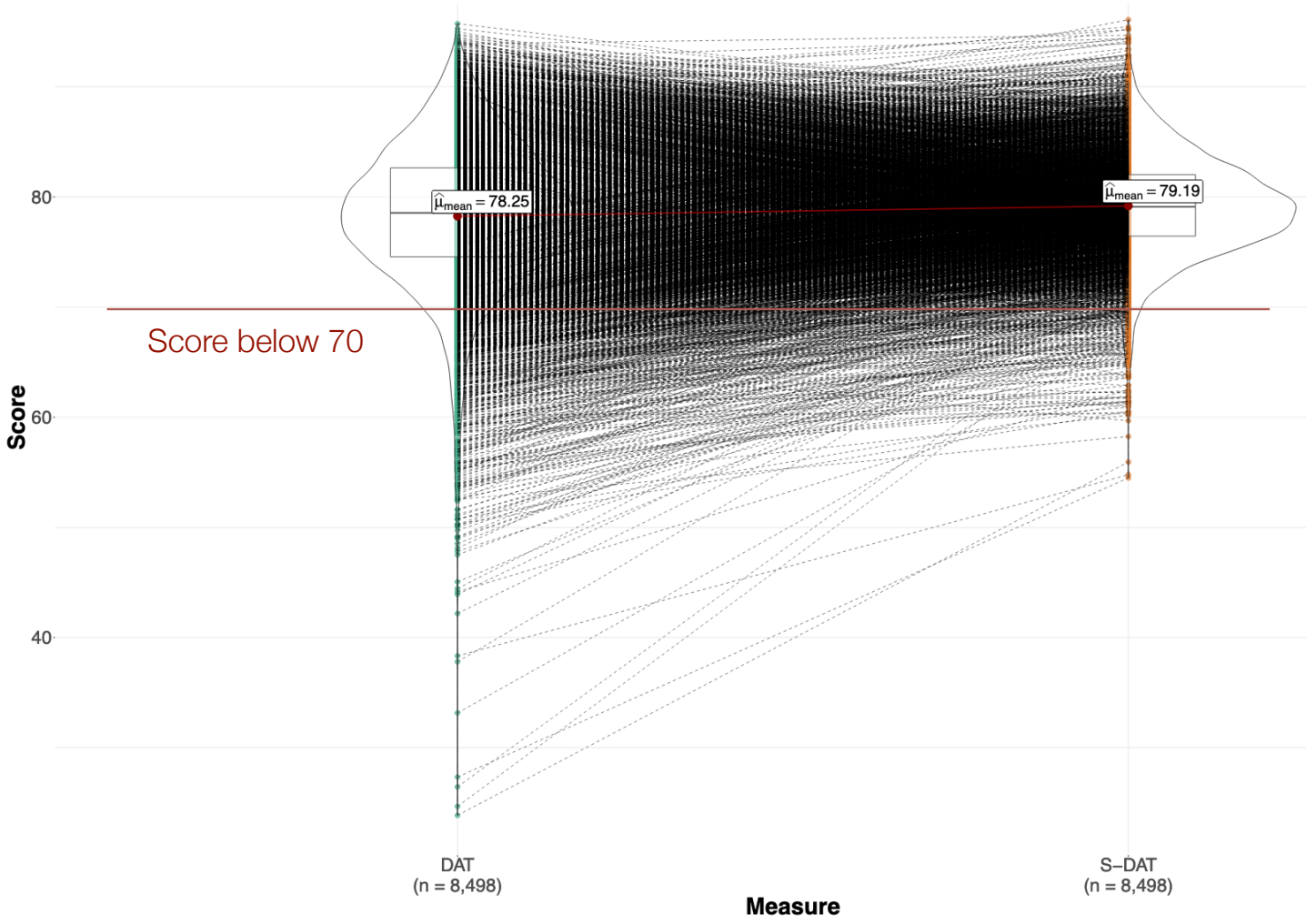


S-DAT: in Calibration to the DAT



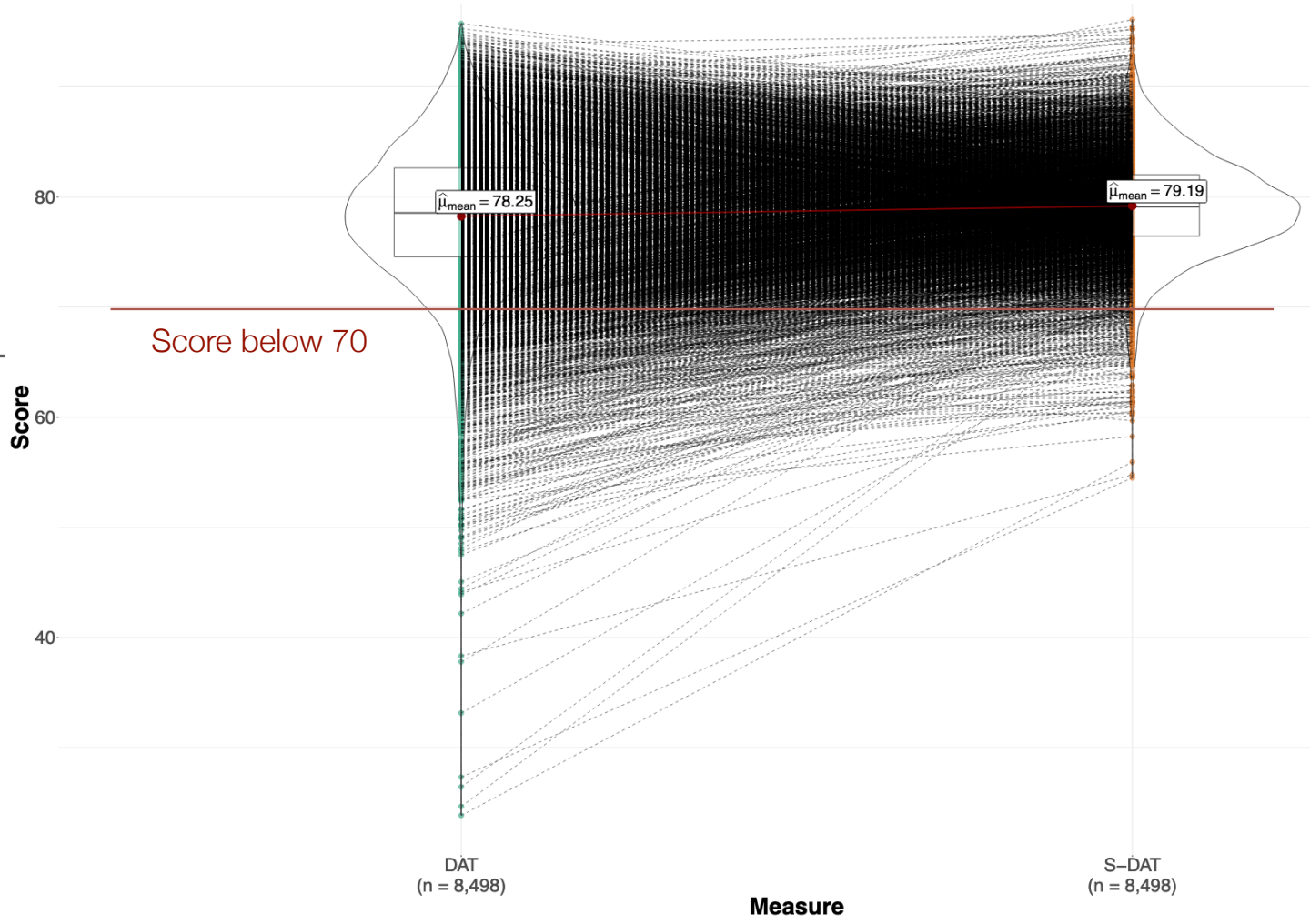
S-DAT: in Calibration to the DAT

- Based analysis on original Olson et al (2021) data, Study 2 with n = 8,498 entries



