Industrial and Organizational Psychology (2026), **19**, 203–207
doi:10.1017/iop.2026.10080

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

The information needed to refine constructs is already there—we just need to use it

Andrew Samo[1,2] and Scott Highhouse[1]

[1]Department of Psychology, Bowling Green State University, USA and [2]RHR International, USA
Corresponding Author: Andrew Samo; Email: asamo@rhrinternational.com

Construct proliferation has long been viewed as a problem in psychology. In 1937, Gordon Allport observed an abundance of motivational constructs, calling it "the problem of motivation" (Allport, 1937, p. 110). In 1958, Robert Guion hosted a symposium on the many meanings of industrial "morale" (Guion, 1958). Today, analyses of the APA's PsycTest database have found over 38,000 constructs with 43,000+ unique measures—with over 80% only used once or twice (Anvari et al., 2025; Elson et al., 2023). Although new constructs often signal the health of a scientific field (Iliescu et al., 2024), they can also fragment research landscapes, prevent standard comparisons, and otherwise make it harder to do good science (Flake et al., 2022; Meehl, 1992; Yarkoni, 2022).

Further complicating matters, Bowling and colleagues (2026) recommend that researchers identifying measures of potentially redundant constructs collect data and then analyze empirical covariation (see also Shaffer et al., 2016). This practice is resource intensive and becomes increasingly prohibitive in terms of cost and data quality as the number of constructs and items increases—it doesn't scale! As the number of constructs and measures increases, it becomes increasingly challenging to explore construct proliferation (Larsen et al., 2008; Wulff & Mata, 2025). In this context, calling for a construct moratorium sounds reasonable.

We believe, however, that advances in natural language processing, particularly with transformer architectures, offer tools uniquely designed to structure large amounts of previously unstructured content (Vaswani et al., 2017). Although the traditional approaches require new data collection for every construct comparison, natural language approaches can evaluate the conceptual similarity of hundreds or thousands of constructs using only the content that already exists: the definitions, scale labels, and item wordings that compose our measures. From a content sampling perspective, each new measure represents an attempt to sample from the universe of possible psychological content (Highhouse, 2009; Loevinger, 1957). Instead of stopping construct development, therefore, we should be using natural language to refine the construct space. In other words, the information needed to refine constructs is already there; we just need to use it.

Sampling psychological content with natural language

Although it is unclear whether psychological experiences are internally represented as natural language (Mandelbaum et al., 2022), our internal experiences are ultimately understood, shared, and then studied as such (Boyd & Schwartz, 2021). Psychometrics uses natural language at every

© The Author(s), 2026. Published by Cambridge University Press on behalf of Society for Industrial and Organizational Psychology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/iop.2026.10080 Published online by Cambridge University Press

stage: conceptualizing new phenomena, delineating content areas, and sampling the universe of content with representative items (Bringmann et al., 2022; Highhouse, 2009; Podsakoff et al., 2016). The language of our concepts and measures therefore occupies a unique position, sitting between the ontological (i.e., constructs as they are) and epistemological (i.e., constructs as empirical models) levels of validity (Borsboom et al., 2004; Messick, 1981)—what we might call the phenomenological level (i.e., constructs as we experience them). Items, at this level, can function as natural language indicators of the conceptual areas they were designed to sample (Arnulf et al., 2024; Bagozzi & Edwards, 1998).

The idea that language maps psychological content is not new (e.g., Allport, 1937; Galton, 1884; Pennebaker & King, 1999), and modern applications have broadly accepted this intuition (e.g., Yarkoni, 2010).[1] Beginning with Larsen and colleagues (2008), researchers have used semantic similarity, with latent semantic analysis (LSA), to predict scale validity, showing that linguistic overlap between survey items predicts observed associations (Arnulf et al., 2014, 2024). This has been directly applied to construct proliferation: detecting jingle and jangle fallacies (Larsen & Bong, 2016); measuring overlap between engagement and satisfaction (Nimon et al., 2016); and building large-scale tools like the Semantic Scale Network (Rosenbusch et al., 2020).

Around this time, the broader NLP field underwent a paradigm shift with transformer architectures (Vaswani et al., 2017), and psychological applications gradually followed, replacing the LSA previously used with transformer-based embeddings that better captured contextual semantics. At the current frontier, Hommel and colleagues (2025) updated the Semantic Scale Network approach with embeddings and introduced the Synthetic Nomological Net; Guenole and colleagues (2024, 2025) developed an embedding-based pseudofactor analysis to recover latent structure directly from item language; and Wulff and Mata (2025) used embeddings to systematically map taxonomic incommensurability across constructs.

