

Dartmouth College
Department of Psychological and Brain Sciences
HB 6207 Moore Hall
Hanover, NH, 03755

February 8, 2023

To the editors of *Nature Human Behavior*:

We have enclosed our manuscript entitled *Text embedding models yield high-resolution insights into conceptual knowledge from short multiple-choice quizzes* to be considered for publication as an *Article*. In our manuscript, we develop and test a mathematical framework, based on natural language processing models, for tracking and characterizing the acquisition of real-world conceptual knowledge.

We asked our participants to watch two lecture videos from the Khan Academy platform and to take short multiple-choice quizzes before and after each lecture. We used a natural language processing model to capture the conceptual content of each moment of the lectures, and of each question. This model represents each lecture as a “trajectory” through a high-dimensional embedding space, and each question as a point in the same space. We designed the embedding space to map conceptually related content onto nearby coordinates. We used these embeddings, along with participants’ automatically graded responses to the quiz questions, to characterize nuances in participants’ knowledge. Essentially, our approach models “knowledge” about a given concept as the weighted proportion of questions a participant answers correctly, where the weights reflect how much each question is “about” the given concept (according to the natural language processing model). Whereas the raw proportion of correctly answered questions can characterize content-nonspecific knowledge (e.g., something akin to “how *much* the participant knows”), these weighted proportions provide estimates about *what* the participant knows.

Our manuscript makes several important contributions. First, we define our framework and show how we can jointly model the content of multiple real-world course lectures and questions. We designed our framework to be scalable to a very broad range of content areas. Second, we use participants’ behavioral data to evaluate the knowledge estimates produced by our framework. We show that the knowledge estimates can reliably differentiate whether participants will answer individual held-out quiz questions correctly versus incorrectly. Third, we use our framework to construct detailed “maps” of what participants know (based on their responses from a single quiz) and how their knowledge changes over time (based on differences between the knowledge maps estimated from successive quizzes). These maps demonstrate how knowledge (and changes in knowledge) about *any* conceptual content expressible by the text embedding model may be estimated from the same small set of multiple-choice questions.

In addition to showing how detailed insights into what people know may be gleaned from a small

number of questions, our work also lays the foundation for several exciting future applications and research areas. For example, the knowledge maps produced by our framework could be used by educators to help guide their lessons and better understand their students. The maps could also be used as a core component of a new generation of automated tutors that adapt to-be-presented content according to the latest estimate of what a student knows. More broadly, we also see potential overlap with other research domains that consider how information “flows” between people. In our manuscript we focus on information flow from teachers (i.e., the course lectures) to students. But we see analogous approaches being used to characterize other social domains as well, such as between friends having a conversation, doctor-patient interactions, romantic partnerships, business meetings, and more. We suggest that the degree of overlap between two individuals’ knowledge maps, in a particular region of content space, may serve as a predictive signal for how effectively those individuals will be able to communicate about content in that region of content space. We expect that this article will be of interest to a broad audience including educators, cognitive psychologists, social psychologists, natural language processing researchers, and others.

Thank you for considering this manuscript, and we hope you will find it suitable for publication in *Nature Human Behavior*.

Sincerely,

Jeremy R. Manning