

Article

School Reform: New Future-Ready Quality Outcomes and Proposed Measures

David Ng  and Daryl Ku

Academic Quality, National Institute of Education, Nanyang Technological University,
Singapore 637616, Singapore; daryl.ku@nie.edu.sg

* Correspondence: david.ng@nie.edu.sg

Abstract: As we increasingly emphasise the importance of developing future-ready outcomes for learners, we will need to also expand new capabilities to measure such outcomes. AI, big data, and analytics are examples of such new capabilities. Ideation is one of six habits of practice we have identified that will prepare students for the future. In this paper, we present a means to computationally appraise ideation quality as one such capability. We have developed a heuristic to appraise the ideation quality of university student essays using natural language processing, a branch of artificial intelligence concerned with the understanding of human languages. Our heuristic allows for ideation quality to be quickly quantified in the form of an ideation score. So, instead of going about the process blindly, we now have a means to provide a point of reference to allow students to give measured consideration to their ideation. Unlike a learning outcome, a future-ready habit is more of a predisposition. Consequently, it is not coherent with conventional assessments, which rather seek to evaluate than to guide. This heuristic represents an outcome of our evaluation of a new problem space in education and is, at the same time, a novel expansion into a space that exploits new capabilities.

Keywords: artificial intelligence; creative problem solving; future-readiness; ideation; innovation; natural language processing



Citation: Ng, D.; Ku, D. School Reform: New Future-Ready Quality Outcomes and Proposed Measures. *Educ. Sci.* **2023**, *13*, 1246. <https://doi.org/10.3390/educsci13121246>

Academic Editors: Youmin Xi,
Dora Ho and James Ko

Received: 1 August 2023
Revised: 27 November 2023
Accepted: 15 December 2023
Published: 17 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As schools strive to be relevant and globally connected, school reform takes on both local and international contexts. International contexts have become widely associated with comparative results from international tests, such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). Countries have assumed that attaining high scores in these tests would be a strong indicator of having a world-class education system. However, education is more than just standardised testing. Thus, education success must be measured beyond these typical achievement standards.

In Ng et al. [1], future-ready learning is encapsulated in a multidimensional framework (see Figure 1) to re-define learning outcomes based on Singapore's trajectories in the economy, society, and environment for the next 10 years. The framework identified six distinct habits of practice required of learners to meet the future challenges of Singapore (see Table 1). In this paper, we would like to propose a means to computationally appraise one of the six habits, ideation. In the context of this study, ideation can be thought of as thinking divergently and convergently, directed at innovation.

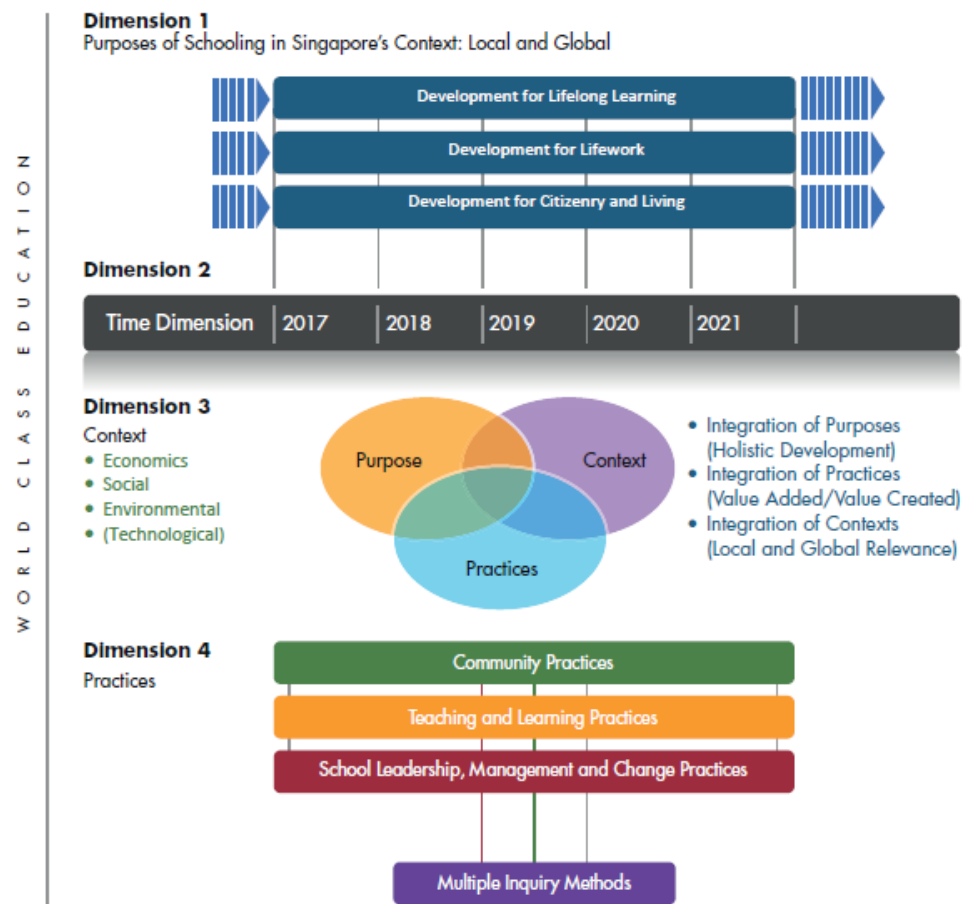


Figure 1. Multi-dimensional framework for educational success.

Table 1. Future-ready outcomes: habits of practices.

Habit	Practices	Habits of Practices Associated with Future-Ready Learning, Lifework, and Living
1. Inquisitiveness	<ul style="list-style-type: none"> Asking various kinds of questions to self and others (higher order questions, metacognitive questions, etc.), which helps individuals make meaning, reflect, and learn. Being curious. 	<ul style="list-style-type: none"> mastery of learning life-long, life-wide, life-deep learning innovation creative thinking
2. Ideation	<ul style="list-style-type: none"> Responding to stimulus (context-dependent; can be through serendipity). Requiring understanding of assumptions of practices and paradigms. Defining, clarifying, and reframing problems. Challenging the status quo. Adopting a wide repertoire of approaches to ideation that leverages on physical and virtual networks. Adopting analytics to answer big questions. 	<ul style="list-style-type: none"> innovation creative thinking critical thinking value creation mastery of learning technology adoption networks

Table 1. Cont.

Habit	Practices	Habits of Practices Associated with Future-Ready Learning, Lifework, and Living
3. Prototyping	<ul style="list-style-type: none"> • Translating ideas into action. • Evaluating, verifying, and communicating ideas and possibilities. • Allowing iteration for feedback, learning from failures. • Testing possibilities, challenges, and potential spin offs. 	<ul style="list-style-type: none"> • innovation • creative thinking • value creation • technology adoption • digitalisation • logical reasoning (critical thinking) skills
4. Entrepreneurship	<ul style="list-style-type: none"> • Seeking opportunities and piecing together information that was previously unconnected. • Developing new products and services that capture new markets/users. • Creating new uses. • Entrepreneurship occurs in a stimulating environment where diversity, multi-cultural, and differences exist. 	<ul style="list-style-type: none"> • value creation • technology adoption • innovation • digitalisation • critical thinking • networks
5. Inter-cultural Acumen	<ul style="list-style-type: none"> • Accepting diversity of values, ethnicity, and religions. • Managing conflicts and seeking optimal solutions. • Cultivating networks of collaboration and deepening ties and friendship. • Deciphering false information from real in a digitalised landscape. 	<ul style="list-style-type: none"> • tolerance • mutual understanding • mutual respect • logical reasoning (critical thinking) skills • rootedness • information assessment literacy
6. Passion	<ul style="list-style-type: none"> • Full immersion in activity and persistence in the face of obstacles. • The passionate activity becomes part of a person's identity. • Finding meaning and purpose in the activity. • Accepting failure as part of learning. 	<ul style="list-style-type: none"> • mastery of learning • life-long, life-wide, life-deep learning • innovation

2. Literature Review

Ideation is identified in the Future-ready Learning Framework [1] as one of the fundamental building blocks for innovation, creative problem solving, and value creation. It involves thinking divergently and convergently and is a part of the bigger process of creative problem solving (CPS). CPS is a type of problem solving that requires divergent thinking and is often directed at innovation.