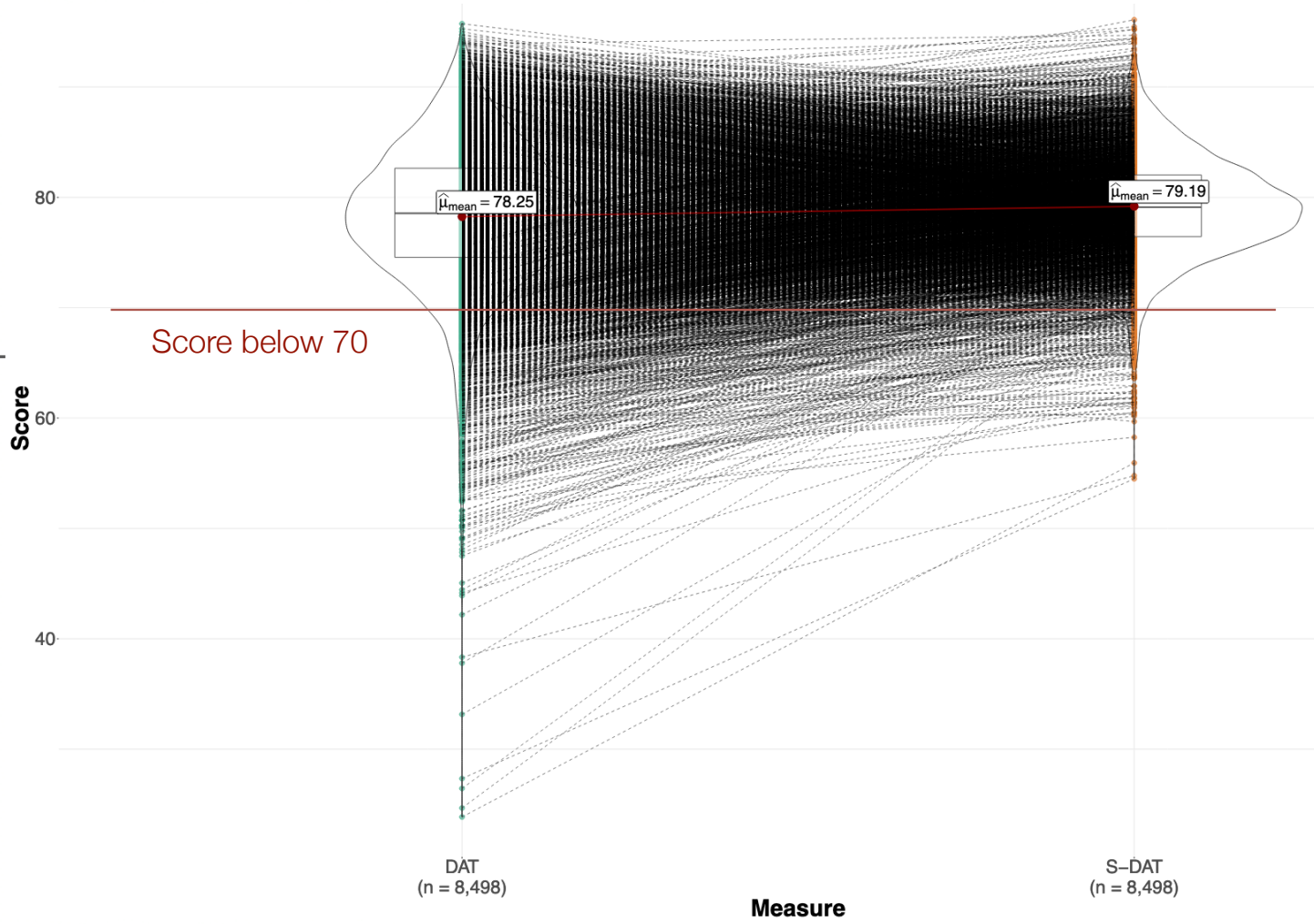
S-DAT: in Calibration to the DAT

- Based analysis on original Olson et al (2021) data, Study 2 with n = 8,498 entries
- S-DAT distribution a bit steeper — scores below 70 are “noise”



S-DAT: in Calibration to the DAT

- Based analysis on original Olson et al (2021) data, Study 2 with n = 8,498 entries
- S-DAT distribution a bit steeper — scores below 70 are “noise”
- Only normed for English yet



S-DAT: in Comparison to other C-measures

Study	Measure	DAT	S-DAT
Olson et al. Study 1a (ns = 138–141)			
	DAT		.67*** [.56, .75]
	AUT: Originality	.32*** [.16, .46]	.17* [−.00, .17]
	AUT: Flexibility	.34*** [.18, .48]	.19* [.03, .35]
	AUT: Fluency	.22** [.06, .37]	.14* [−.03, .30]
	Bridge-the-Associative-Gap Task	.22** [.06, .38]	.11 [−.06, .28]
Olson et al. Study 1b (ns = 205–284)			
	DAT		.60*** [.52, .67]
	AUT: Originality	.32*** [.20, .43]	.24*** [.11, .36]
	AUT: Flexibility	.35*** [.23, .46]	.27*** [.15, .39]
	AUT: Fluency	.30*** [.17, .41]	.16** [.03, .28]
	Bridge-the-Associative-Gap Task	.23*** [.10, .36]	.08 [−.06, .22]
Olson et al. Study 2 (ns = 355–8,498)			
	DAT		.65*** [.64, .66]
	AUT: Originality	.13** [.03, .23]	.13** [.02, .23]

S-DAT: in Comparison to other C-measures

- Bridge-the-Associative-Gap Task = RAT-extension = convergent thinking

Study	Measure	DAT	S-DAT
Olson et al. Study 1a (ns = 138–141)			
	DAT		.67*** [.56, .75]
	AUT: Originality	.32*** [.16, .46]	.17* [−.00, .17]
	AUT: Flexibility	.34*** [.18, .48]	.19* [.03, .35]
	AUT: Fluency	.22** [.06, .37]	.14* [−.03, .30]
	Bridge-the-Associative-Gap Task	.22** [.06, .38]	.11 [−.06, .28]
Olson et al. Study 1b (ns = 205–284)			
	DAT		.60*** [.52, .67]
	AUT: Originality	.32*** [.20, .43]	.24*** [.11, .36]
	AUT: Flexibility	.35*** [.23, .46]	.27*** [.15, .39]
	AUT: Fluency	.30*** [.17, .41]	.16** [.03, .28]
	Bridge-the-Associative-Gap Task	.23*** [.10, .36]	.08 [−.06, .22]
Olson et al. Study 2 (ns = 355–8,498)			
	DAT		.65*** [.64, .66]
	AUT: Originality	.13** [.03, .23]	.13** [.02, .23]

S-DAT: in Comparison to other C-measures

- Bridge-the-Associative-Gap Task = RAT-extension = convergent thinking
- Relatively low correlations with AUT

Study	Measure	DAT	S-DAT
Olson et al. Study 1a (ns = 138–141)			
	DAT		.67*** [.56, .75]
	AUT: Originality	.32*** [.16, .46]	.17* [−.00, .17]
	AUT: Flexibility	.34*** [.18, .48]	.19* [.03, .35]
	AUT: Fluency	.22** [.06, .37]	.14* [−.03, .30]
	Bridge-the-Associative-Gap Task	.22** [.06, .38]	.11 [−.06, .28]
Olson et al. Study 1b (ns = 205–284)			
	DAT		.60*** [.52, .67]
	AUT: Originality	.32*** [.20, .43]	.24*** [.11, .36]
	AUT: Flexibility	.35*** [.23, .46]	.27*** [.15, .39]
	AUT: Fluency	.30*** [.17, .41]	.16** [.03, .28]
	Bridge-the-Associative-Gap Task	.23*** [.10, .36]	.08 [−.06, .22]
Olson et al. Study 2 (ns = 355–8,498)			
	DAT		.65*** [.64, .66]
	AUT: Originality	.13** [.03, .23]	.13** [.02, .23]

S-DAT: in Comparison to other C-measures

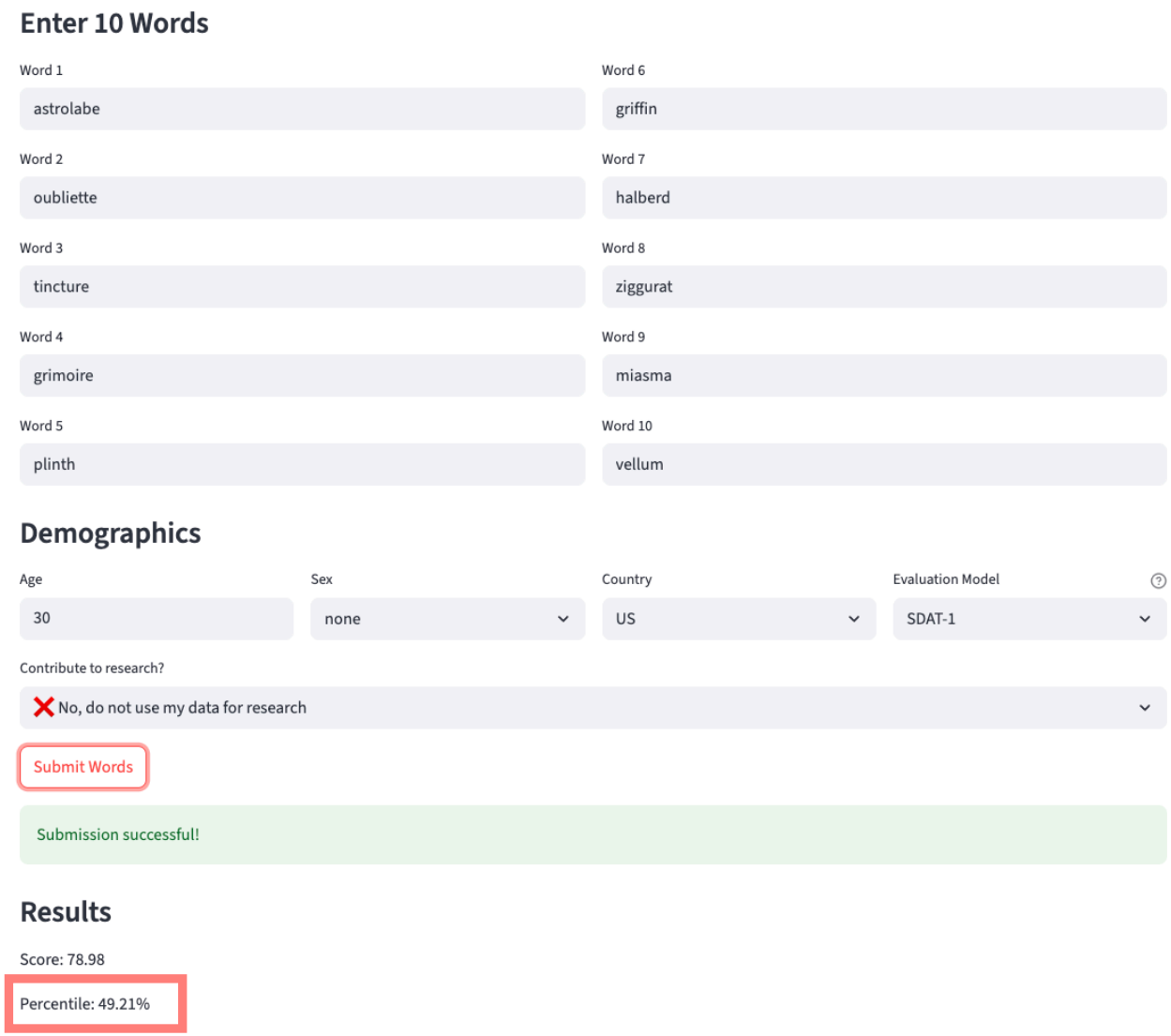
- Bridge-the-Associative-Gap Task = RAT-extension = convergent thinking
- Relatively low correlations with AUT
- Better discriminant validity for convergent thinking

