TECHNION
Israel Institute
of Technology

FACULTY OF DATA
AND DECISION SCIENCES
MAKING SENSE OF IT

CREATIVE THINKING

Yoed N. Kenett
yoedk@Technion.ac.il
Faculty of Data and Decision Sciences
Technion – Israel Institute of Technology

February 9th, 2025

Measuring Creativity

- How can creativity be quantified?
- Human creativity judgements are subjective, and can be unreliable
- Can machines objectively assess human creativity?

Key take-home message

Focus on open-ended tasks that facilitate divergent thinking

Link to tools

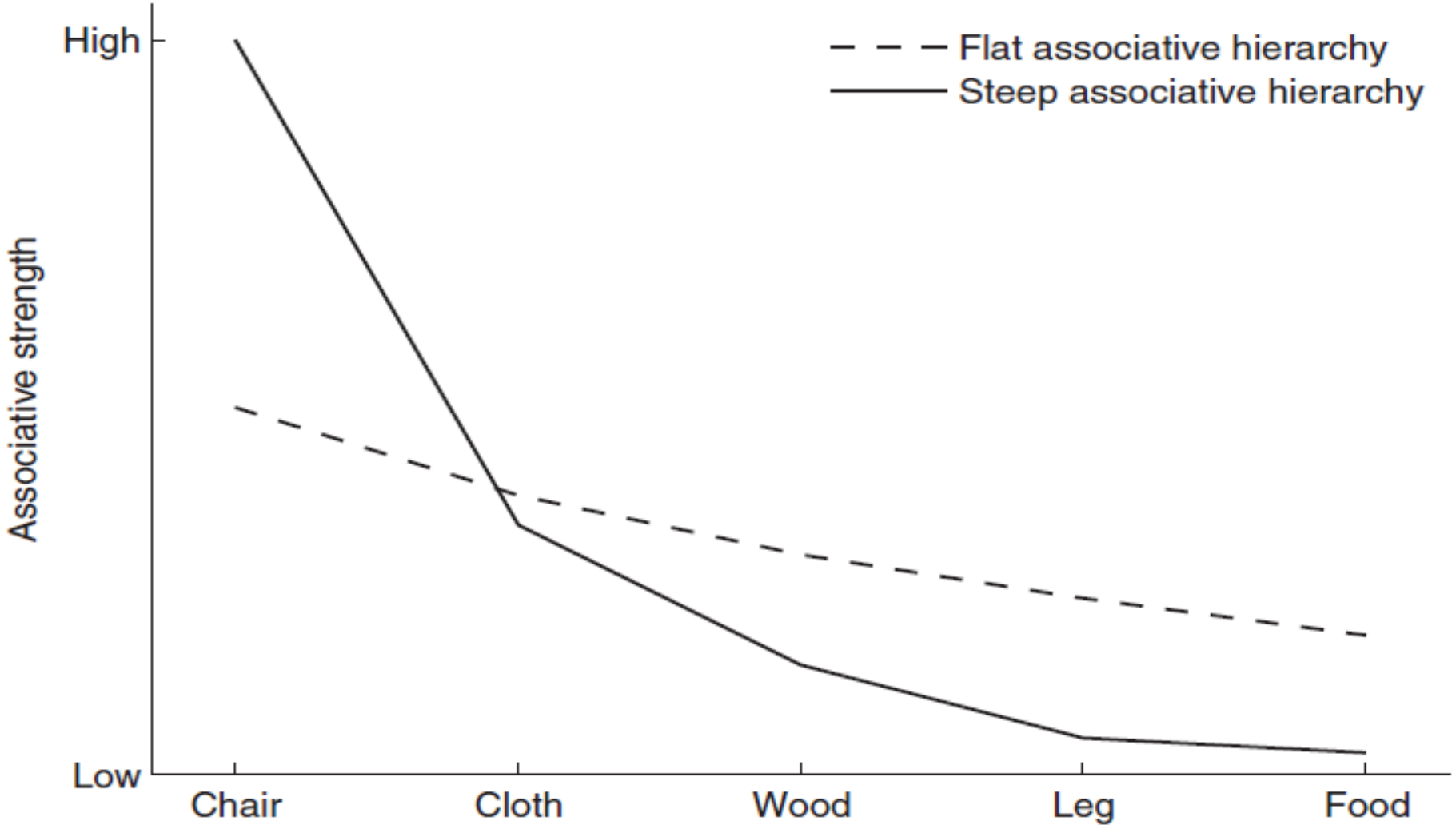
- Forward Flow
  - https://osf.io/7p5mt/
- DSI
  - https://osf.io/ath2s/
- OCSAI
  - https://openscoring.du.edu/scoringllm
- AUDRA
  - https://osf.io/kqn9v/
- Bloom complexity model
  - https://osf.io/823ak/
- Creative problem solving
  - https://osf.io/45veq/?view_only=9fcb23a2564b4190a4ceb1f2f1707129
- CAP
  - https://cap.ist.psu.edu/

A computational assessment revolution

- The application of computational lingutistic models **revolutionized** how we assess creativity

- Issues to be aware of
  - Language issues
  - Model verification
  - Task issues
  - Tools issues

The associative theory of creativity (Mednick, 1962)



How do we search our memory?

nature reviews psychology

https://doi.org/10.1038/s44159-023-00158-z

Perspective

Check for updates

The role of memory in creative ideation

Mathias Benedek[1], Roger E. Beaty[2], Daniel L. Schacter[3,4] & Yoed N. Kenett[5]

Abstract

Creativity reflects the remarkable human capacity to produce novel and effective ideas. Empirical work suggests that creative ideas do not just emerge out of nowhere but typically result from goal-directed memory processes. Specifically, creative ideation is supported by controlled retrieval, involves semantic and episodic memory, builds on processes used in memory construction and differentially recruits memory at different stages in the creative process. In this Perspective, we propose a memory in creative ideation (MemiC) framework that describes how creative ideas arise across four distinguishable stages of memory search, candidate idea construction, novelty evaluation and effectiveness evaluation. We discuss evidence supporting the contribution of semantic and episodic memory to each stage of creative ideation. The MemiC framework overcomes the shortcomings of previous creativity theories by accounting for the controlled, dynamic involvement of different memory systems across separable ideation stages and offers a clear agenda for future creativity research.