Conceptually, ideation means coming up with ideas. But, as part of the bigger process of CPS, ideation goes beyond just coming up with ideas. When problems are complex or wicked, ideation can be directed at clarifying and framing problems during the problematising phase (where problems are originated or clarified). But it can also be directed at seeding solutions during the solutioning phase (where solutions are explored and refined). It is an activity so fundamentally intertwined with creativity and innovation that it must be unpacked and disentangled to be understood better. To address our objective, we reviewed the cognate literature with a focus on interpreting how theories can be computationally operationalised. The review inspired how we subsequently designed and developed a heuristic to computationally appraise ideation quality.

2.1. Creativity

Creativity is generally associated with avant-garde art, luck, novel ideas, or epiphanies. In the creativity literature, however, creativity is viewed less as serendipity [2,3] and more as a methodical process of *practical* problem solving [4–9]. Smith [10] emphasised that

innovation is fundamentally about progressive change. Consequently, he argued that outside of the art domain, an “idea’s craziness per se is not a virtue” (p. 350). As also asserted by Boden [11], genuinely creative ideas are about *possibilities* and not *probabilities*; they are not simply “unusual combinations” or “statistical surprises”.

Creativity is the ability to produce work that is novel, high-quality, useful, and which satisfies task constraints [9]. It involves “the production of high-quality, original, and elegant solutions to complex, novel, ill-defined, or poorly structured, problems” [12]. Ideas are therefore considered creative not just because they are avant-garde but also because they are practical.

2.2. Ideation in Creative Problem Solving

In the literature, CPS broadly refers to the employment of creativity to solve problems in the service of innovation, guided or underpinned by an organising system such as a process, model, framework, or methodology.

CPS is distinct from problem solving because it emphasises divergent thinking [5] and divergent-convergent activities [13].

If the convergent phase of problem solving is what drives us toward solutions, the objective of divergent thinking is to multiply options to create choices. . . . By testing competing ideas against one another, there is an increased likelihood that the outcome will be bolder, more creatively disruptive, and more compelling. Linus Pauling said it best: “To have a good idea, you must first have lots of ideas”—and he won two Nobel Prizes [14].

As cautioned by Smith [8], problem solving is not innovation—some problems require tried and true solutions. Innovation, on the other hand, requires a multiplicity of solutions and this is a recurring theme in the literature. One of Osborn’s [15] basic principles of ideative efficacy specifically stated to “reach for quantity”. Isaken et al.’s [16] CPS framework also specifically described striving “to produce many options, varied possibilities, and novel, or new, ideas for solving a problem or effecting a change”.

As highlighted by Mumford et al. [12], there are various CPS processes for various domains of work. We were not aiming to conduct an exhaustive review of the literature. We sought to only acquire sufficient appreciation to unpack and disentangle ideation as a constituent activity of CPS. This was needed to inform how we could computationally appraise ideation quality through algorithms instead of relying on personalities [17,18]. We focused on three main CPS organising systems in our review: Osborn’s [15] model, Isaken et al.’s [16] framework, and Mumford et al.’s [12] model.

Osborn’s [15] model consisted of three main procedures: *fact-finding*, *idea-finding*, and *solution-finding*. Ideation is mainly present in fact-finding and idea-finding. During fact-finding, the focus is on being informed while simultaneously coming up with and refining ideas about the nature of the problem. As stated by Osborn, important new ideas are only occasionally stumbled upon, so we must sometimes “originate the problem itself” (p. 87). During idea-finding, the focus is on coming up with ideas to seed solutions, then selecting feasible ones and refining them.

Isaken et al.’s [16] framework consisted of three main process components and one management component. The process components are: *understanding the challenge*, *generating ideas*, and *prepare for action*. The management component is *planning the approach*. Ideation is mainly present during understanding the challenge and generating ideas. Understanding the challenge is concerned with coming up with broad abstract ideas about what the challenge, opportunity, or concern is, as well as how it can be problematised as practical statements. This was stated to be important because problems might initially be indeterminate, with people experiencing different aspects of the same problem and therefore perceiving it differently. Generating ideas is concerned with producing many, varied, and unusual ideas, then selecting potential ones to develop further.

Mumford et al.’s [12] empirically driven model consisted of eight sequential but also interdependent processing activities. These eight activities are *problem definition*, *information*

gathering, concept selection, conceptual combination, idea generation, idea evaluation, implementation planning, and solution monitoring. There are relatively more activities ostensibly, but this is because the model does not subsume related activities under broader ones, as with Osborn's fact-finding and Isaken et al.'s understanding the challenge. In Mumford et al.'s model, ideation is predominantly present during problem definition and idea generation. Consistent with the previous two models, the definition of the problem was emphasised as important. The researchers explained that it was because an acute understanding of the problem was found to be positively correlated with solution creativity. Interestingly, Mumford et al.'s model does not explicitly elaborate on idea generation. In the journal paper [19] where this model was initially proposed, idea generation was not listed as a key activity. In their updated model, it was implied to be the act of coming up with ideas to seed and refine solutions in a manner coherent with the outcomes of having performed antecedent activities.

While there are operational differences, these three CPS organising systems all involve problematising, solutioning, and implementing the solution. It was made explicit that the production and refinement of ideas into a potential solution are distinct from practically implementing the solution. This is coherent with Couger's [13] argument that while creativity and innovation are similar, innovation is more specifically about translating ideas into products or services.

The literature implicitly recognises the role of ideation in the problematising phase but does not explicitly discuss how ideas play a role in originating or clarifying problems. Instead, emphasis is placed on how ideas facilitate solutioning. Similarly, the three organising systems above also emphasised ideation in the solutioning phase. In the context of solutioning, ideation does not simply involve coming up with ideas. Instead, it involves the seeding of solutions as well as the refinement of solutions. Some researchers explicitly differentiate between nascent ideas and refined ones. Osborn [15], for example, referred to nascent ideas as "possible leads". Isaken et al. [16], on the other hand, referred to them as "tentative images". What this also means is that an idea can be granular. It can be a discrete construct, or a composite one. This is important as unravelling what constitutes an idea is critical to its operationalisation.

2.3. Appraising Ideation Quality

Reinig et al. [20] defined ideation quality as "the degree to which an ideation activity produces ideas that are helpful in attaining a goal". While idea quantity is a key metric to evaluate ideation quality in terms of solutioning, it is not the only one. However, it is the only metric that is "objective and easy to measure" [21]. Idea quality, idea novelty, and idea variety are other metrics that can be used to evaluate ideation quality. We highlight some methods in which creativity researchers unpacked and operationalised creativity across various domains of work, starting from more theoretical perspectives and then transitioning to more practical ones.

Isaken et al. [16] proposed that in the generating ideas process, four qualities are important: *fluency, flexibility, originality, and elaboration*. Fluency is the ability to generate many ideas. Flexibility is the ability to generate a different variety of ideas. Originality is the ability to generate novel ideas. Elaboration is the ability to flesh out ideas, making them "richer, fuller, more complete, or more interesting" (p. 89). Fluency, flexibility, originality, and elaboration correspond to how the Torrance test of creative thinking [22] expresses creativity in terms of fluency, variety, originality, and elaboration, respectively.

Reinig et al.'s [20] empirical study, situated in the context of organisational problem solving, examined ideation quality in brainstorming sessions, but focused primarily on quantity and quality. In addition, they unpacked and expressed quality in terms of quantitative measurements they referred to as *sum-of-quality, average-quality, and good-idea-count*. These involved experts evaluating and scoring individual ideas on a one to four ordinal scale; the scores were then summed and averaged to determine sum-of-quality and

average-quality, respectively. Good-idea-count was determined by the number of ideas that exceeded predetermined thresholds.

Shah et al. [10] proposed that in engineering design, ideation quality can be measured in terms of *novelty*, *variety*, *quality*, and *quantity*. However, unlike Reinig et al.'s study, where experts subjectively scored ideas on an arbitrary scale, Shah et al.'s study objectively calculated novelty, variety, and quality using weighted formulas. The weights could be positive or negative in order to reward or penalise a design element. Even so, only the calculation aspect was objective. The assignment of the weights similarly required experts. The calculation of quantity was not weighted as the exception. It was simply the number of ideas aggregated. The researchers had considered the possibility that, as a metric, quantity could be inflated if ideas are not discrete enough. They subsequently explained that discreteness should be addressed through variety instead. They further advised against consolidating novelty, variety, quality, and quantity into an overall metric, as they each measure a different aspect of creativity and "adding them directly makes no sense" (p. 133).