Some practical recommendations for using natural language to refine construct space

In practice, there are a few basic ways to use natural language to explore construct space at multiple levels of analysis depending on your goals:

1. ***Situating a new construct.*** If you have an idea for a new construct but want to check how similar it is to existing constructs (or identify dis/similar constructs to include in a nomological network) before investing in validation, you can write a description or definition of the construct and input it into the Semantic Scale Network (Rosenbusch et al., 2020) or the Synthetic Nomological Net (Hommel et al., 2025). This allows you to estimate your novel construct's similarity to existing constructs and identify other measures to consider when refining your construct space or to include when establishing convergent and discriminant validity (Meehl, 1992).
2. ***Comparing between-construct space.*** If you are examining potential overlap, redundancy, or taxonomic incommensurability (i.e., jingle and jangle; Larsen & Bong, 2016) between constructs, then embedding-based semantic similarity offers a scalable alternative to empirical covariation. Following Wulff and Mata (2025), you can encode scale labels and item content, compute pairwise semantic similarity, and systematically identify jingle (i.e., high label similarity, low item similarity) or jangle fallacies (i.e., low label similarity, high item similarity) to investigate.

[1]A very brief Google Scholar search for "natural language" and "psychology" yields 17,800+ results between 2000–2025.

https://doi.org/10.1017/iop.2026.10080 Published online by Cambridge University Press

Table 1. Content and Semantic Overlap Framework for Within-Construct Overlap

		Content area overlap	
		Low	High
Semantic similarity	High	Jangle risk *(sampling different content areas with similar item language)*	Convergence *(sampling similar content areas with similar language)*
	Low	Distinct *(sampling unique areas with different language)*	Jingle risk *(sampling similar content areas with different language)*

3. ***Examining within-construct space.*** If you want to look within a construct space, when there is a particularly ambiguous construct or multiple measures claim to capture the same thing, a more granular analysis can identify specific areas of content convergence (or divergence) within a construct space. One approach we propose integrates content overlap (whether measures sample similar conceptual domains) with semantic similarity (whether item language is similar) yielding the 2 × 2 framework shown in Table 1. This can reveal which content areas within a construct space are a source of jingle or jangle and whether the issue lies in content sampling, natural language operationalizations, or both, providing a diagnostic.

Each of these approaches benefits from large, open-source repositories of psychological content. As starting points, the APA's PsycTests database (e.g., used by Hommel et al., 2025), the International Personality Item Pool (IPIP) content (e.g., used by Wulff & Mata, 2025), or domain-specific inventories like the judgment and decision-making (JDM) content from the Decision-Making Individual Differences Inventory (DMIDI; Appelt et al., 2011) provide rich pools of psychological content for exploration. To illustrate the accessibility of these methods, as shown in Figure 1, we collected the 200+ JDM measures cataloged in the DMIDI, encoded their available items as embeddings using open-source sentence transformers, and projected them to two-dimensional embedding space with UMAP reduction. The resulting visual shows clusters of linguistically similar measures, organized by JDM categories, and potential areas of redundancy within JDM measurement. We would expect that category (color) groups cluster more closely together, as they should share more similar content. Any points that are overlapping may share redundant content. Similarly, any unexpected colors in a larger cluster may need closer inspection. As we can see, on the right side of the figure, risk attitude measures do seem to cluster together, yet there are several decision-making styles and one personality measure that are closely grouped with risk attitudes. On the left side, we can see a mix of personality and risk attitude categories, with a small cluster of motivation measures toward the top. This is only a brief illustration, but there are a number of unexpected groupings to explore in more detail (see Figure 1).

Bowling et al. call for a moratorium until steps are taken to address construct proliferation. We outline several steps based in natural language analysis: situating new constructs within semantic nomological networks, comparing existing constructs with embedding-based semantic similarity (Figure 1), and examining within-construct heterogeneity through an integrated content-semantic overlap framework (Table 1). We also propose a systematic effort to take advantage of the natural language of psychological measurement: a large-scale, open-source inventory of psychological content, similar to the APA's PsycTests, IPIP, and DMIDI. So instead of reading lips to say "no new constructs," I-O psychology should, in fact, continue reading people's lips, using natural language to refine our understanding of psychological constructs.

https://doi.org/10.1017/iop.2026.10080 Published online by Cambridge University Press