Sections

[1]Department of Psychology, University of Graz, Graz, Austria. [2]Department of Psychology, Pennsylvania State University, University Park, PA, USA. [3]Department of Psychology, Harvard University, Cambridge, MA, USA. [4]Center for Brain Science, Harvard University, Cambridge, MA, USA. [5]Faculty of Data and Decision Sciences, Technion — Israel Institute of Technology, Haifa, Israel. e-mail: mathias.benedek@uni-graz.at

Trends in Cognitive Sciences

CellPress

Review

Associative thinking at the core of creativity

Roger E. Beaty[1,*] and Yoed N. Kenett[2]

Creativity has long been thought to involve associative processes in memory: connecting concepts to form ideas, inventions, and artworks. However, associative thinking has been difficult to study due to limitations in modeling memory structure and retrieval processes. Recent advances in computational models of semantic memory allow researchers to examine how people navigate a semantic space of concepts when forming associations, revealing key search strategies associated with creativity. Here, we synthesize cognitive, computational, and neuroscience research on creativity and associative thinking. This Review highlights distinctions between free- and goal-directed association, illustrates the role of associative thinking in the arts, and links associative thinking to brain systems supporting both semantic and episodic memory – offering a new perspective on a longstanding creativity theory.

Highlights

Creativity involves associative thinking: linking concepts from memory. Computational models of semantic memory allow researchers to quantify associative thinking as movement through a semantic space of concepts.

Associative thinking reflects a search process operating on a semantic memory network structure. Highly creative people travel further in semantic space, switch between more semantic subcategories, and make larger leaps between associations.

Creativity involves both free associations, in which concepts are spontaneously connected, and goal-directed associations, in which concepts are strategically combined.

Free association is relevant for artistic creativity, implicating domain-general cognitive abilities in domain-specific creative expertise.

Neuroscience evidence indicates associative thinking engages brain regions involved in both semantic and episodic memory.

Revisiting the role of associative thinking in creativity

What is the first word that comes to mind when you think of creativity? Some people may associate creativity with art; others with imagination, novelty, or expression. This exercise illustrates the phenomenon of association – how one concept links to others in memory, and how people vary in the associations they make [1]. Creative thinking has long been conceptualized as involving an associative process over memory, where concepts are combined to form new and effective ideas [2–6]. However, associative thinking has been historically challenging to study, due in part to methodological limitations in modeling memory and the retrieval processes operating on it. Prior studies have relied on simple measures of association, such as counting the total number of associations produced (i.e., the product of associative thinking) – obscuring the process of associative thinking and limiting our understanding of its role in creativity.

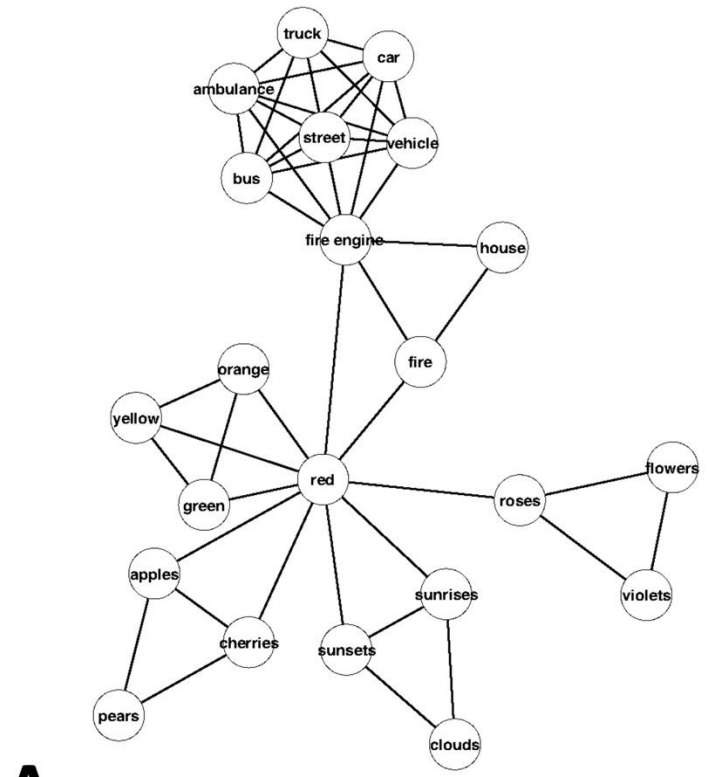
Recent advances in the computational modeling of semantic memory – the vast database of concepts, and the relationships between them [7] – have begun to overcome these limitations, yielding new insights into associative thinking and its contribution to creativity. In particular, **distributional semantic models** (see Glossary) provide powerful tools for quantifying **semantic distance**, allowing researchers to quantitatively measure how far people travel in a semantic space (or network) of concepts when searching memory. Cognitive and neuroimaging studies of associative thinking have yielded additional insights, from disentangling spontaneous/**free association** versus controlled/**goal-directed association**, to demonstrating the role of associative thinking in the arts and sciences, to linking associative thinking to brain systems implicated in creative thought.

In this Review, we integrate cognitive, computational, and neuroscience research on creativity and associative thinking. We operationalize creativity as a cognitive process involving the generation of new and effective ideas [8], while recognizing that creativity is complex, multifaceted, and often domain specific. We focus on the role of associative thinking as a general mechanism driving the early stage of idea generation. We therefore view creative thinking as a form of high-level cognition – a product of lower-level cognitive systems, including cognitive control, attention, and memory [9,10].