Kudrowitz and Wallace's [21] empirical study, situated in the context of product design, also investigated ideation quality, but focused on post-early-stage ideation, with commercial viability as a focus. They proposed that *novelty*, *usefulness*, and *feasibility* can be used as metrics to quickly filter down a large selection of blue-sky ideas following prolific, early-stage ideation. Coherent with previous studies, they found that idea quantity was highly correlated with creativity, even arguing that "prolific idea generation is creative idea generation" (p. 137). Unlike Shah et al.'s study, however, Kudrowitz and Wallace's study employed laymen recruited from Amazon Mechanical Turk to conduct evaluations instead of experts. However, expertise may not be as relevant here, as the products being evaluated were toasters, umbrellas, and toothbrushes.

Kudrowitz and Wallace's study was particularly interesting because it shed further light on the notion of idea granularity that we highlighted previously. Ideas are typically portrayed as discrete and abstract concepts in the literature. But in practice, the line between an idea and a solution can be vague in some domains of work. Osborn's [15] idea association technique can be used to rearrange the elements of abstract ideas in order to produce more ideas, but it may be difficult to do the same if the idea is an actual object, material, chemical element, etc. In addition, Kudrowitz and Wallace's study also inspired us in thinking about ideation as comprising various distinct stages as well as how and where the four metrics could be more applicable at each stage. We eventually concluded that without expert personalities, only idea quantity can be used as a metric to computationally appraise ideation quality.

3. Methodology

Our objective was to develop a heuristic to computationally appraise ideation quality, which allows for human expertise to be incorporated. Ideation in this study was conceptualised as the act of thinking divergently/laterally and convergently/vertically, directed at innovation. This section discusses the thinking behind the development of our heuristic.

We employed a *distant reading* approach to analyse our data because we were primarily interested in exploring macro-level trends and patterns. This is an approach described in digital humanities as the computational study of text [23,24]. Close reading "entails close and in-depth attention to the details of a smaller section of text", whereas distant reading "involves processing (information in or about) large corpora of texts with the help of computational analysis" [25]. Distant reading "relies on automated procedures whose design involves strategic human decisions about what to search for, count, match, analyse, and then represent as outcomes in numeric or visual form" [26]. In *How We Think: Digital Media and Contemporary Technogenesis*, Hayles [27] referred to this as human-assisted computer reading, where humans use computer algorithms to "analyse patterns in large textual corpora where size makes human reading of the entirety impossible" (p. 70). According to Bode [28], "data-rich analysis has the potential to explore large-scale patterns and connections in ways that non-data-rich research cannot". Drouin [23] stated that distant

reading can reveal “concepts that might have escaped a reader’s experience, but without acknowledging the work’s historical and discursive context”. The ability to read large volumes of text objectively and quickly to explore macro-level trends and patterns was an important criterion, hence the use of distant reading.

We chose to appraise the ideation quality of university students’ essays. In our study, students had to write an academic essay on an educational leadership topic as part of their summative assessment. The assessment evaluated their ability to connect cognate concepts with the topic (a process we consider as the lateral association of concepts), and their ability to expand concepts (a process we consider as the vertical expansion of concepts). This corresponds to divergent and convergent thinking, respectively.

There was a compelling motivation behind our decision to investigate university student essays. Some courses in our university have an enrolment of more than 2000 students, making the task of marking their essays with tight deadlines challenging. We collaborated with our colleagues conducting these courses to investigate how thousands of essays could be analysed computationally. Out of the six habits we highlighted earlier, ideation quality was a metric that we deemed most suitable for exploration.

Aside from having to assess thousands of essays efficiently, another important reason was to address variability in marking standards. A course with over 2000 students and tight deadlines will require a great number of human markers with relevant expertise. Different markers will have different backgrounds and expertise, so this will inevitably translate into different interpretations of the marking rubrics, and subsequently, different marking standards. Students’ scores consequently reflect not just their performance, but also the artefacts of markers’ subjectivity. This problem is similarly present when evaluating creativity [29]. However, with the use of computational methods, instead of personalities, the artefacts of subjectivity can be attenuated while still respecting the fact that creativity does not exist independently, but is instead part of a complex system of interacting personal, social, and cultural factors [17,18,29–31].

3.1. Data

The essay assignments submitted by postgraduate students in educational leadership courses constituted part of our data. The following exemplified an actual essay assignment instruction from a doctoral-level course:

Write a scholarly paper of 10–12 pages (not including the title page and reference page(s)), consistent with expectations for graduate-level writing. The topic of the paper is “Educational Leadership for Developed Countries (Advanced Economy)”. In your paper, you are expected to discuss, critique, and explore at least three major concepts from the class discussions, presentations and reading list.

Addressing the assignment would require students to think divergently in terms of the major concepts related to the theories and practices of educational leadership. They would also be required to expand upon the concepts (i.e., think convergently) by discussing, critiquing, and exploring how these concepts and practices will meet the challenges of developed countries. The challenge imposed by such an assignment is coherent with the process of ideation, as the students could not simply just discuss, critique, and explore, without also advancing a thesis.

Our study investigated two essay datasets. *Dataset 1* is a corpus that consists of all five essays submitted by students for a doctoral-level course on “Current and Emerging Educational Leadership Theories” in 2015. *Dataset 2* is a corpus that consists of all 23 essays submitted by students for a Master’s level course on “Educational Leadership and Principals’ Theories” in 2016. These two courses shared the same assignment topic and were conducted by the first author, who also graded the 28 essays.

As our study was not a typical quantitative or qualitative research study, what constituted data must be clarified. The study involved analysing the essays submitted by 28 postgraduate participants—as part of the development of a heuristic. We were specifically interested in the artefacts produced from a treatment of their essays. Therefore, it

would not be appropriate to simply describe our study as involving a population size of 5 and 23, respectively, since there were at least 50,000 and 110,000 data points, respectively. Instead, it would be more accurate to state that our unit of analysis is an essay, and our unit of observation is a word pair—a construct we elaborate later in Section 3.2.

In addition, it must be highlighted that computers cannot understand text. To analyse text using NLP, it must first be transformed into vectors, also called word embeddings. A vector in this case would be the representation of a word. As it was not possible to collect additional data due to ethics and privacy challenges, we were limited to 28 essays. This was nowhere near enough to generate robust word embeddings. Consequently, we decided to employ the largest English language word embeddings provided by spaCy (<https://spacy.io/>, accessed on 1 January 2020). This came packaged as *en_core_web_lg-3.2.0*, which stood at approximately 741 MB. This comprised approximately 684,830 unique vectors, each with 300 dimensions. Doing so allowed us to better determine the semantic similarity of word pairs, in comparison to word embeddings generated from only 28 essays—with text all associated with the same assignment topic. As such, while there were only 28 essays, these were also essays interpreted through the lens of an exponentially larger collection of text.

3.2. Method

The discussion of our methods will be conducted at a more conceptual level as it is beyond the scope of this journal to go into the specificity of data operations.

To reiterate, we aimed to computationally appraise ideation quality in the context of university student essays. Consequently, we had to review the literature to unpack what constitutes an idea, and how the quality of an idea in the context of CPS can be evaluated, given there are no personalities, only algorithms.

We operationally defined a discrete idea as a word pair in the context of university student essays. A word pair is the result of taking what is basically the Cartesian product of the list of words in the title and the list of words in the content of an essay; but not the mathematical sense of Cartesian product where items in a set must be unique. For example, if we have an essay with the title words “T1 T2 T3”, and the content words “C1 C2 C2”, the Cartesian product would be as follows: T1-C1, T1-C2, T1-C2, T2-C1, T2-C2, T2-C2, T3-C1, T3-C2, T3-C2. Word pairs may not necessarily be unique since the same word can appear multiple times in an essay. As we were concerned with idea quantity and not idea variety, this was not considered problematic [10].