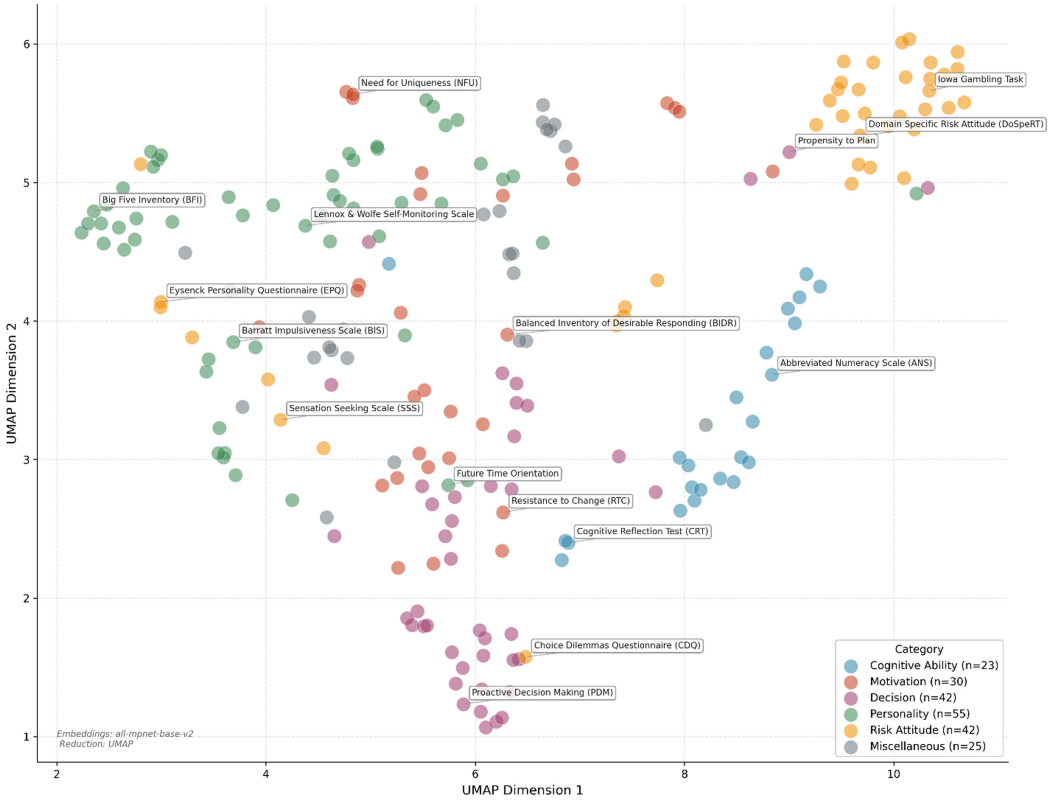


Figure 1. Semantic Embeddings of Decision-Making Individual Difference Inventory (DMIDI) Measure Categories
Note. Embeddings were encoded with an open-source all-mpnet-base-v2 sentence transformer and projected with a UMAP reduction. Measures were chosen to label by identifying the most central and extreme category members based on UMAP dimension centroid distances.

References

Anvari, F., Alsalti, T., Oehler, L. A., Marion, Z., Hussey, I., Elson, M., & Arslan, R. C. (2025). A fragmented field: Construct and measure proliferation in psychology. *Advances in Methods and Practices in Psychological Science*, **8**(3), Article 25152459251360642. https://doi.org/10.1177/25152459251360642

Allport, G. W. (1937). *Personality: A psychological interpretation*. Holt.

Appelt, K. C., Milch, K. F., Handgraaf, M. J. J., & Weber, E. U. (2011). The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, **6**(3), 252–262.

Arnulf, J. K., Larsen, K. R., Martinsen, Ø. L., & Bong, C. H. (2014). Predicting survey responses: How and why semantics shape survey statistics on organizational behaviour. *PloS One*, **9**(9), e106361.

Arnulf, J. K., Olsson, U. H., & Nimon, K. (2024). Measuring the menu, not the food: "Psychometric" data may instead measure "lingometrics" (and miss its greatest potential). *Frontiers in Psychology*, **15**, 1308098.

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, **1**(1), 45–87. https://doi.org/10.1177/109442819800100104

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, **111**(4), 1061.

Bowling, N. A., Sessa, V. I., Shaffer, J. A., & Banks, G. C. (2026). Read my lips: No new constructs! Construct proliferation as a threat to the future of I-O psychology. *Industrial-Organizational Psychology*, **19**(2), 186–202.