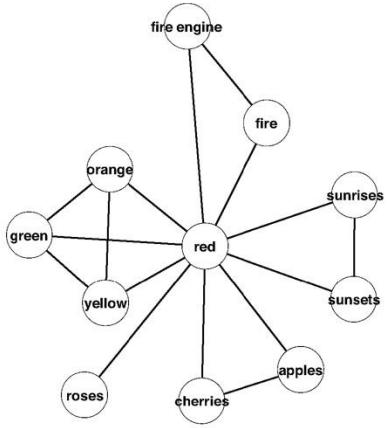
[1]Department of Psychology, Pennsylvania State University, University Park, PA, USA
[2]Faculty of Data and Decision Sciences, Technion – Israel Institute of Technology, Haifa, Israel

*Correspondence: rebeaty@psu.edu (R.E. Beaty).

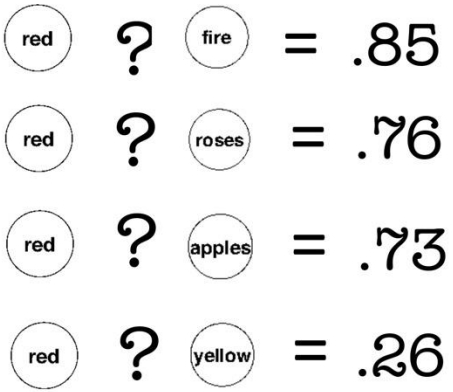
Corpus- & Network-based models



A



B

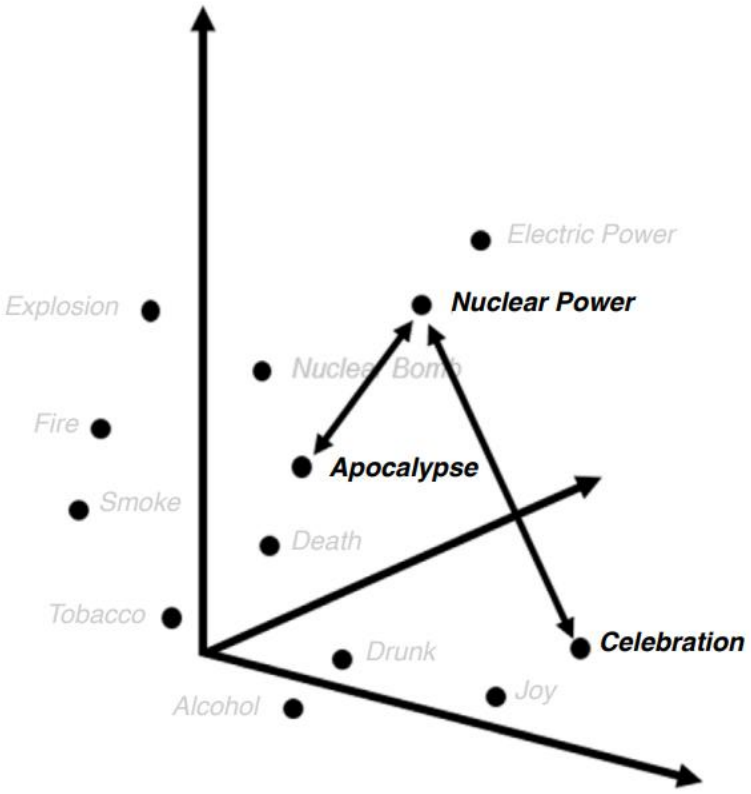


C

Kenett (2019). *Current Opinion in Behavioral Sciences*, 27, 11-16.

Semantic Distance

- Semantic distance quantifies conceptual (dis)similarly using distributional semantic models
- Computes co-occurrence of words (semantic similarity) in large text corpora (e.g., books)
- Consistent with associative theory of creativity (Mednick, 1962)



Bhatia et al. (2019), *Current Opinion in Behavioral Sciences;* Mednick (1962), *Psychological Review*