Even though our method focused only on the relationship between title and content words, we are not suggesting that ideation cannot take place locally at the sentence or paragraph level. However, investigating ideation quality at the sentence or paragraph level can become unwieldy due to the ambiguity of language. NLP problems such as coreference resolution and word sense disambiguation can be challenging not just to computers, but also humans [32]. This challenge can be further compounded given the nature of academic writing. Whether a sentence should be qualified syntactically or semantically also has to be considered, in addition to *how*. Introducing different levels of analysis for individual essays will also result in different baselines. Essays with more sentences and paragraphs will translate into a higher number of word pairs. We previously explored heuristics that involved student-prescribed titles, if any, as opposed to the assignment-prescribed title. We also explored incorporating essay-specific salient words to complement the assignment prescribed title. These resulted in some essays yielding an exponentially higher number of word pairs—enough to significantly impact final outcomes. Doing so no longer presented the analogue equivalent of students, positioned on platforms characterised by the same task parameters and context, creatively solving problems and innovating. As highlighted earlier, creativity does not exist independently. To study creativity with respect, we instead sought to indirectly observe variances in the path treaded. Therefore, we chose to focus only on the relationship between title and content words after having considered the implications.

To appraise the ideation quality of a single word pair, we first quantified the semantic similarity between its title word and its content word. This involved computing the cosine similarity between the vector representation of each word—the cosine of the angle between two vectors. The result would be a value between -1.0 and 1.0 , but most values were positive with the word embeddings we employed. Without taking into consideration the artefacts of the word embeddings employed, the cosine similarity of a word pair can be interpreted as follows:

- 1.0: Semantically the same
- 0.5: 50% semantically similar
- 0.0: No similarity

Real-world data require interpretation to take relevant factors and contexts into consideration.

Computers cannot understand text as we highlighted earlier. This is why text must first be transformed into vectors. The validity of this transformation is premised on the linguistic principle of distributional hypothesis put forward by Harris [33]. The underlying idea is that the distribution of words in a language is not random, and that the meaning of a word can be inferred through its co-occurrence with other words (p. 146). The reason we can use the cosine similarity between two vectors to determine semantic similarity is because the direction of a word's vector representation is indicative of its meaning [34].

Given that creativity and CPS fundamentally involve divergent and convergent thinking, we decided to similarly discriminate between divergent and convergent ideas. To do so, we further weighed the cosine similarity of each word pair so as to yield a more coherent unit of analysis. Table 2 below demonstrates how word pairs were grouped and ascribed a weight.

Table 2. Weights of word pairs.

Cosine Similarity (CS)	Idea Type	Weight
$0.60 < CS \leq 1.00$	Divergent	0.01
$0.45 < CS \leq 0.60$	Divergent-Convergent	0.10
$0.30 < CS \leq 0.45$	Convergent	1.00

The cosine similarity ranges indicated above are ranges that we felt best reflect the data, following discussion with the first author as domain expert. This was based on a rule of thumb approach. As we were primarily interested in the macro-level trends and patterns that would emerge, precision was not a concern and also not as relevant. This aspect of our heuristic is where we acknowledge that creativity does not exist independently, and where we incorporate human expertise.

To facilitate the calibration of these ranges by a domain expert, we constructed a “semantic ruler”—essentially a simple NLP application to determine the cosine similarity between select word pairs. See Figure 2 below for an example.

In this particular example, the first author as domain expert selected the title word “leadership” and a set of 11 content words. The content words included words the first author considers to be prominent to the essay topic, as well as words known not to be associated with leadership. This helped to establish points of reference. Explicitly knowing the cosine similarity of potential word pairs allowed the first author to better determine the cosine similarity ranges to qualify convergent, divergent-convergent, and divergent ideas.

We discarded word pairs with values below 0.30 because we considered such word pairs to be artefacts of linguistic expression rather than ideation per se. Again, this followed discussion with the first author as domain expert. As indicated earlier, we are not suggesting that ideation cannot take place locally at the sentence or paragraph level. Words that are part of a compound subject, verb, or object of the sentence may be directed at elaborating the local subject, verb, or object. Congruently, not all words in a sentence will serve to expand directly upon the title. Examples include closed-class words such as “it” and

“neither”, and proper nouns such as “John” and “Brown”. We decided that 0.30 should be the threshold where word pairs would no longer be considered sensible in the context of our data and should be treated as noise.

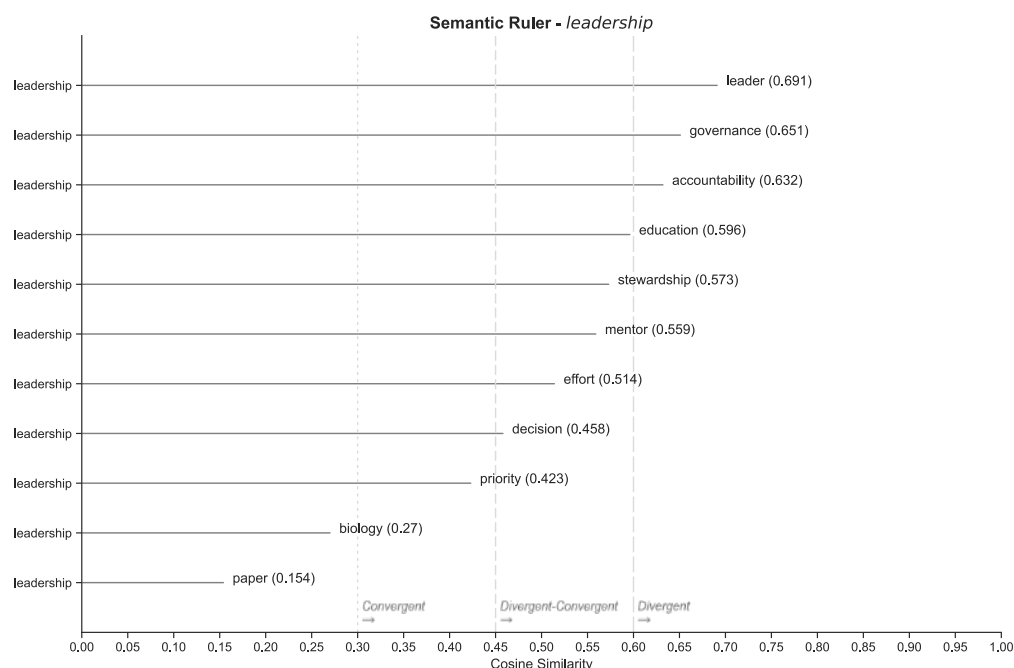


Figure 2. Semantic ruler.

Table 3 below illustrates how we conceived of divergent and convergent ideas given the title word “leadership” as an example.

Table 3. Examples of idea type.

Idea Type	Title Word	Content Word
Divergent	Leadership	Accountability
Divergent-Convergent	Leadership	Mentor
Convergent	Leadership	Priority

Divergent thinking is a critical component of creativity and CPS. It can also be thought of as the ability to think laterally. Conversely, convergent thinking can be thought of as the ability to think vertically. But divergent thinking in itself is only a means to an end. As highlighted in the literature review, creative ideas on their own are simply unusual combinations or statistical surprises. They are merely probabilities and inconsequential. In the same vein, simply presenting an assortment of ideas from the literature does not constitute an academic essay. We expect postgraduate students to additionally engage in a coherent treatment of the literature they reviewed, to engage in higher level discourse, and to ultimately advance a thesis in relation to the assignment topic. As such, we weighed convergent ideas more as these play a role that is commensurate.

To appraise the ideation quality of an essay, we subsequently added the weights ascribed to all word pairs. Table 4 below illustrates how an essay with three different word pairs would be computationally appraised for ideation quality.

Table 4. Computation of ideation quality for an essay.

Word Pair	Cosine Similarity	Weight	No. of Occurrences	Sum of Weights
Leadership-Priority	0.42	1.00	3	3
Leadership-Mentor	0.56	0.10	4	0.4
Leadership-Accountability	0.63	0.01	9	0.09
			Total	3.49

Since it can be difficult to gauge the distance in ideation quality between essays, we additionally normalised the sum of weights within a dataset so as to scale the values between 0 and 1. We refer to this final outcome as the ideation score. Table 5 below illustrates how the sum of weights for a sample dataset with only three essays would be normalised.

Table 5. Normalisation of the sum of weights across the dataset.

Essay	Sum of Weights	Ideation Score
Essay 1	3.49	0.000
Essay 2	6.21	1.000
Essay 3	3.75	0.096

As highlighted earlier, personalities were not directly involved in the appraisal process, only algorithms. In addition, we cannot independently observe creativity. There are no objective benchmarks, or upper and lower bounds in which we can take reference. However, we do know that the essays within a dataset share the exact same task parameters and context. This effectively means that the students in a course were engaged in the same innovative process and dealing with very similar constraints. This can be exploited to indirectly observe variances in the path treaded as there will be inherent differences in the students' ideation ability. So, instead of arbitrarily establishing upper and lower bounds, we let them emerge organically. A numerical score can be too granular, so it may be more meaningful to represent this categorically. For example, 0.00 to 0.33 as "Low", 0.34 to 0.66 as "Medium", and 0.67 to 1.0 as "High".