Study	Measure	DAT	S-DAT
Olson et al. Study 1a (ns = 138–141)			
	DAT		.67*** [.56, .75]
	AUT: Originality	.32*** [.16, .46]	.17* [−.00, .17]
	AUT: Flexibility	.34*** [.18, .48]	.19* [.03, .35]
	AUT: Fluency	.22** [.06, .37]	.14* [−.03, .30]
	Bridge-the-Associative-Gap Task	.22** [.06, .38]	.11 [−.06, .28]
Olson et al. Study 1b (ns = 205–284)			
	DAT		.60*** [.52, .67]
	AUT: Originality	.32*** [.20, .43]	.24*** [.11, .36]
	AUT: Flexibility	.35*** [.23, .46]	.27*** [.15, .39]
	AUT: Fluency	.30*** [.17, .41]	.16** [.03, .28]
	Bridge-the-Associative-Gap Task	.23*** [.10, .36]	.08 [−.06, .22]
Olson et al. Study 2 (ns = 355–8,498)			
	DAT		.65*** [.64, .66]
	AUT: Originality	.13** [.03, .23]	.13** [.02, .23]

S-DAT: Examples for its Stability

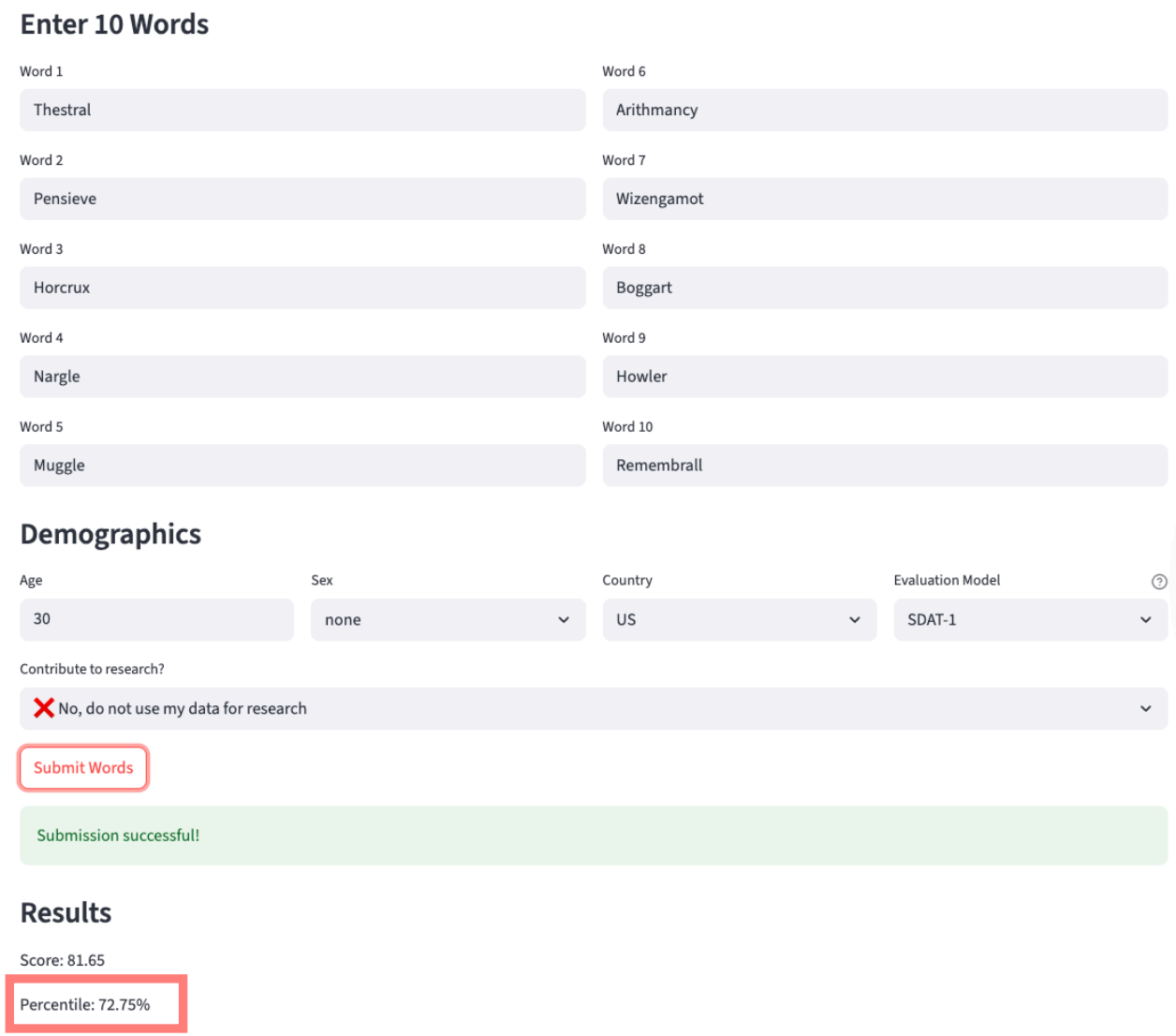
S-DAT: Examples for its Stability

- Using ultra-rare words



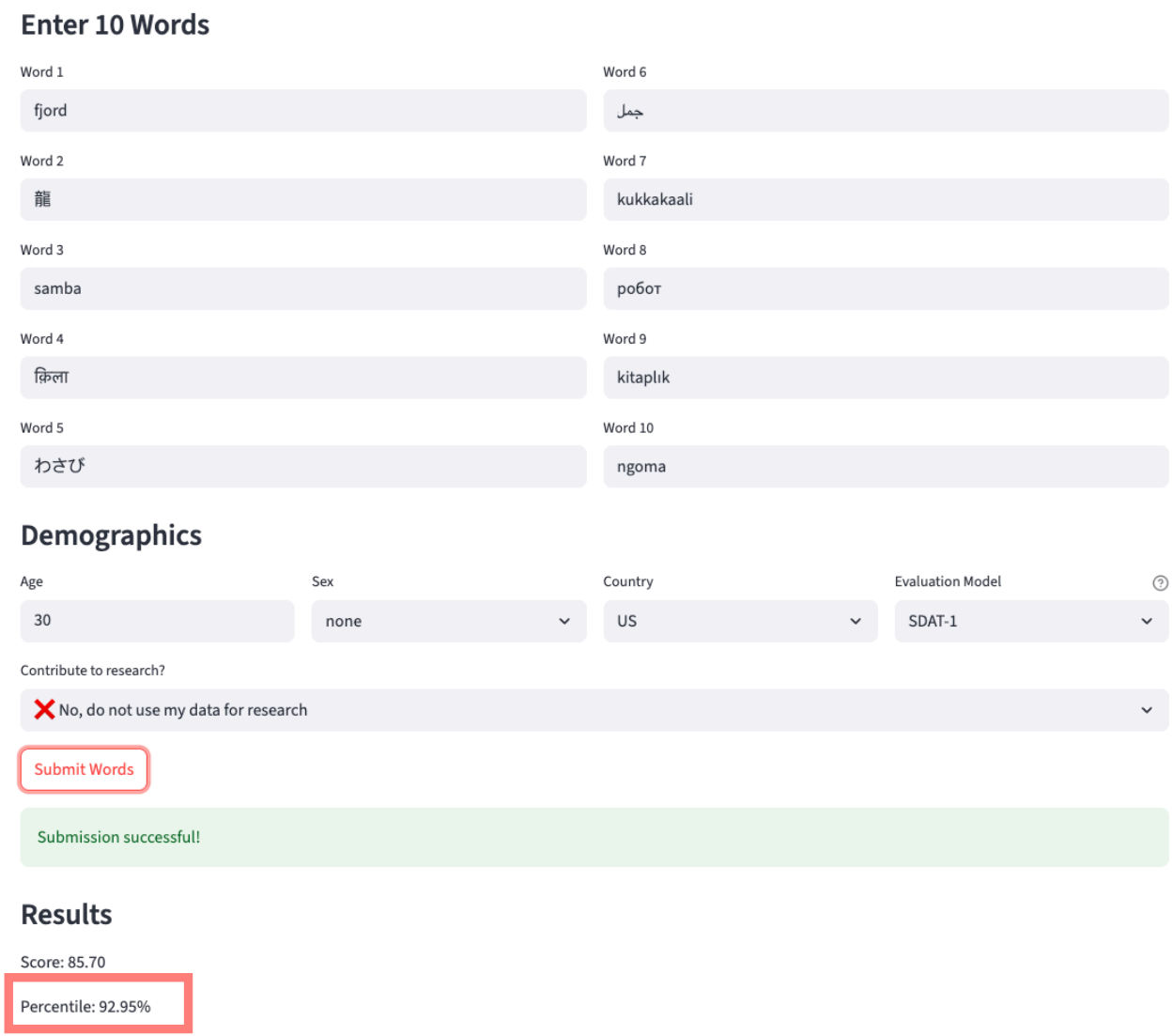
S-DAT: Examples for its Stability

- Using ultra-rare words
- Using very niche words, e.g. from Harry Potter universe



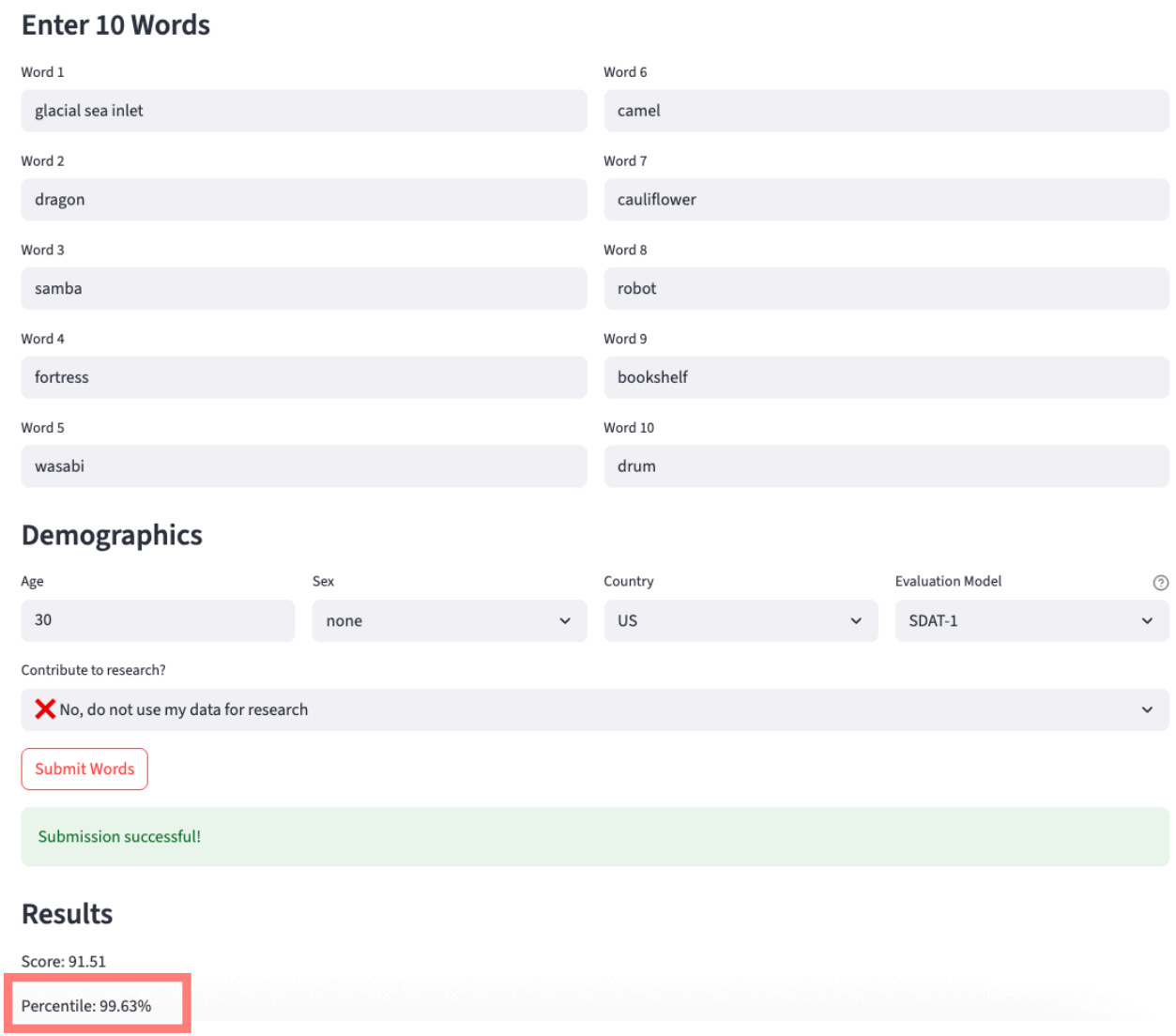
S-DAT: Examples for its Stability

- Using ultra-rare words
- Using very niche words, e.g. from Harry Potter universe
- Language-Hopping (= language-independent representation)



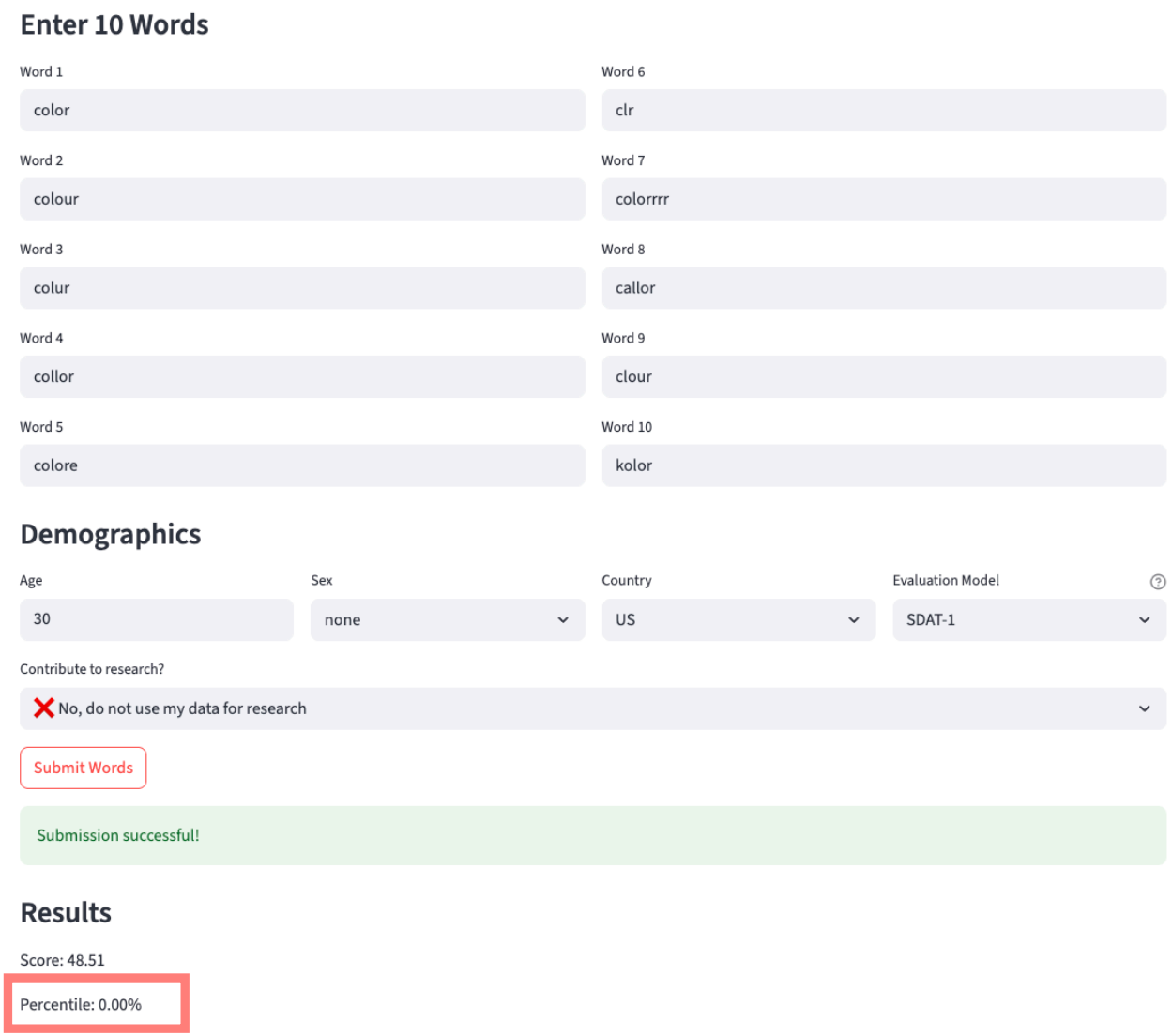
S-DAT: Examples for its Stability

- Using ultra-rare words
- Using very niche words, e.g. from Harry Potter universe
- Language-Hopping (= language-independent representation)