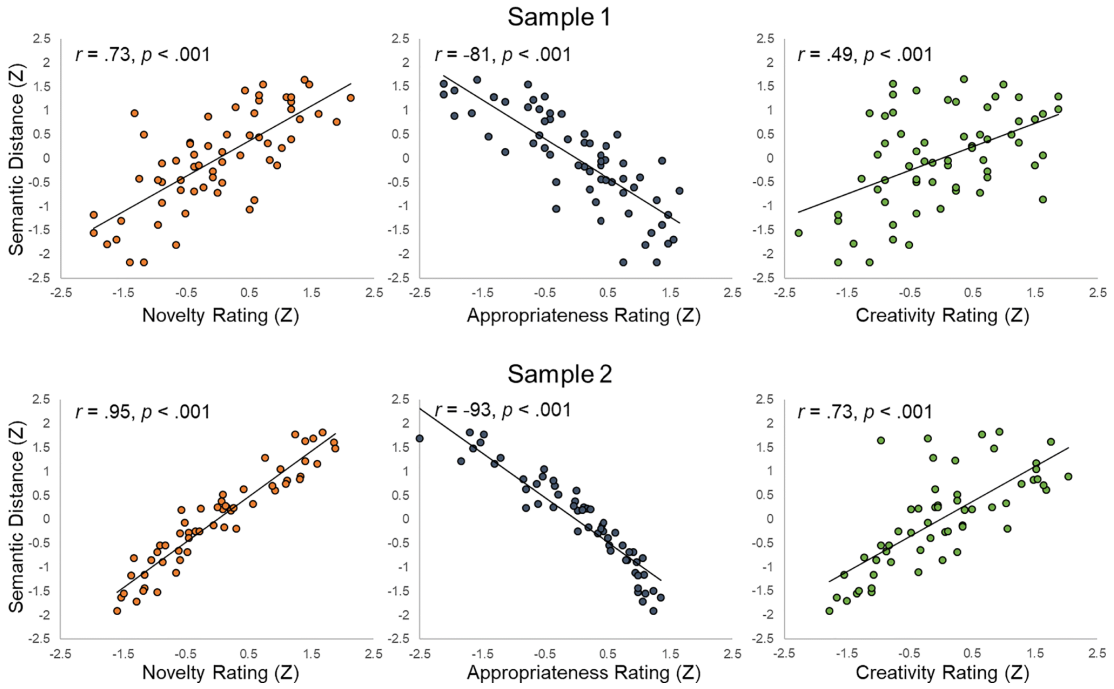
Semantic Distance Demo

- Let's see how semantic distance works

- Think of associations to the word: coffee

Validating Semantic Distance

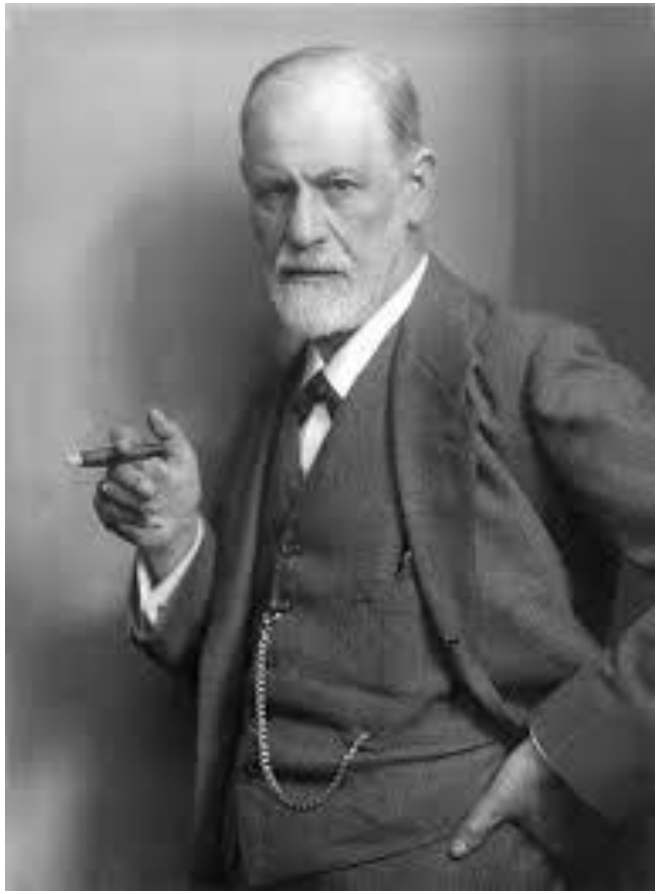
- Semantic distance correlates with human ratings of novelty, appropriateness, and creativity on simple word association tasks



Beaty & Johnson (2021), *Behavior Research Methods*

Forward flow

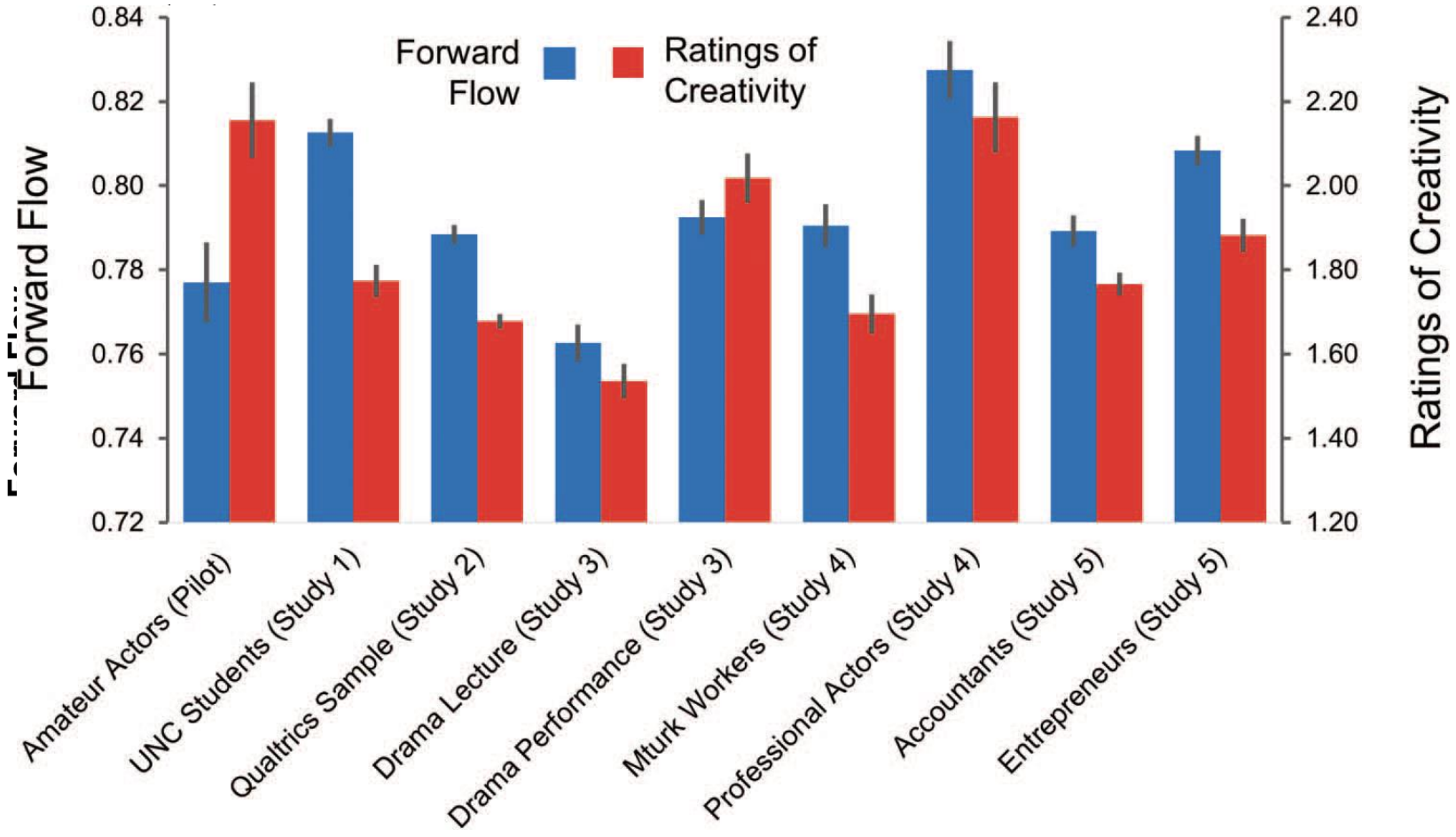
Chain free association task



Write down the next word that follows in your mind from the previous word.
Press TAB after each word. Continue when all text boxes are complete.