4. Analysis

To reiterate, Dataset 1 is a corpus that consists of 5 essays, and Dataset 2 a corpus that consists of 23 essays. To facilitate the development of our heuristic, we performed exploratory data analysis (EDA) on the datasets separately. This was partly to determine if they were coherent with our expectations, and the experience of the first author who graded the essays. This helped to also calibrate the parameters for our treatment.

Given all the essays were interpreted using the same word embeddings, and that they were based on very similar task parameters and context, we decided that it would not be incoherent to make basic comparisons between the two datasets.

The means of the title words and content words for the two datasets are provided in Table 6. The essays were fed into our data processing pipeline after academic references were manually removed. This constituted the data seen by the computer. In NLP, tokenisation is a preliminary process whereby text is broken up into the smallest semantic units for subsequent processing. Each unit is a "token". The tokens in Table 6 refer to the number of tokens after input data underwent pre-processing. Specifically, title-content word pairs were generated based on the tokens in the title and the tokens in the content. The number of tokens was lower than the number of words because non-meaningful tokens were discarded. The reduction was not always proportionate. The essays with the highest and lowest word count incidentally retained their positions as the essays with the highest and lowest token count. In Dataset 1, the remaining essays retained their positions. In Dataset 2, however, the majority of the remaining essays did not.

Table 6. Mean words and tokens.

	Dataset 1	Dataset 2
Mean title word	7.00	7.09
Mean title token	6.00	3.00
Mean content word	3219.80	3128.17
Mean content token	1730.60	1659.26

Figure 3 is a histogram of the frequency distribution of the cosine similarities of the word pairs in Dataset 1. Figure 4 is the histogram for Dataset 2. They were plotted with a bin width of 0.01. The specific descriptive statistics are provided in Table 7.

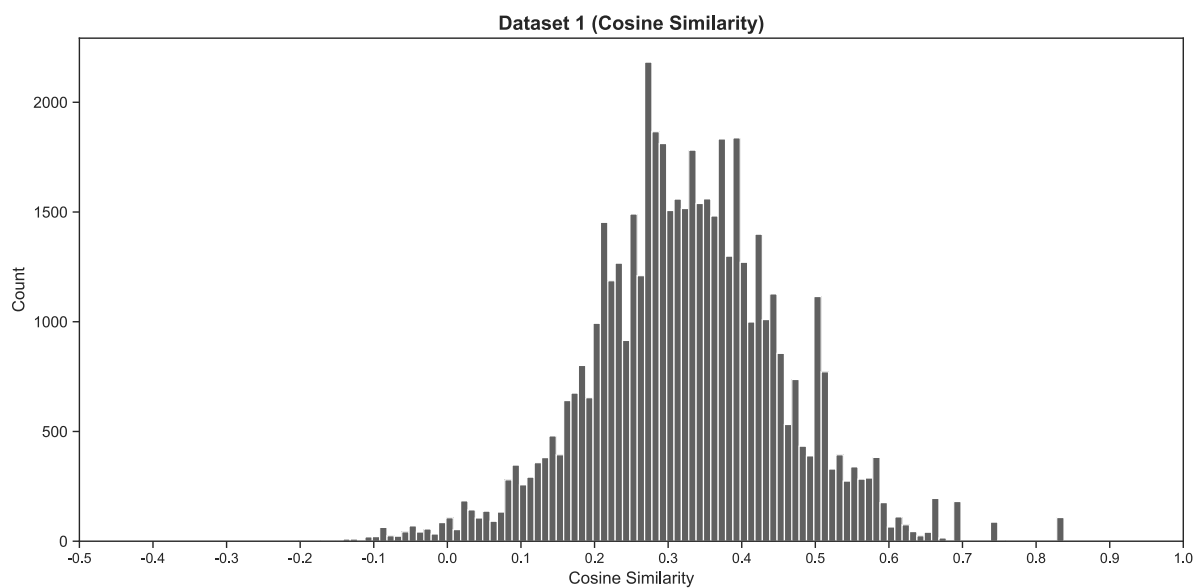
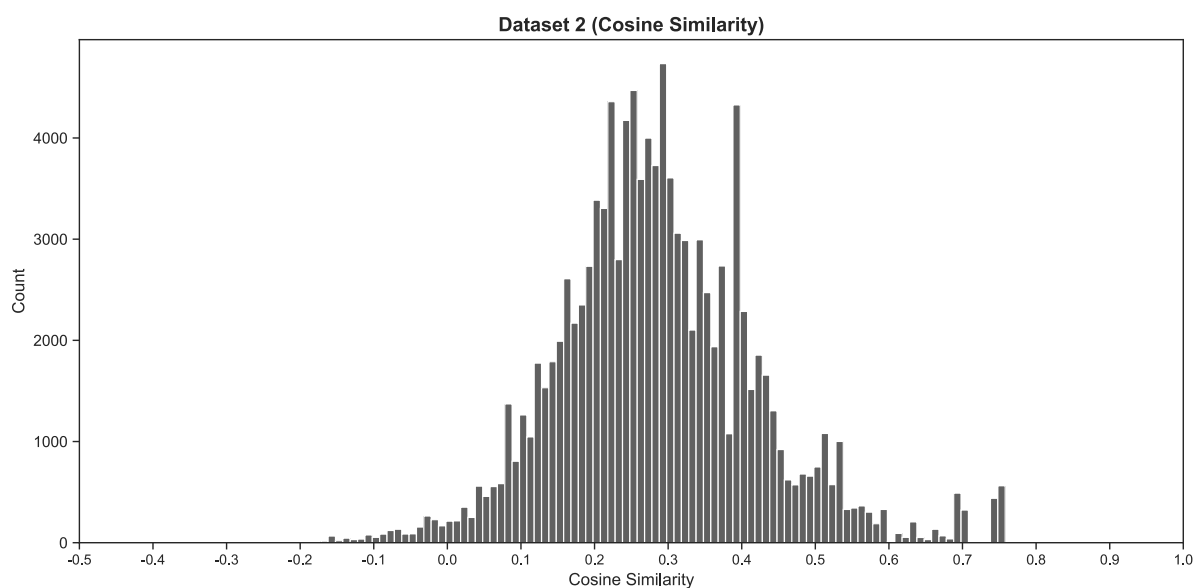
**Figure 3.** Distribution of the cosine similarities of word pairs in Dataset 1.**Figure 4.** Distribution of the cosine similarities of word pairs in Dataset 2.

Table 7. Descriptive statistics of the cosine similarity of word pairs.

	Dataset 1	Dataset 2
Count	51,379	111,968
Mode	0.28	0.39
Mean	0.33	0.29
Median	0.33	0.28
Standard deviation	0.13	0.14
Minimum	−0.22	−0.25
Maximum	0.83	0.78
Skew	0.05	0.47
Kurtosis	0.73	1.14

Figure 5 is a bar plot of the count of weights by idea type in Dataset 1. Figure 6 is the bar plot for Dataset 2.