S-DAT: Examples for its Stability

- Using ultra-rare words
- Using very niche words, e.g. from Harry Potter universe
- Language-Hopping (= language-independent representation)
- Misspellings



IOL Lab
at ZIB and TUB

English

- English
- Arabic
- Deutsch
- Français
- Italiano
- Japanese
- Turkish

Your submitted data will only be used for research purposes if you explicitly consent to it. It does not include any personal information (no IP addresses, no names, no email addresses) and as such no personal information will be shared with third parties.

[ZIB's Privacy Policy and Imprint]

Research supported by

University of Essex

Powered by

RISE
powered by Zuse Institute Berlin

Multi-Lingual Divergent Association Task

How creatively do you think?

The Divergent Association Task (DAT) is a quick, research-based test that gives you insight into your creative potential.

Your task: Enter 10 words—each as different in meaning and association as possible. The more unrelated the words, the better.

Why it works:

The DAT measures divergent thinking, a core aspect of creativity. It does this by analyzing the semantic distance between the words you provide. Greater distance suggests greater cognitive flexibility---a key trait in creative thought.

Good to know:
The DAT works in multiple languages! Basically, any language and script is supported, although results might vary across languages due availability of training data.

Ready to explore how your mind makes connections?

Give it a try below!

Read more (english only)

Submit Your Words

Enter 10 Words

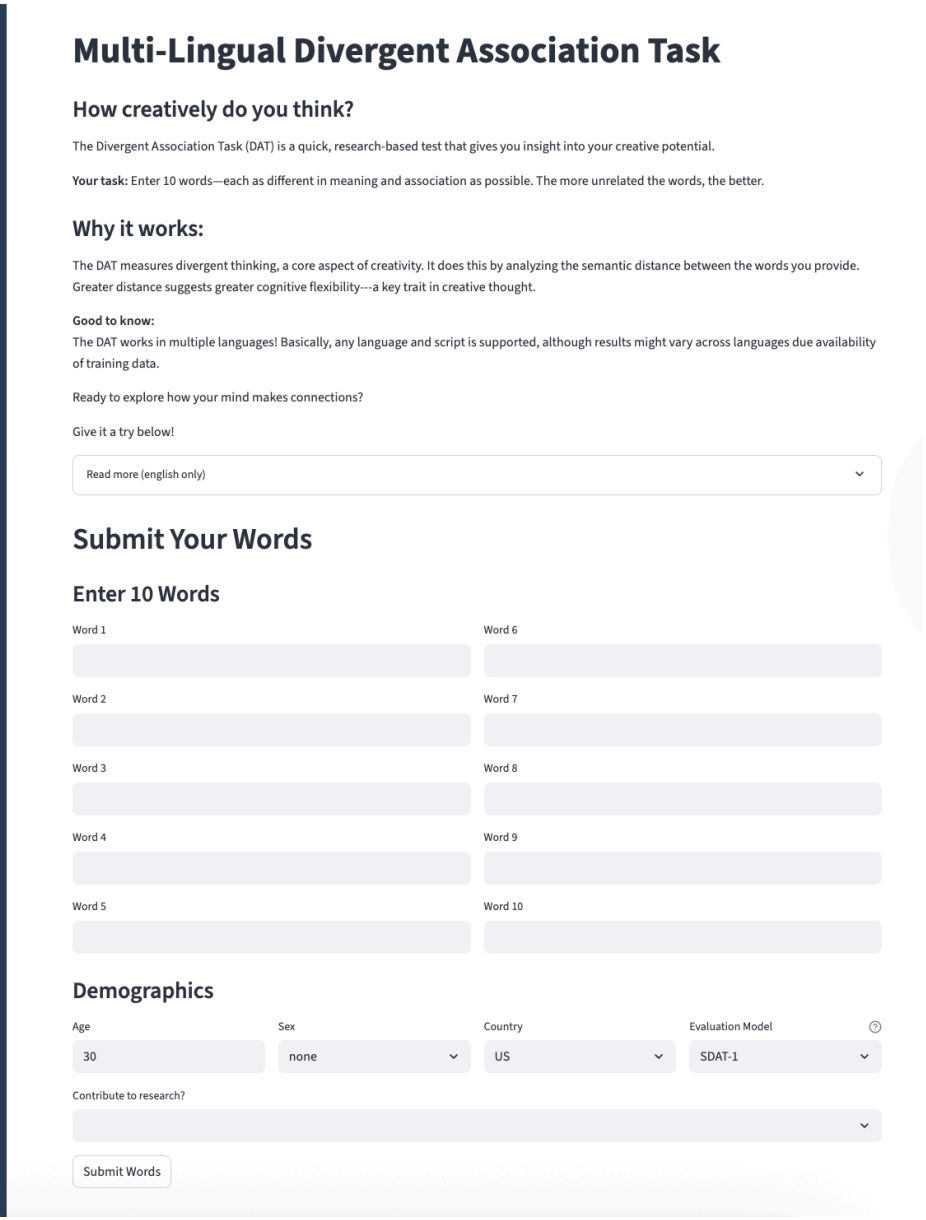
Word 1
Word 2
Word 3
Word 4
Word 5
Word 6
Word 7
Word 8
Word 9
Word 10

Demographics

Age: 30
Sex: none
Country: US
Evaluation Model: SDAT-1

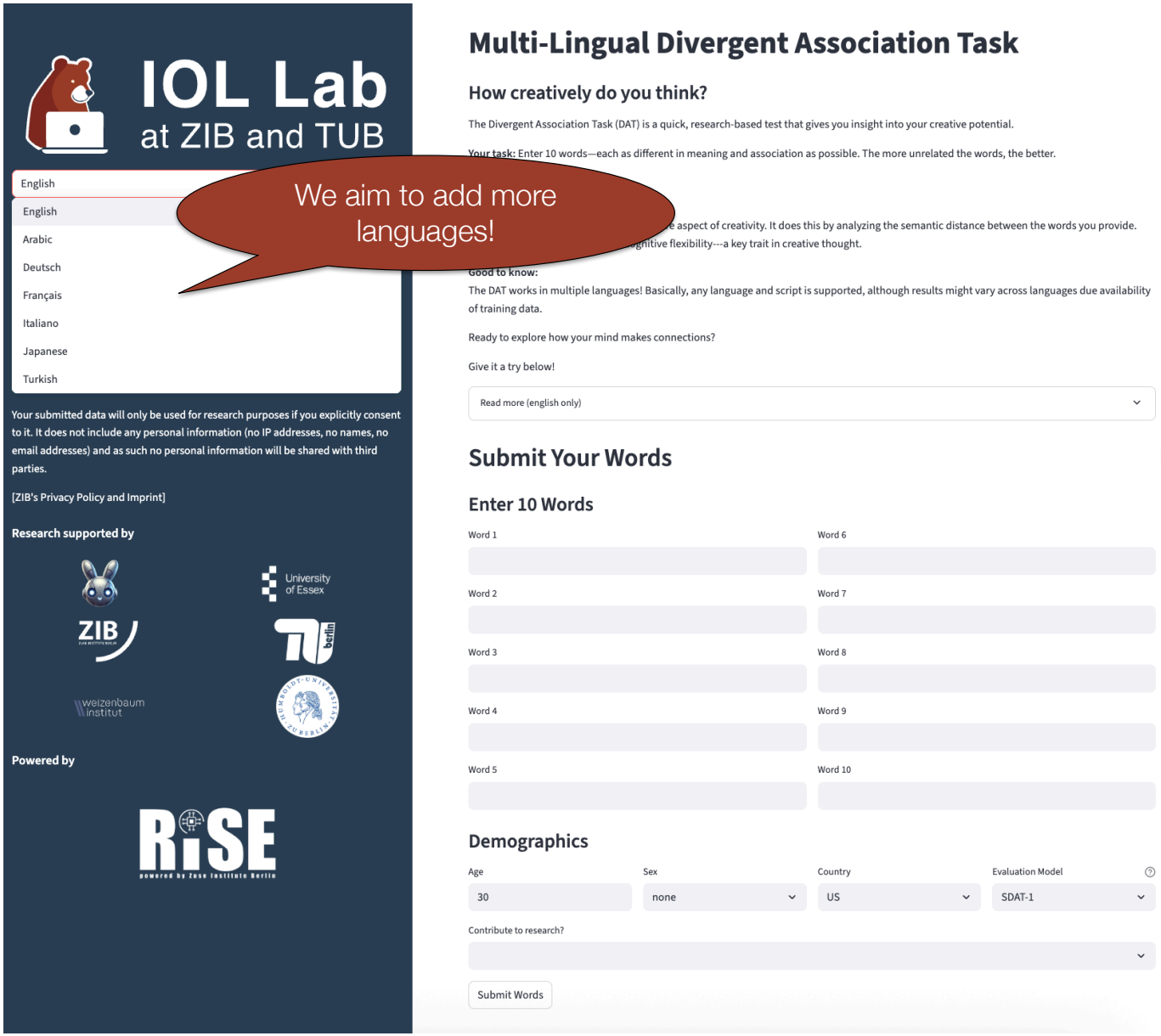
Contribute to research?

