STEP

1 **GENERATING DATA**

With situational judgement *items* or semi-structured interview questions prompt participants to generate ecologically valid descriptions of construct dimensions.

Output: A list of **sentences** or short texts reflecting different dimensions of the construct of interest.

2 **EMBEDDING SENTENCES**

Using NLP models calculate vector representations of **sentences** such that semantic relationships are preserved.

Output: A **high dimensional space** with number of points = length of sentence list.

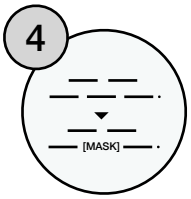
3 **MEASURING ITEM BIAS**

In the **high dimensional space** inspect the bias that *items* in the data generation process introduce to the embedding.

Output: **Estimate of bias** for each item in the embedding space.

* only if bias significant

4 **REDUCING ITEM BIAS**



Based on the **estimate of bias** apply text alterations to reduce bias in the embedding space.

Output: A **bias-corrected embedding space** comparable to 3).

5 **REDUCING DIMENSIONALITY**

Reduce dimensionality of the **bias-corrected embedding space** in order to prepare for clustering.

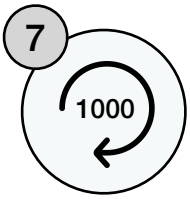
Output: A **low dimensional space** with equal number of points.

6 **CLUSTERING**

Perform clustering on the reduced and **low dimensional space**.

Output: One possible **clustering solution** reflecting dimensions of the construct of interest.

7 **CHECKING ROBUSTNESS**



Repeat steps 3) and 6) with different seeds to ensure that the **clustering solution** does not depend on stochastic components of the process.

Output: An estimate about the most probable number of **clusters**.

8 **VALIDATING CLUSTERS**

Compute final solution and measure consistency of **clusters** with sentence intrusion tasks and humans in the loop.

Output: Indicators for **cluster coherence**.

9 **INTERPRETING & COMPARING**

Interpret meaning of **coherent clusters** through expert judgement. Compare to alternative solutions from previous research.

Final result: Interpretable dimensions of the construct of interest.