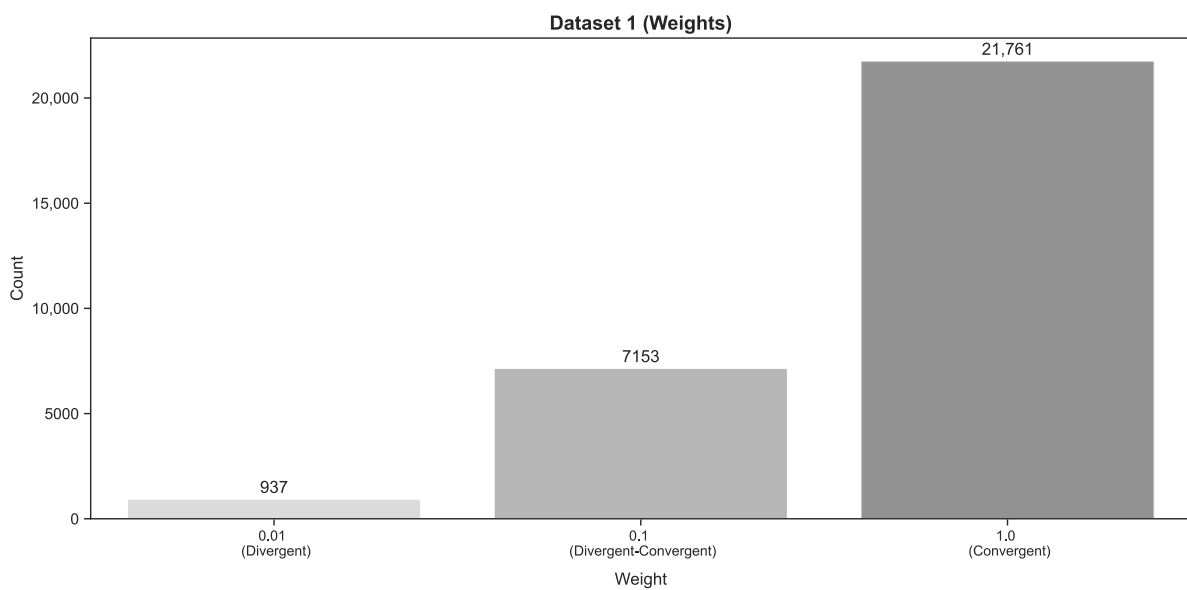
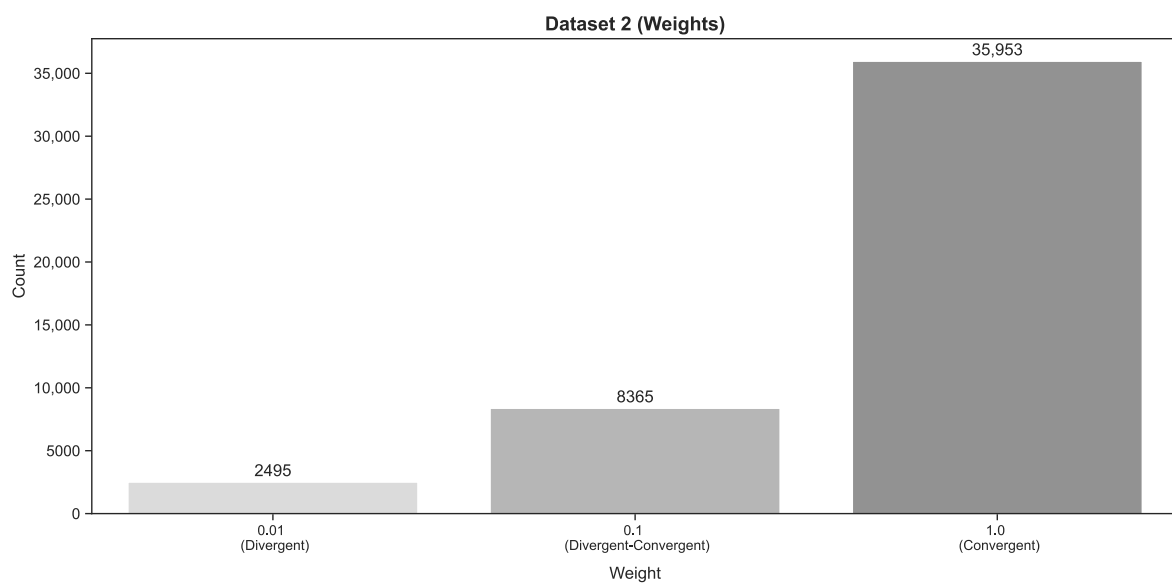
**Figure 5.** Count of weights by idea type in Dataset 1.**Figure 6.** Count of weights by idea type in Dataset 2.

Figure 7 is a bar plot of the sum of weights by essays, sorted from highest to lowest, in Dataset 1. Figure 8 is the bar plot for Dataset 2.

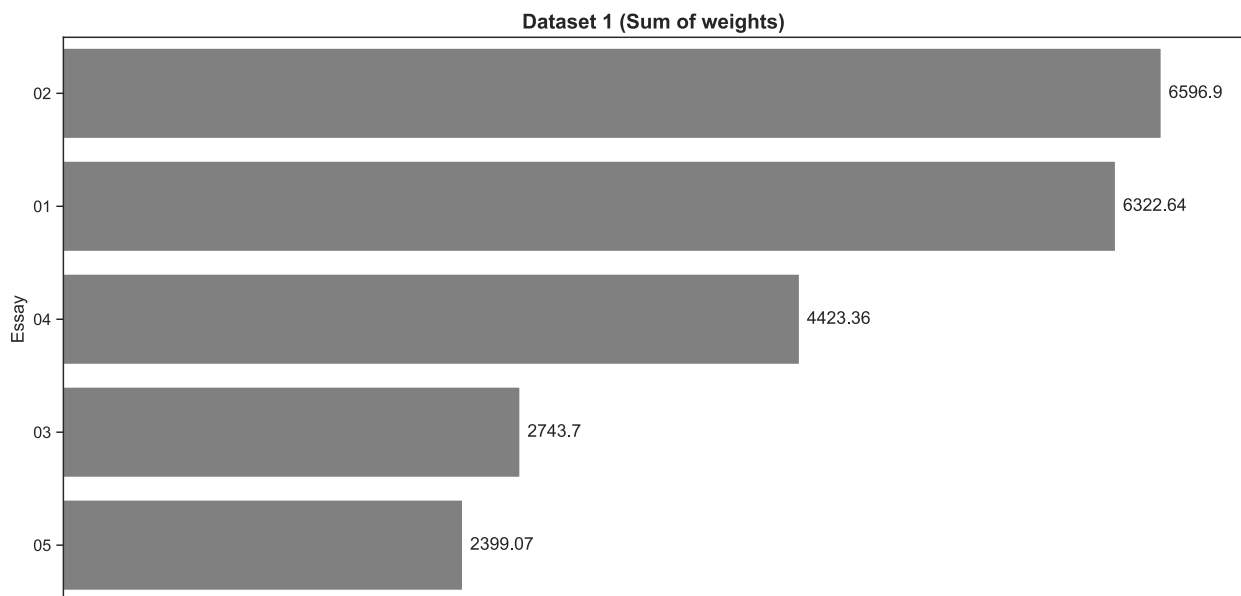


Figure 7. Sum of weights by essay in Dataset 1.

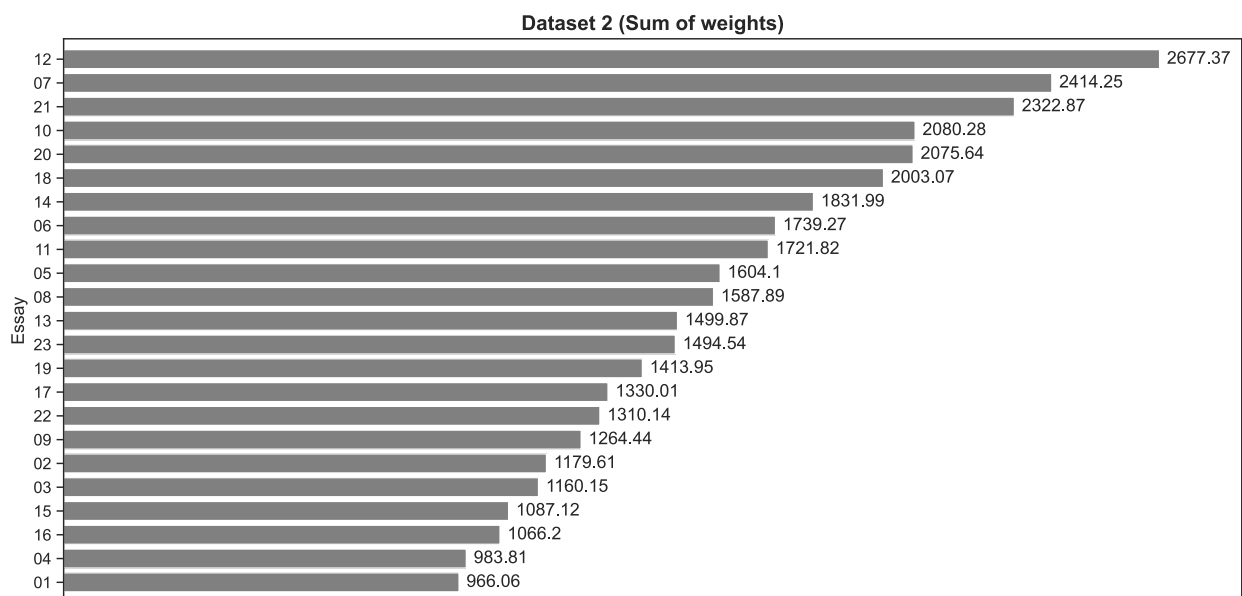


Figure 8. Sum of weights by essay in Dataset 2.