Submit Words



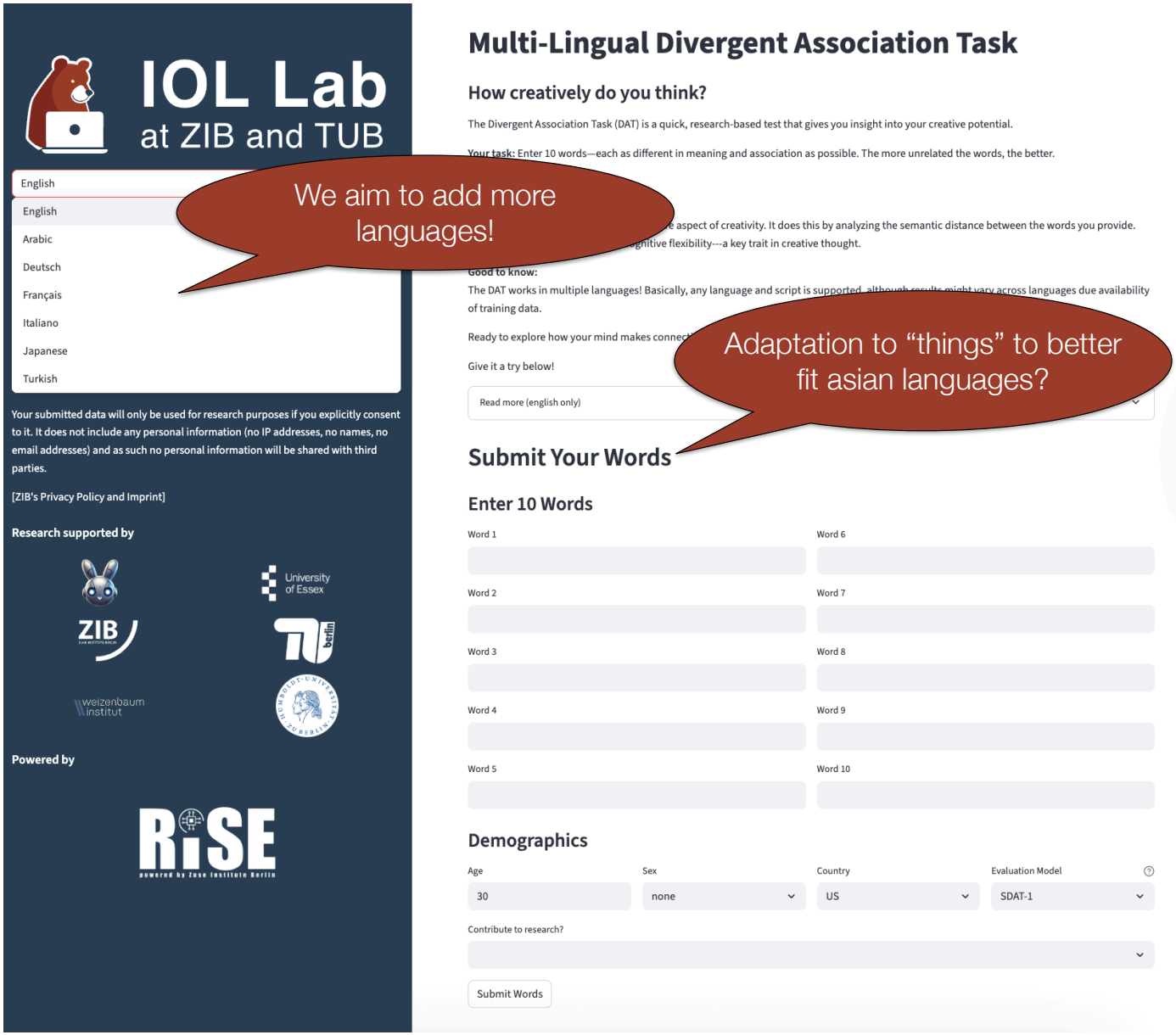
S-DAT

https://sdat.iol.zib.de/



S-DAT

https://sdat.iol.zib.de/



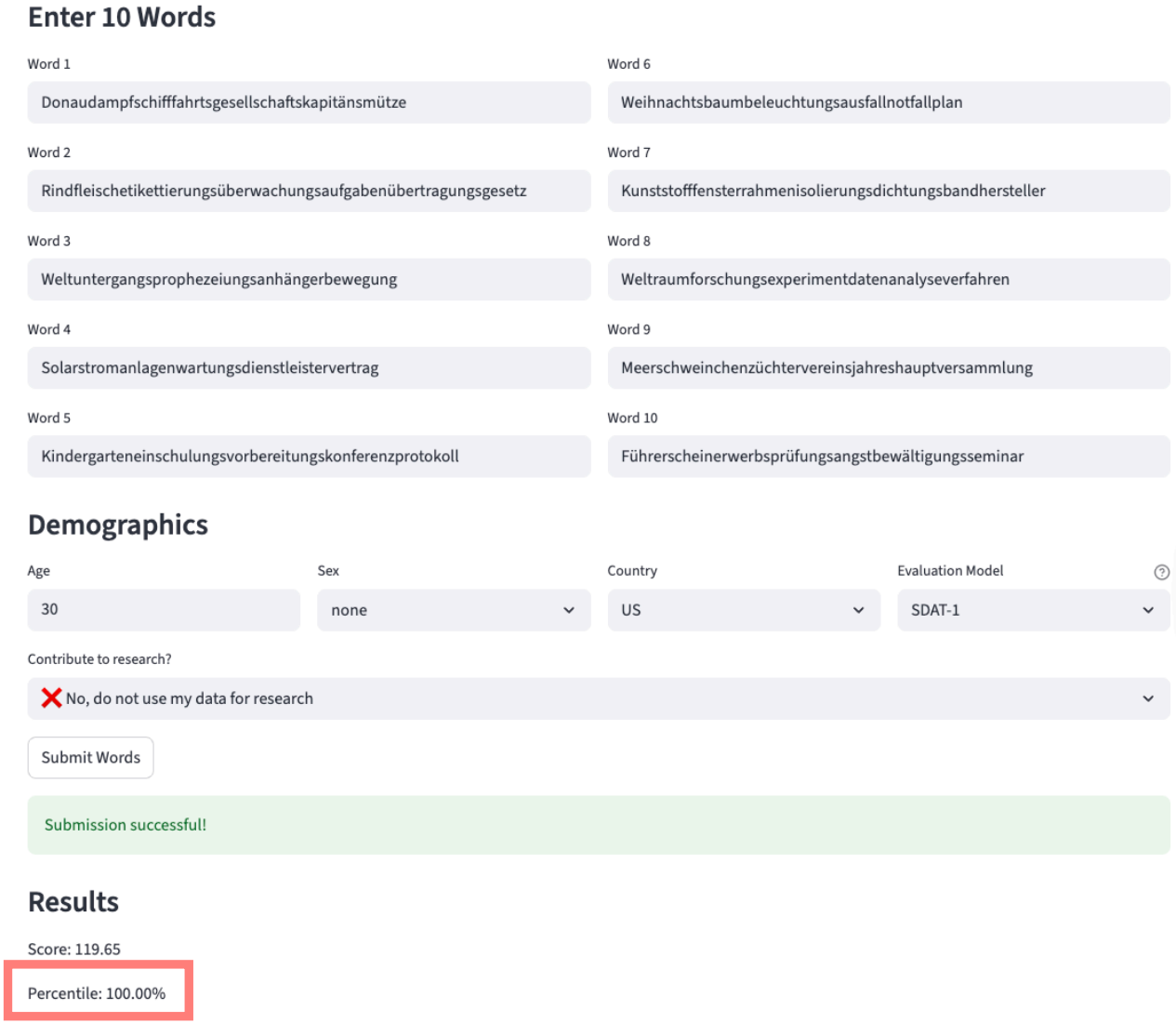
S-DAT

https://sdat.iol.zib.de/

S-DAT: Gaming the System

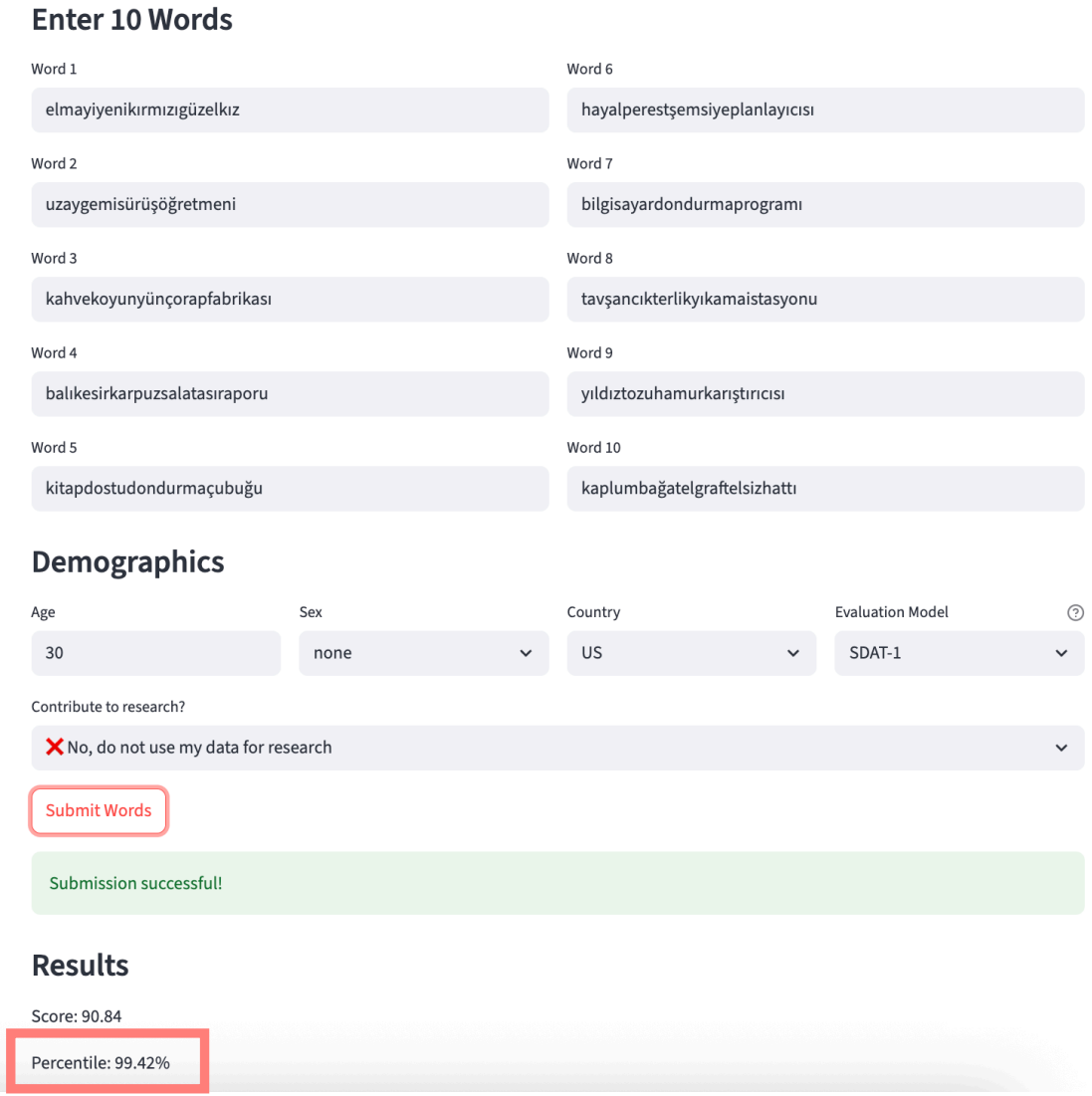
S-DAT: Gaming the System