Your starting word is 'Table'

Forward flow



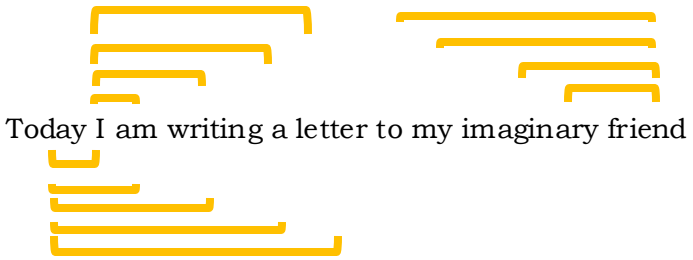
Gray et al., (2019). *American Psychologist, 74*(5), 539-554.

Divergent semantic integration

DSI 1/3

- Can semantic distance capture creativity in narrative texts (e.g., short stories) via divergent semantic integration (DSI)?

DSI computes the average of all pairwise **word-to-word** semantic distances in a text



Indexes how much story incorporates **divergent contexts, situations,** and **concepts**

3-word story prompt:
stamp-letter-send

High DSI Story

My husband had just left for the war, and I had mixed feelings about it. I have always imagined from the movies, writing a **letter** would be romantic. But I was wrong once I put the **stamp** on that envelope to **send**.

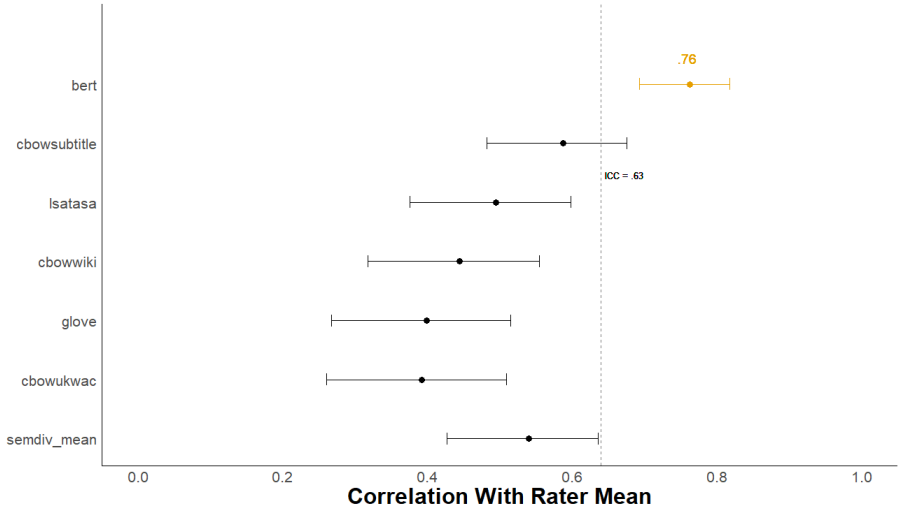
Low DSI Story

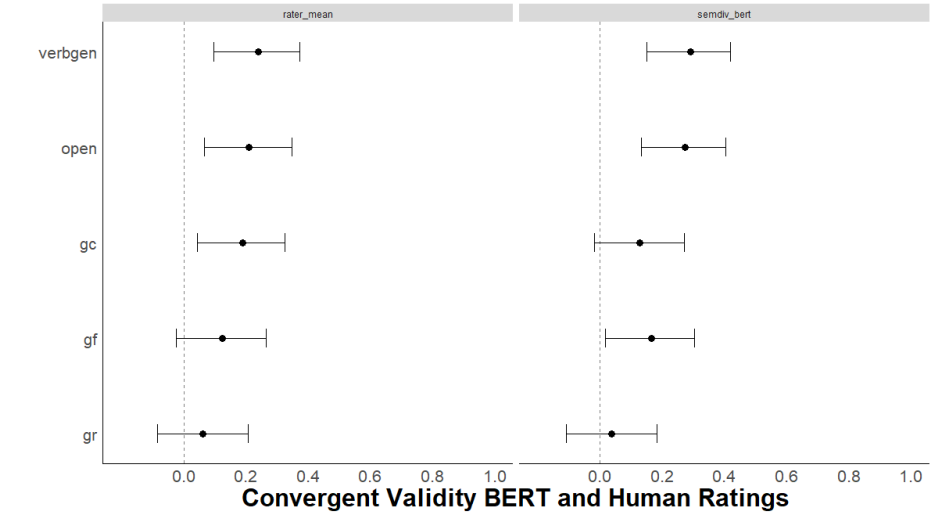
I wrote a **letter** to my aunt. I went to the post office and bought a **stamp**. I put the stamp on the letter and gave it to the mailman to **send**.

Johnson et al. (under review)

DSI 2/3

- N = 179 undergrads
- Writing prompt: stamp-letter-send
- 5 creativity raters
- 6 DSI models (e.g., LSA, BERT)
- Validation measures (openness, intelligence, word association)