Figure 9 is a box plot of the ideation scores in Dataset 1. Figure 10 is the box plot for Dataset 2.

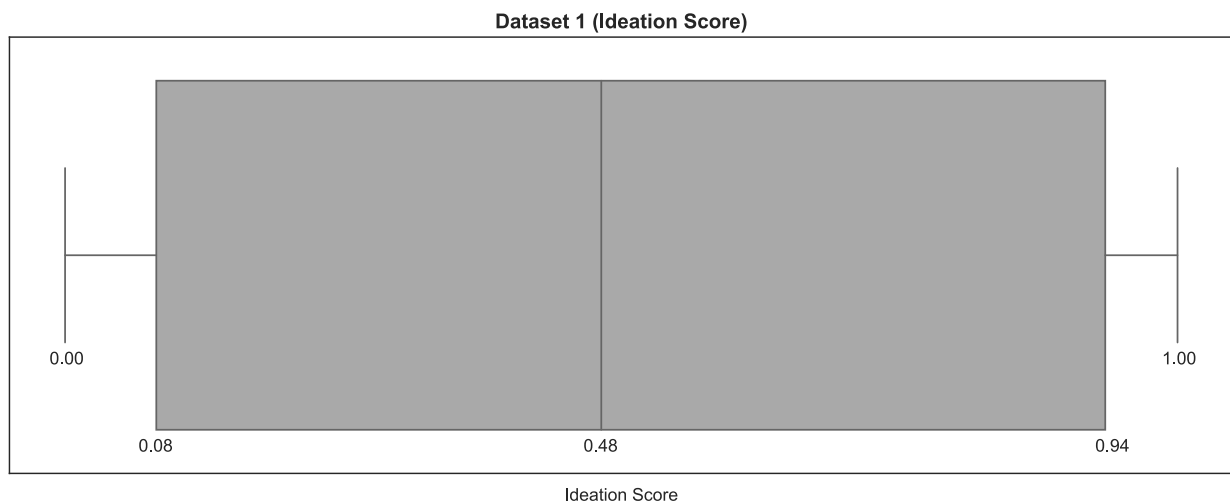


Figure 9. Box plot of ideation scores for Dataset 1.

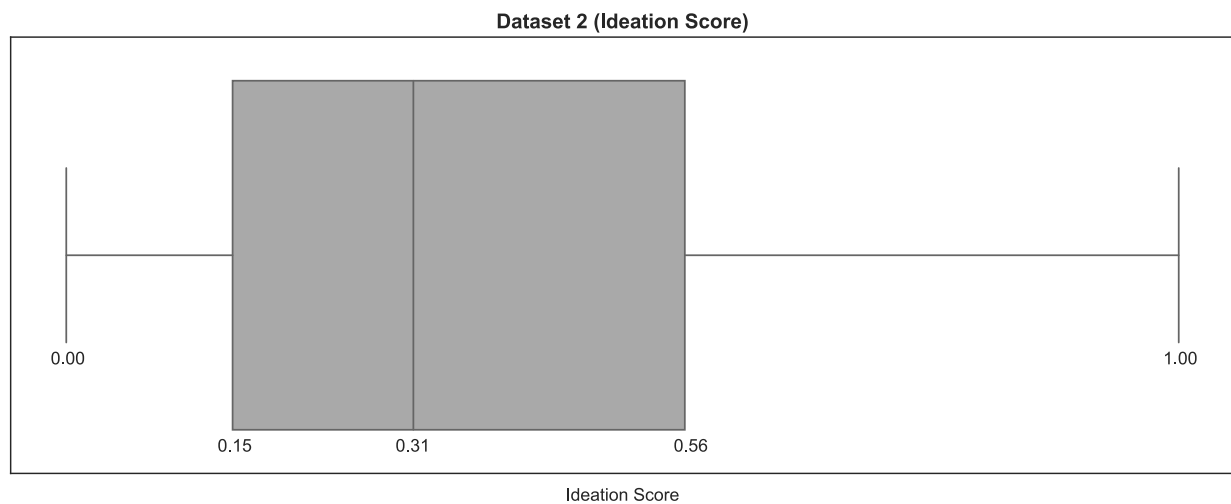


Figure 10. Box plot of ideation scores for Dataset 2.

5. Discussion

As part of the development of our heuristic, we computationally appraised ideation quality using natural language processing. The process involved consultation with the first author as domain expert. In this section, we share our interpretations of the results, as well as the insights gleaned.

5.1. Distributions of Cosine Similarities

The mean and median of the cosine similarities in Dataset 1 and Dataset 2 are almost the same. However, their distributions are only fairly symmetrical. Both datasets are slightly right skewed—more so in Dataset 2 than Dataset 1. Both datasets are light-tailed, so most values are close to the median.

Harris' [33] distributional hypothesis indicated that the distribution of words in a language is not random. As such, we were surprised to learn that the distribution of the cosine similarities were fairly symmetrical in both the datasets. If new datasets exhibit similar behaviour, this can be further exploited.

Within each dataset, we compared the distributions of the cosine similarities of the essays with the highest and the lowest ideation score, and we did not observe any significant differences. We noted that the essays with the highest and the lowest ideation score are not the essays with the highest and the lowest word or token count, although the higher scored

essay does have a relatively higher word and token count. Subsequently, we compared the distributions of the cosine similarities of the essays with the highest and the lowest word and token count. Again, we did not observe any significant differences. This indicates that the ideation score is not solely a function of word or token count in the essay, despite what we initially suspected. Nonetheless, very large quantities of word or token count can significantly impact the ideation score as we highlighted earlier in Section 3.2. In both datasets, the higher scored essay has a higher number of divergent, divergent-convergent, and convergent ideas compared to the lower scored essay.

An examination of the sum of weights in both datasets suggests that ideation quality may be higher in Dataset 1 compared to Dataset 2 (see Figures 7 and 8). The sum of weights in Dataset 1 ranges from 2399.07 to 6596.9, whereas in Dataset 2, it ranges from 966.06 to 2677.37. Given that Dataset 1 is a corpus consisting of essays submitted by students for a doctoral-level course, the results appear coherent.

We additionally examined the relationship between ideation scores and grades. The mean and standard deviation of the grades in Dataset 1 are $M = 9.2$ and $SD = 1.17$. In Dataset 2, they are $M = 8.34$ and $SD = 1.13$. Notwithstanding the small number of essays, there is a strong positive correlation between the two for Dataset 1, $n = 5$, $r = 0.77$, $p = 0.13$. However, this is not statistically significant. The grades of these five essays in Dataset 1 are provided in Table 8 below. As can be seen, the grade of an essay is not necessarily a reflection of its ideation quality.

Table 8. Grades of essays in Dataset 1¹.

Essay	Grade	Grade Point ¹	Ideation Score
01	A	10	0.94
02	A+	11	1.00
03	B+	8	0.08
04	B+	8	0.48
05	A-	9	0.00

¹ This is the grade point system used for our analysis and not the university's prescribed grade point system that does not differentiate between "A" and "A+".

For Dataset 2, there is a positive but moderate correlation, $n = 23$, $r = 0.47$, $p = 0.02$. This appears statistically significant. Nonetheless, we were not expecting academic grades to be correlated with ideation scores since ideation is more of a predisposition and therefore not congruent with academic performance.

Since there is no widely accepted reference or standard to determine the ability to ideate, we created our own standard for this study. We used the sum of weights to compare the two datasets because of their underlying similarities. As stated earlier, ideation scores facilitate comparison within a dataset. However, it would also not be incoherent to make a basic comparison between the two, given that the essays were treated the same, and that they were based on very similar task parameters and context. Given the results, we propose the following categorical representations to interpret the students' ability to ideate:

- High ideation range: sum of weights > 4500
- Medium ideation range: sum of weights 2000–4500
- Low ideation range: sum of weights < 2000

This is based on Dataset 1 as a paragon. A graphical representation of these categories for Dataset 1 and Dataset 2 is illustrated in Figures 11 and 12, respectively.