- Other languages allow to create high scores easily —> e.g., due to compound words, e.g., in German, Finnish



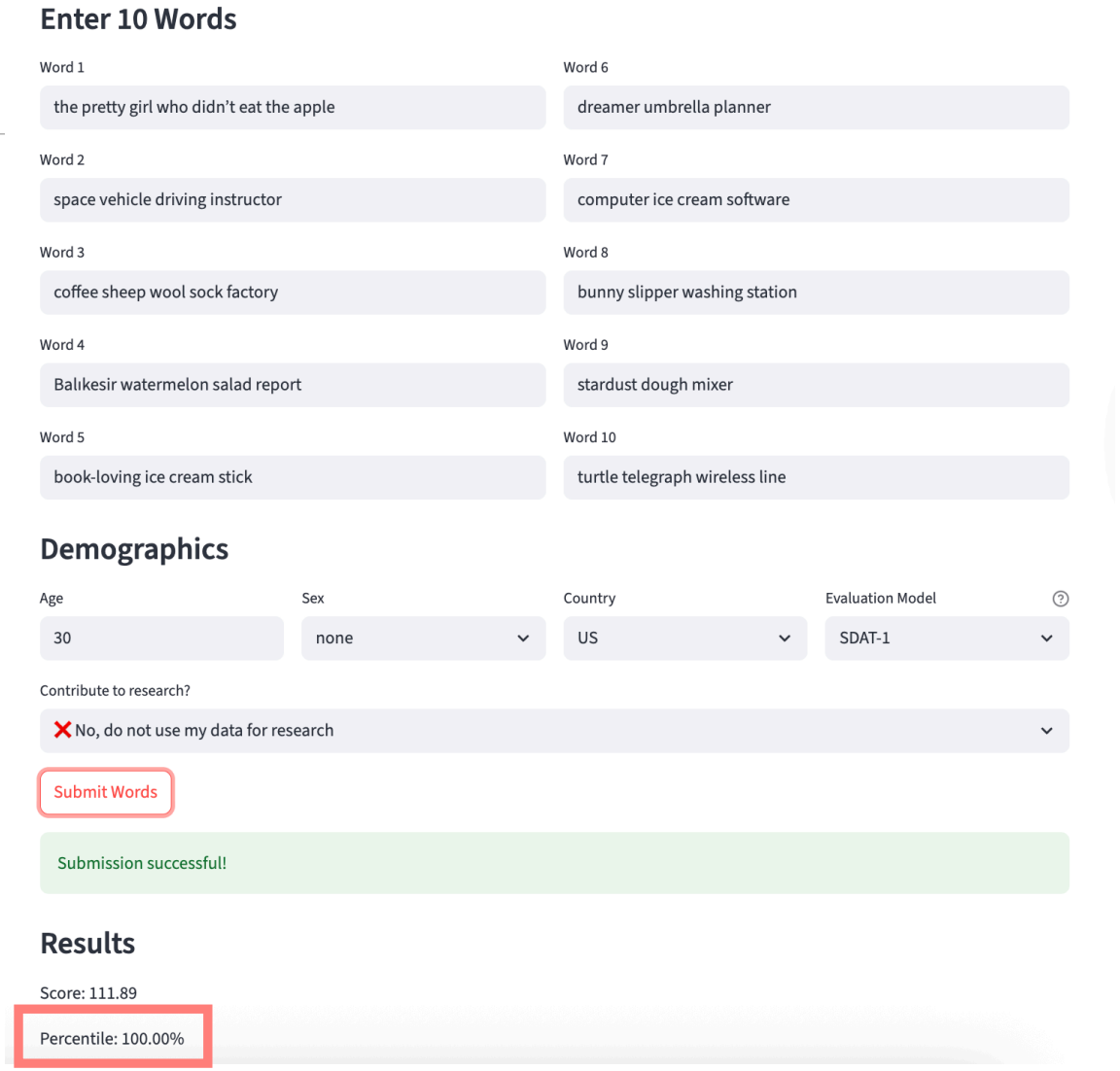
S-DAT: Gaming the System

- Other languages allow to create high scores easily —> e.g., due to compound words, e.g., in German, Finnish
- Suffix-chaining to create rare words, e.g., Turkish