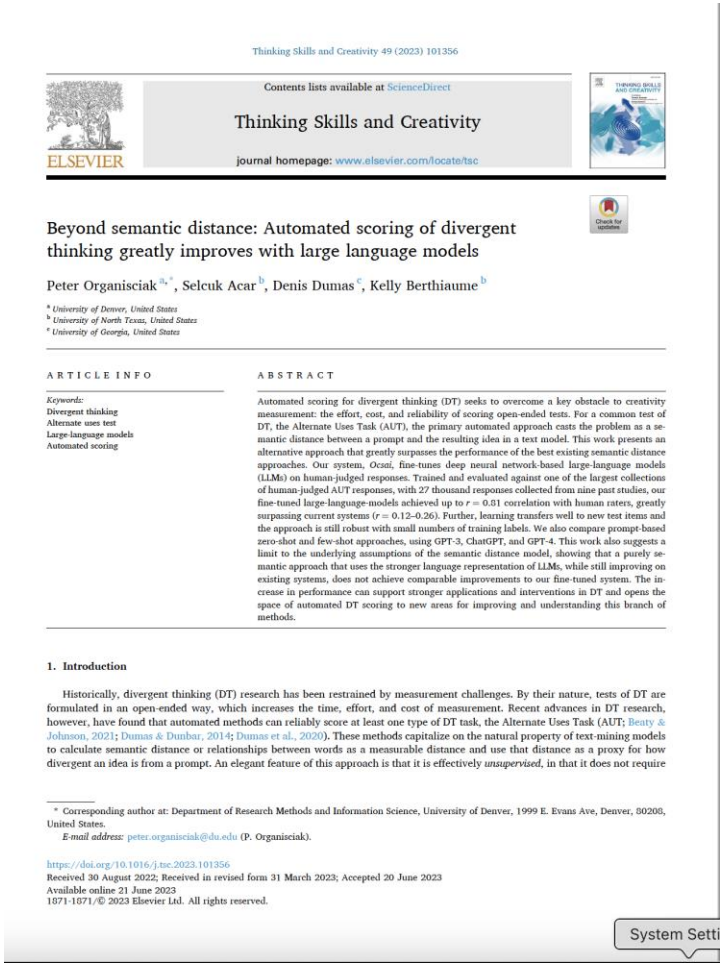
DSI 3/3

- DSI can reliably and automatically assess narrative creativity on (very) short stories

- Results replicated across 27 story prompts; >3,500 stories; and ethnically diverse samples

Large Language Models

Open Creativity Scoring with AI (OCSAI) 1/2



Thinking Skills and Creativity 49 (2023) 101356

Contents lists available at ScienceDirect

Thinking Skills and Creativity

journal homepage: www.elsevier.com/locate/tsc

Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models

Peter Organisciak [a,*], Selcuk Acar [b], Denis Dumas [c], Kelly Berthiaume [b]

[a] University of Denver, United States
[b] University of North Texas, United States
[c] University of Georgia, United States

ARTICLE INFO

Keywords:
Divergent thinking
Alternate uses test
Large-language models
Automated scoring

ABSTRACT

Automated scoring for divergent thinking (DT) seeks to overcome a key obstacle to creativity measurement: the effort, cost, and reliability of scoring open-ended tests. For a common test of DT, the Alternate Uses Task (AUT), the primary automated approach casts the problem as a semantic distance between a prompt and the resulting idea in a text model. This work presents an alternative approach that greatly surpasses the performance of the best existing semantic distance approaches. Our system, *Ocsai*, fine-tunes deep neural network-based large-language models (LLMs) on human-judged responses. Trained and evaluated against one of the largest collections of human-judged AUT responses, with 27 thousand responses collected from nine past studies, our fine-tuned large-language-models achieved up to $r = 0.81$ correlation with human raters, greatly surpassing current systems ($r = 0.12–0.26$). Further, learning transfers well to new test items and the approach is still robust with small numbers of training labels. We also compare prompt-based zero-shot and few-shot approaches, using GPT-3, ChatGPT, and GPT-4. This work also suggests a limit to the underlying assumptions of the semantic distance model, showing that a purely semantic approach that uses the stronger language representation of LLMs, while still improving on existing systems, does not achieve comparable improvements to our fine-tuned system. The increase in performance can support stronger applications and interventions in DT and opens the space of automated DT scoring to new areas for improving and understanding this branch of methods.

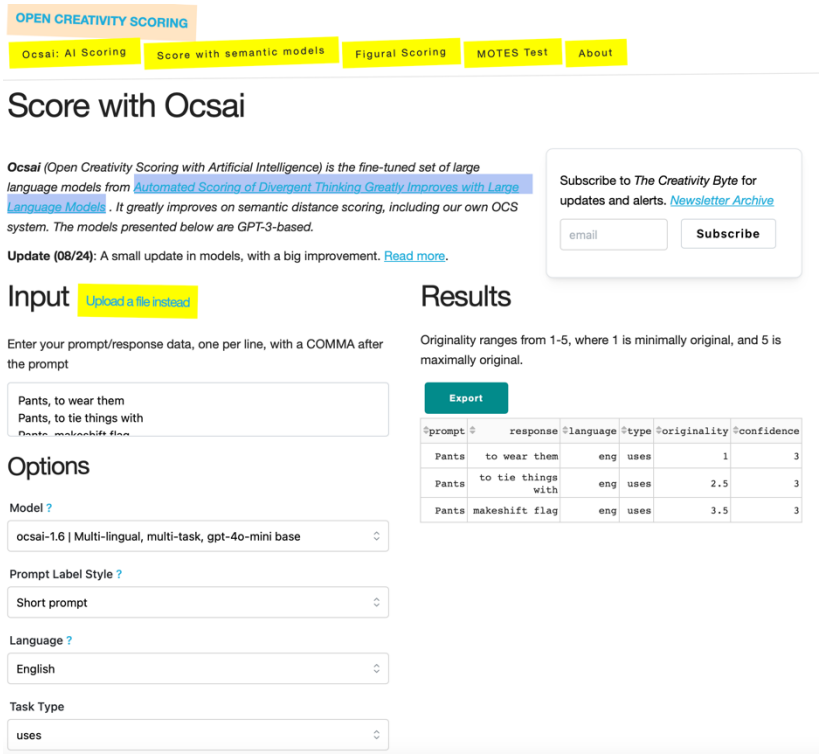
1. Introduction

Historically, divergent thinking (DT) research has been restrained by measurement challenges. By their nature, tests of DT are formulated in an open-ended way, which increases the time, effort, and cost of measurement. Recent advances in DT research, however, have found that automated methods can reliably score at least one type of DT task, the Alternate Uses Task (AUT; Beaty & Johnson, 2021; Dumas & Dunbar, 2014; Dumas et al., 2020). These methods capitalize on the natural property of text-mining models to calculate semantic distance or relationships between words as a measurable distance and use that distance as a proxy for how divergent an idea is from a prompt. An elegant feature of this approach is that it is effectively *unsupervised*, in that it does not require