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, **40**(1), 21–41. https://doi.org/10.1177/0261927X20967028

Bringmann, L. F., Elmer, T., & Eronen, M. I. (2022). Back to basics: The importance of conceptual clarification in psychological science. *Current Directions in Psychological Science*, **31**(4), 340–346. https://doi.org/10.1177/09637214221096485

https://doi.org/10.1017/iop.2026.10080 Published online by Cambridge University Press

Elson, M., Hussey, I., Alsalti, T., & Arslan, R. C. (2023). Psychological measures aren't toothbrushes. *Nature Communications Psychology*, **1**(25), 1–4. https://doi.org/10.1038/s44271-023-00026-9

Flake, J. K., Davidson, I. J., Wong, O., & Pek, J. (2022). Construct validity and the validity of replication studies: A systematic review. *American Psychologist*, **77**(4), 576.

Galton, F. (1884). Measurement of character. *Fortnightly Review*, **36**, 179–185.

Guenole, N., Samo, A., & Sun, T. (2024). Pseudo factor analysis of language embedding similarity matrices: New ways to model latent constructs. https://doi.org/10.31234/osf.io/vf3se

Guenole, N., D'Urso, E. D., Samo, A., Sun, T., & Haslbeck, J. (2025). Enhancing scale development: Pseudo factor analysis of language embedding similarity matrices. https://doi.org/10.31234/osf.io/vf3se_v2

Guion, R. M. (1958). Industrial morale (a symposium): I. The problem of terminology. *Personnel Psychology*, **11**, 59–94.

Highhouse, S. (2009). Tests don't measure jobs: The meaning of content validation. *Industrial & Organizational Psychology*, **2**(4), 493–495.

Hommel, B., Külpmann, A., & Arslan, R. (2025). The synthetic nomological net: A search engine to identify conceptual overlap in measures in the behavioral sciences. https://doi.org/10.31234/osf.io/nfgmv_v1

Iliescu, D., Greiff, S., Ziegler, M., Nye, C., Geisinger, K., Sellbom, M., Samuel, D., & Saklofske, D. (2024). Proliferation of measures contributes to advancing psychological science. *Communications Psychology*, **2**(1), 19.

Larsen, K. R., Nevo, D., & Rich, E. (2008). Exploring the semantic validity of questionnaire scales. Proceedings of the 41st annual Hawaii International Conference on System Sciences (HICSS 2008). 440–440.

Larsen, K. R., & Bong, C. H. (2016). A tool for addressing construct identity in literature reviews and meta-analyses. *MIS Quarterly*, **40**(3), 529–551.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, **3**(3), 635–694.

Mandelbaum, E., Dunham, Y., Feiman, R., Firestone, C., Green, E. J., Harris, D., Kibbe, M. M., Kurdi, B., Mylopoulos, M., Shepherd, J., Wellwood, A., Porot, N., & Quilty-Dunn, J. (2022). Problems and mysteries of the many languages of thought. *Cognitive Science*, **46**(12), e13225. https://doi.org/10.1111/cogs.13225

Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*, **60**(1), 117–174. https://doi.org/10.1111/j.1467-6494.1992.tb00269.x

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, **89**(3), 575–588. https://doi.org/10.1037/0033-2909.89.3.575

Nimon, K., Shuck, B., & Zigarmi, D. (2016). Construct overlap between employee engagement and job satisfaction: A function of semantic equivalence? *Journal of Happiness Studies*, **17**(3), 1149–1171.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality & Social Psychology*, **77**(6), 1296–1312.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. *Organizational Research Methods*, **19**(2), 159–203.

Rosenbusch, H., Wanders, F., & Pit, I. L. (2020). The semantic scale network: An online tool to detect semantic overlap of psychological scales and prevent construct proliferation. *Psychological Methods*, **25**(3), 380–392.

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods*, **19**(1), 80–110.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, **30**, 5998–6008.

Wulff, D. U., & Mata, R. (2025). Semantic embeddings reveal and address taxonomic incommensurability in psychological measurement. *Nature Human Behaviour*, **9**, 944–954.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, **44**(3), 363–373.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, **45**, e1. https://doi.org/10.1017/S0140525X20001685

Cite this article: Samo, A. & Highhouse, S. (2026). The information needed to refine constructs is already there—we just need to use it. *Industrial and Organizational Psychology* **19**, 203–207. https://doi.org/10.1017/iop.2026.10080

https://doi.org/10.1017/iop.2026.10080 Published online by Cambridge University Press