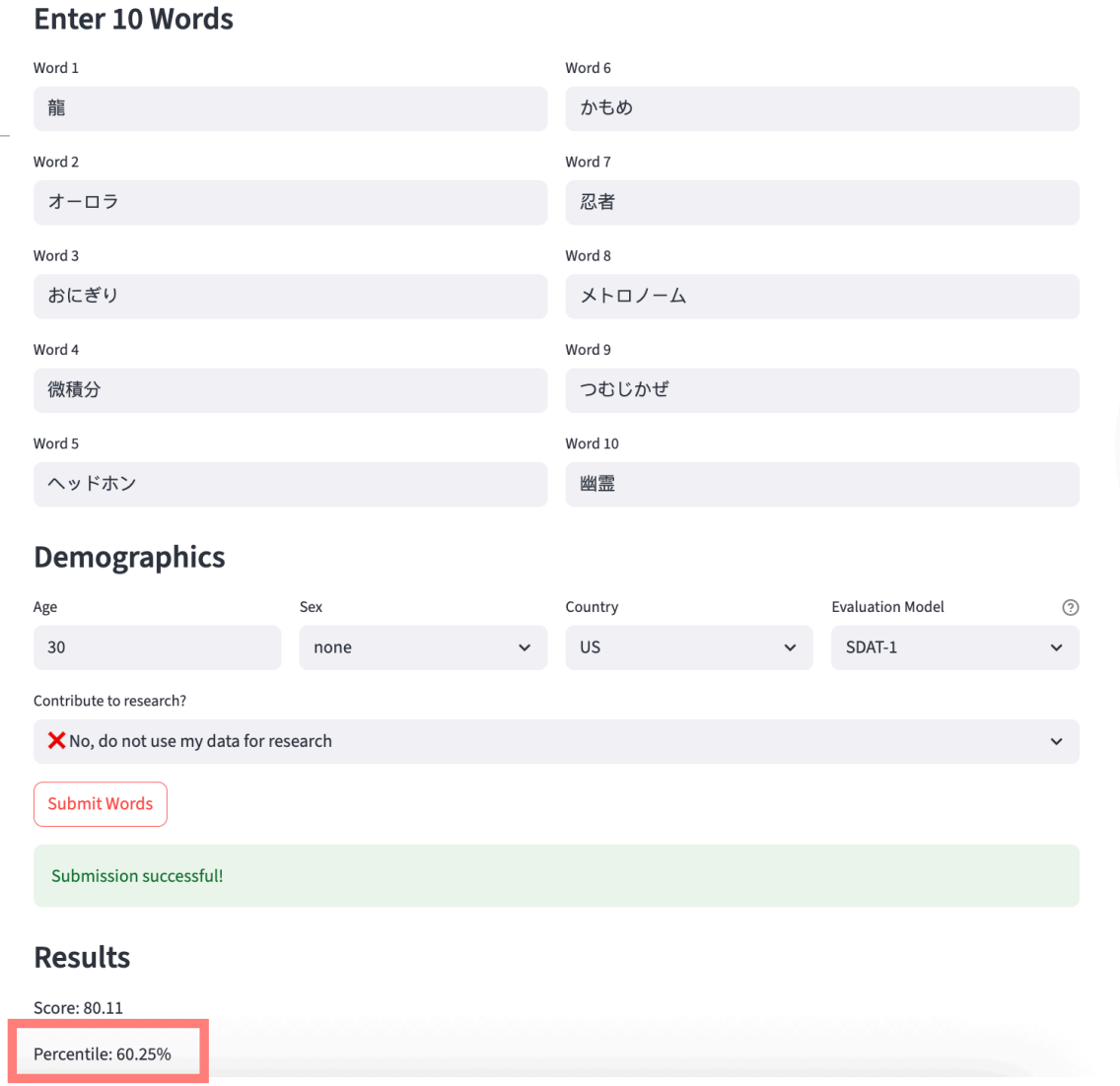
S-DAT: Gaming the System

- Other languages allow to create high scores easily —> e.g., due to compound words, e.g., in German, Finnish
- Suffix-chaining to create rare words, e.g., Turkish



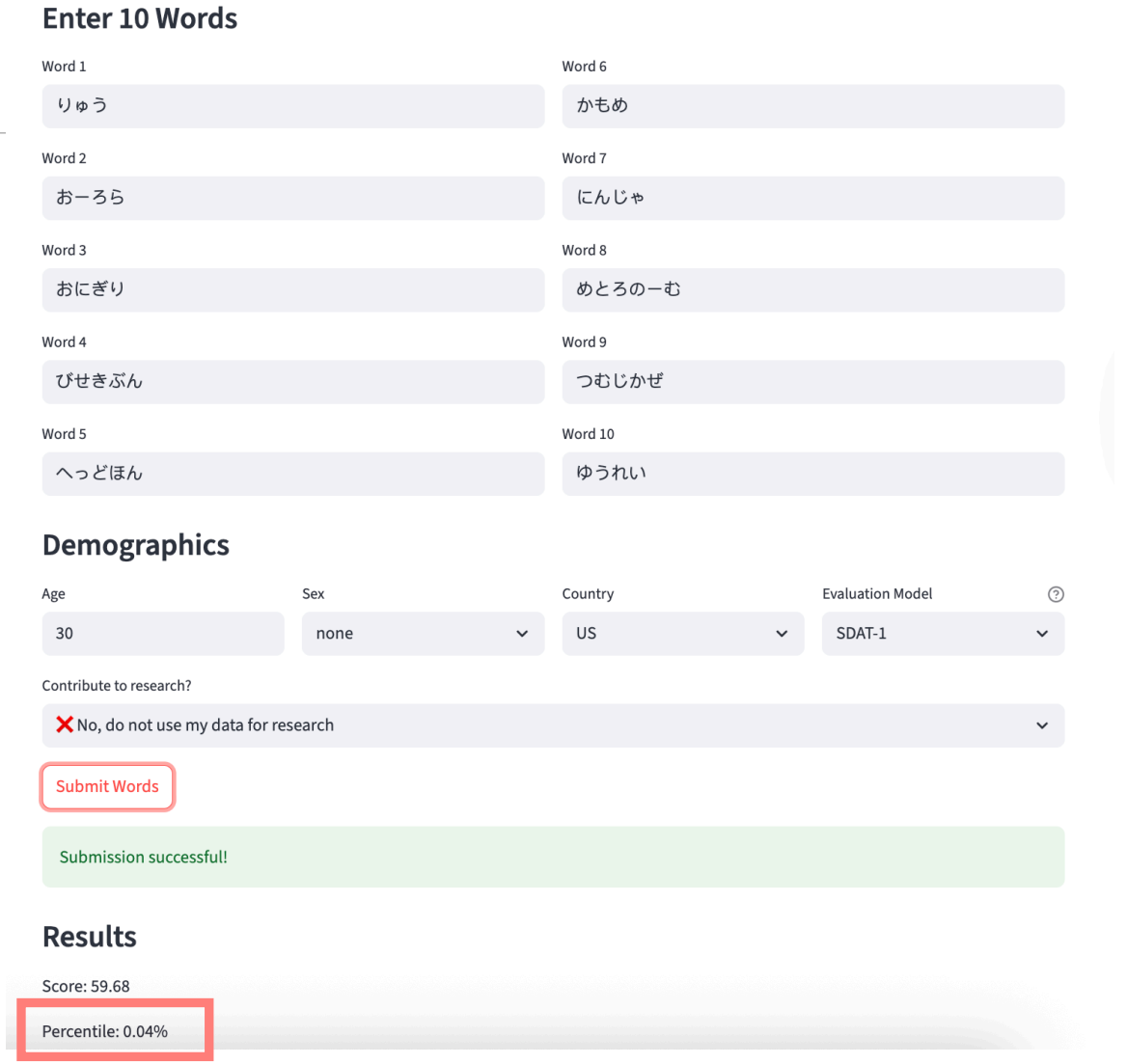
S-DAT: Gaming the System

- Other languages allow to create high scores easily —> e.g., due to compound words, e.g., in German, Finnish
- Suffix-chaining to create rare words, e.g., Turkish
- Mix between Scripts within a language, e.g., Japanese



S-DAT: Gaming the System

- Other languages allow to create high scores easily —> e.g., due to compound words, e.g., in German, Finnish
- Suffix-chaining to create rare words, e.g., Turkish
- Mix between Scripts within a language, e.g., Japanese



Looking for Collaboration!

https://sdat.iol.zib.de/

S-DAT

Looking for Collaboration!

- Join S-DAT Validation Project

https://sdat.iol.zib.de/

S-DAT

Looking for Collaboration!

- Join S-DAT Validation Project

https://sdat.iol.zib.de/

S-DAT

Looking for Collaboration!

- Join S-DAT Validation Project
  - Add a Language
  - Assess Data for Validation
  - Just approach us or via email: jennifer.haase@hu-berlin.de

https://sdat.iol.zib.de/

S-DAT

Literature

1. Haase, J., Hanel, P. H. P., & Pokutta, S. (2025). S-DAT: A Multilingual, GenAI-Driven Framework for Automated Divergent Thinking Assessment (No. arXiv:2505.09068). arXiv. https://doi.org/10.48550/arXiv.2505.09068

2. Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, *118*(25), e2022340118. https://doi.org/10.1073/pnas.2022340118

3. Organisciak, P., Acar, S., Dumas, D., & Berthiaume, K. (2023). Beyond Semantic Distance: Automated Scoring of Divergent Thinking Greatly Improves with Large Language Models. *Thinking Skills and Creativity*, 101356. https://doi.org/10.1016/j.tsc.2023.101356