* Corresponding author at: Department of Research Methods and Information Science, University of Denver, 1999 E. Evans Ave, Denver, 80208, United States.
E-mail address: peter.organisciak@du.edu (P. Organisciak).

https://doi.org/10.1016/j.tsc.2023.101356
Received 30 August 2022; Received in revised form 31 March 2023; Accepted 20 June 2023
Available online 21 June 2023
1871-1871/

System Setti



OPEN CREATIVITY SCORING

Ocsai: AI Scoring | Score with semantic models | Figural Scoring | MOTES Test | About

Score with Ocsai

Ocsai *(Open Creativity Scoring with Artificial Intelligence) is the fine-tuned set of large language models from Automated Scoring of Divergent Thinking Greatly Improves with Large Language Models . It greatly improves on semantic distance scoring, including our own OCS system. The models presented below are GPT-3-based.*

Update (08/24): A small update in models, with a big improvement. Read more.

Subscribe to *The Creativity Byte* for updates and alerts. Newsletter Archive

email | Subscribe

Input

Upload a file instead

Enter your prompt/response data, one per line, with a COMMA after the prompt

Pants, to wear them
Pants, to tie things with
Pants, makeshift flag

Options

Model ?

ocsai-1.6 | Multi-lingual, multi-task, gpt-4o-mini base

Prompt Label Style ?

Short prompt

Language ?

English

Task Type

uses

Results

Originality ranges from 1-5, where 1 is minimally original, and 5 is maximally original.

Export

prompt	response	language	type	originality	confidence
Pants	to wear them	eng	uses	1	3
Pants	to tie things with	eng	uses	2.5	3
Pants	makeshift flag	eng	uses	3.5	3

Open Creativity Scoring with AI (OCSAI) 2/2

Automated Drawing Assessment (AUDRA)

Behavior Research Methods (2024) 56:3619–3636
https://doi.org/10.3758/s13428-023-02258-3

ORIGINAL MANUSCRIPT

AuDrA: An automated drawing assessment platform for evaluating creativity

John D. Patterson[1] · Baptiste Barbot[2,3] · James Lloyd-Cox[4] · Roger E. Beaty[1]

Accepted: 25 September 2023 / Published online: 2 November 2023

Abstract
The visual modality is central to both reception and expression of human creativity. Creativity assessment paradigms, such as structured drawing tasks Barbot (2018), seek to characterize this key modality of creative ideation. However, visual creativity assessment paradigms often rely on cohorts of expert or naïve raters to gauge the level of creativity of the outputs. This comes at the cost of substantial human investment in both time and labor. To address these issues, recent work has leveraged the power of machine learning techniques to automatically extract creativity scores in the verbal domain (e.g., SemDis; Beaty & Johnson 2021). Yet, a comparably well-vetted solution for the assessment of visual creativity is missing. Here, we introduce AuDrA – an Automated Drawing Assessment platform to extract visual creativity scores from simple drawing productions. Using a collection of line drawings and human creativity ratings, we trained AuDrA and tested its generalizability to untrained drawing sets, raters, and tasks. Across four datasets, nearly 60 raters, and over 13,000 drawings, we found AuDrA scores to be highly correlated with human creativity ratings for new drawings on the same drawing task (r = .65 to .81; mean = .76). Importantly, correlations between AuDrA scores and human raters surpassed those between drawings' elaboration (i.e., ink on the page) and human creativity raters, suggesting that AuDrA is sensitive to features of drawings beyond simple degree of complexity. We discuss future directions, limitations, and link the trained AuDrA model and a tutorial (https://osf.io/kqn9v/) to enable researchers to efficiently assess new drawings.

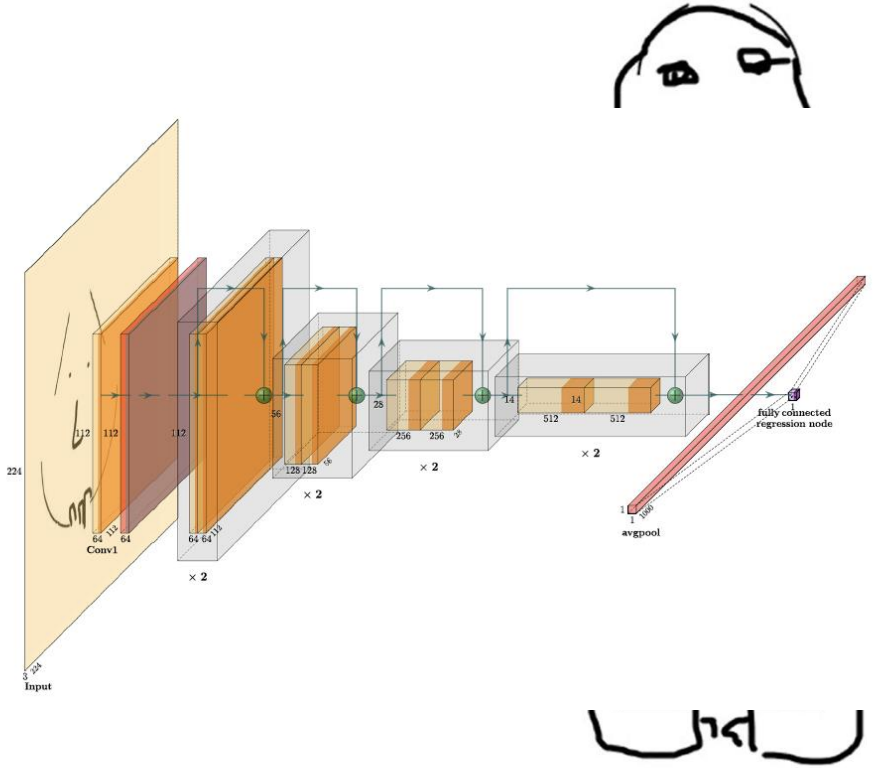
Keywords Automated creativity scoring · Computational creativity · Divergent thinking · Drawing assessment · Visual creativity

Introduction

How can human creativity be quantified? Researchers commonly administer tests of creative thinking – spanning verbal tasks (e.g., word association) to visual tasks (e.g., sketches) – yet they are confronted with the vexing question of how to quantify creative outputs from such tests. A common approach is to ask human raters to provide subjective judgements for each response, in the spirit of the classic Consensual Assessment Technique (CAT; Amabile, 1982). Although subjective scoring can be reliable and valid (Amabile, 1982; Kaufman et al., 2007; Myszkowski & Storme, 2019), it is also time-consuming and resource-intensive, slowing the pace of research, and acting as a barrier for researchers and practitioners without the human resources to support subjective scoring methods such as the CAT. Recently, researchers have begun to rigorously test whether verbal creativity assessment can be automated using machine learning, with encouraging signs of progress, including strong correlations between computational metrics and human ratings (Acar et al., 2021; Beaty & Johnson, 2021; Buczak et al., 2023; Dumas et al., 2021; Stevenson et al., 2020). This work builds on a seminal study of automatic assessment of verbal creativity tests (Forthmann & Doebler, 2022; Paulus et al., 1970).

✉ John D. Patterson
jpttrsn@psu.edu

[1] Department of Psychology, Pennsylvania State University, University Park, PA, USA
[2] Psychological and Educational Sciences Research Institute, UCLouvain, Louvain-la-Neuve, Belgium
[3] Child Study Center, Yale University, New Haven, USA
[4] Department of Psychology, Goldsmiths, University of London, London, UK



Automatic scoring of question complexity 1/2

Automated Scoring of Open-Ended Question Complexity: A Large Language Model Approach

Tuval Raz
tuval.raz@campus.technion.ac.il
Technion—Israel Institute of Technology
Simone Luchini
Pennsylvania State University
Roger Beaty
Pennsylvania State University
Yoed Kenett
Technion—Israel Institute of Technology

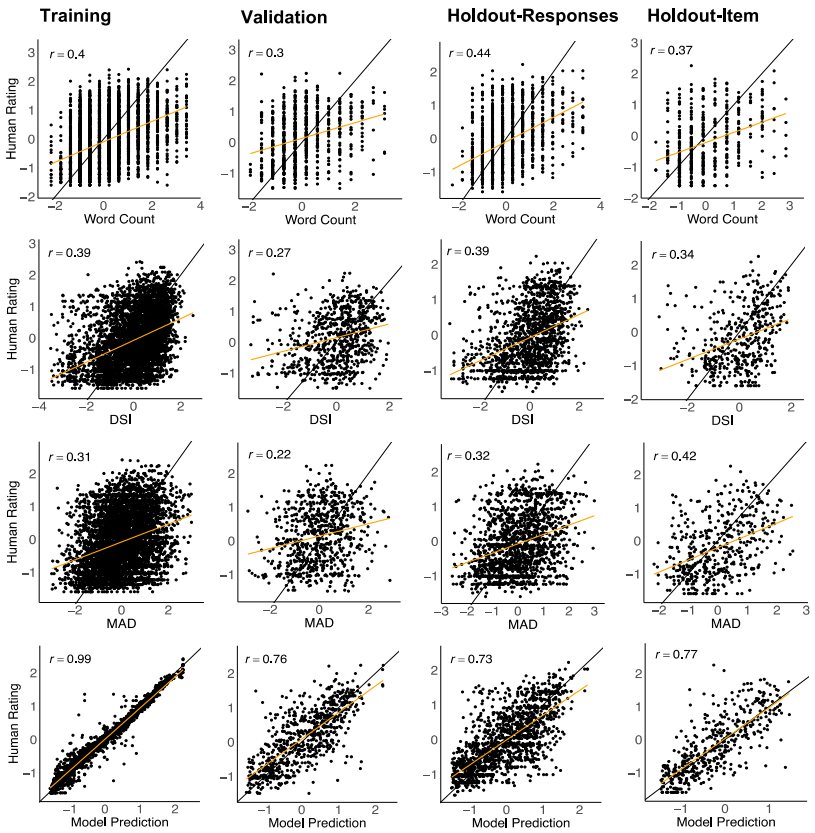
Research Article

Keywords: Question asking, Bloom taxonomy, NLP, LLM

Posted Date: February 19th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3890828/v1

Additional Declarations: No competing interests reported.



Automatic scoring of question complexity 2/2

Complex problem solving

Automatic Scoring of Creative Problem-Solving with Large Language Models: A Comparison of Originality and Quality Ratings

Simone Luchini[1], Nadine T. Maliakkal[2], Paul V. DiStefano[1], John D. Patterson[1], Roger E. Beaty[1], & Roni Reiter-Palmon[2]

[1] Pennsylvania State University, USA

[2] University of Nebraska Omaha, USA

Author Note

R.E.B. is supported by grants from the National Science Foundation [DRL-1920653; DUE-2155070].

Correspondence should be addressed to Simone Luchini or Roger E. Beaty, 140 Moore Building, University Park, PA 16802. Email: skl5875@psu.edu.

Becky is a college student who works part-time at Mark's Pizzeria. Mark, the owner of the restaurant, has treated Becky very well. He gave her a job that she needs to help pay her rent when no other business would employ her because she was arrested for shoplifting three years ago. Mark also lets Becky work around her school schedule and has asked if she wants to be a shift manager in the summers. Becky's roommate Jim also works at the pizzeria, but Jim has been causing a lot of problems at work. He always avoids doing his job, treats customers rudely, and makes a lot of mistakes with orders. Jim recently began stealing food from the pizzeria. Two days ago, the pizzeria was short-staffed, so Jim and Becky were the only employees left at closing time. Jim made 10 extra pizzas and took them home to a party he was hosting without paying for them. Becky feels like she needs to do something about Jim's behavior. However, Becky is hesitant to tell Mark about Jim because Jim is a good friend to Becky. Becky also needs Jim to have a job so he can pay his portion of their rent. Becky does not know what to do.

Creatvity assessment platform

CREATIVITY
ASSESSMENT
PLATFORM