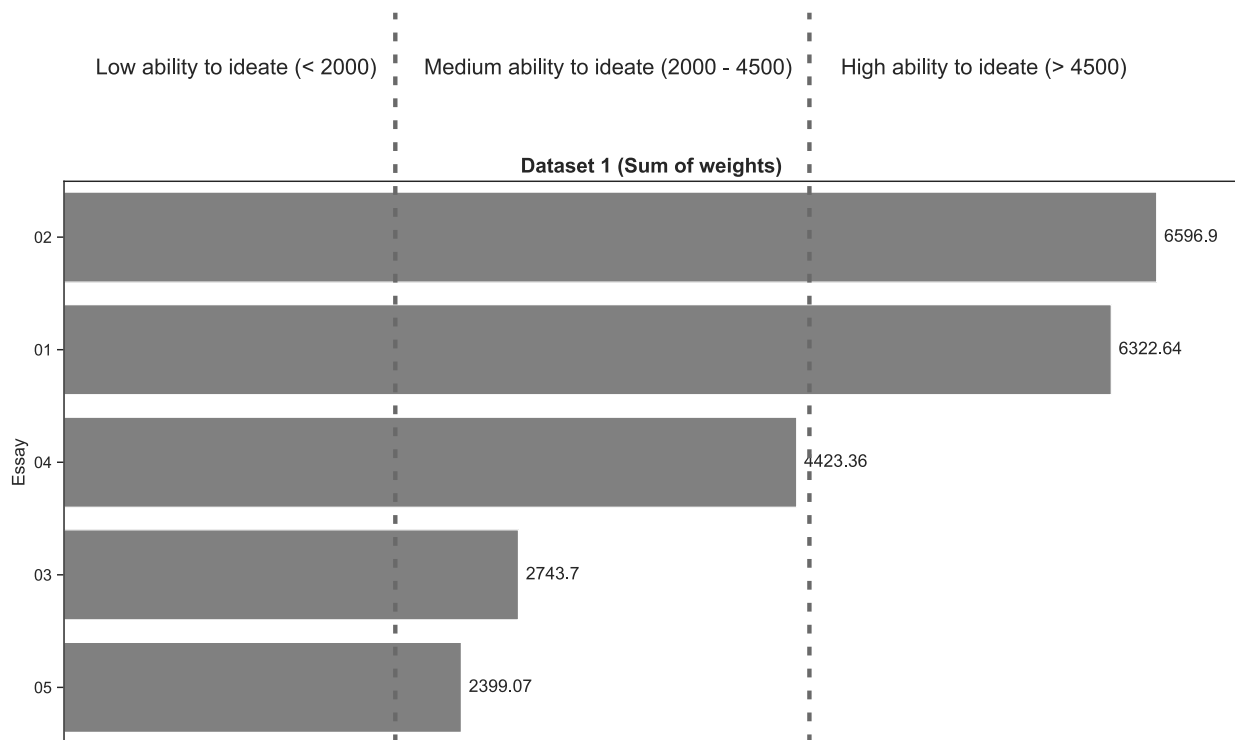


Figure 11. Ideation abilities in Dataset 1.

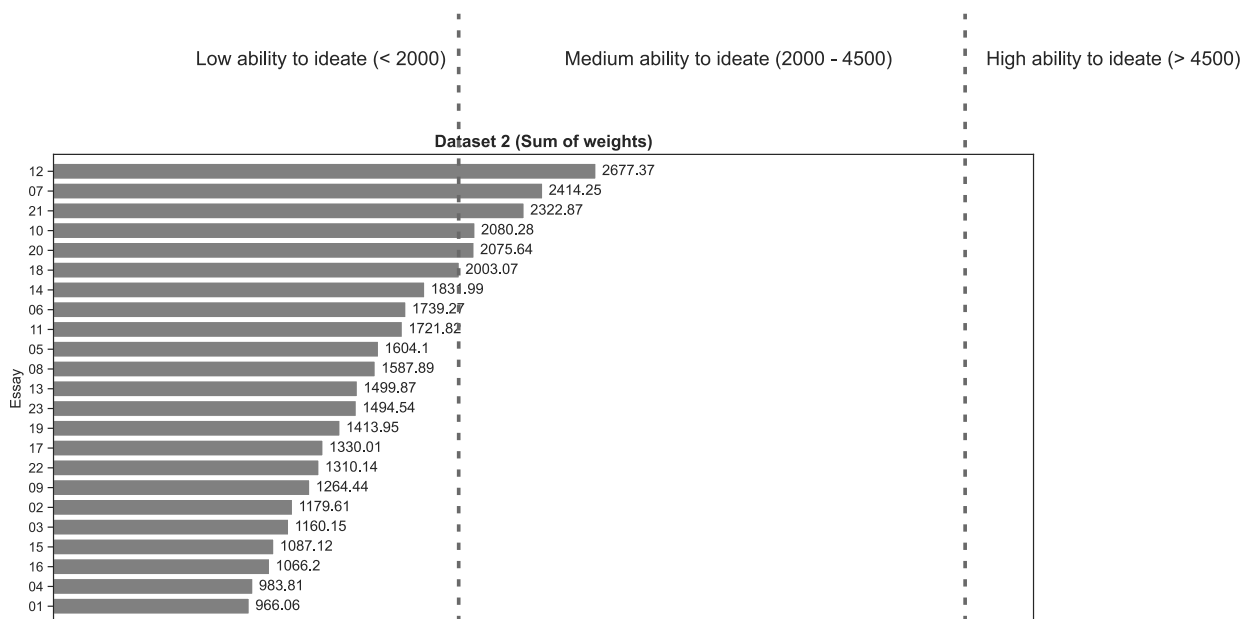


Figure 12. Ideation abilities in Dataset 2.

The results appear coherent when interpreted using our standard. We did not expect Master’s students to ideate at the level of doctoral students.

5.2. Data Quality

Data cleaning is a typical part of the data processing pipeline. It is necessary because unstructured data consists of noise that could interfere with analytical objectives. As such, it is ideal to reduce noise to a minimum. In our case, we had to manually clean part of the

data for two main reasons. First, not all essays used an orthodox citation style such as APA 7th. Some even used what appeared to be a commingling of orthodox and unorthodox styles. As such, it was not possible for academic references, which were not relevant to our analysis, to be identified and removed as noise through the pipeline. Second, data could not be readily or reliably ingested. Some essays relied on the visual layout of Microsoft Word elements such as table cells and shapes, instead of paragraphing, to present narrative coherence. Imagine constructing a sentence using characters written individually on a piece of paper. A gust of wind will easily cause the papers to be re-arranged or moved out of place. In this case, visual changes such as the font or font size will result in similar consequences. One essay in particular was written entirely in a table cell. This prevented the data from being ingested. Some essays also conflated content with presentation. For example, instead of using the appropriate indentation and spacing functionalities to style text, new paragraphs were indented using the tab key, words were spaced horizontally using double spaces, and sentences were spaced vertically using empty lines. Doing so fundamentally altered the underlying data and impacted on its integrity even though the change might not be visible to the human eye.

The scholarship quality of the essays impacted how ideation quality could be computationally appraised. Not being able to identify and remove academic references meant that our algorithm could potentially be analysing data not congruent with ideation. Our definition of an idea is a title–content word pair. Therefore, if metadata such as academic references are not removed, additional word pairs not coherent with ideation will be generated. As a result, ideation scores can be inflated.

5.3. Digital Literacy

Digital literacy impacted how ideation quality could be computationally appraised. Definitions of digital literacy [35–37] generally refer to it as the ability to use computing technology to fulfil a goal. We argue that a better definition is required. In the context of future-readiness, we assert that digital literacy is the informed use of computing technology, with due consideration for best practices in the fulfilment of a goal. Using Adobe Photoshop to write a chapter as the co-author of a book, for example, would not satisfy this definition as it would be an anti-pattern. The presentations of the essays in the two datasets indicated that the doctoral and Master’s students did not understand Microsoft Word’s role as a word processing tool—as opposed to a desktop publishing tool. Consequently, desktop publishing techniques that were antithetical to word processing were used to manipulate text. While they ostensibly attained their goal, manipulating text at the visual level is fundamentally different from manipulating text at the narrative level. Computers will increasingly work alongside humans to access, interpret, and meaningfully transform data in this age of algorithm transformation. However, computers and humans fundamentally perceive and interact with digital data differently. Whitespace characters that are invisible to humans are perceived the same as non-whitespace characters by computers. The Latin letter “A”, the Greek letter “Α”, and the Cyrillic letter “А” are indistinguishable to human eyes but not to computers. The essay that was written entirely in a table cell exemplifies the consequences of this difference in perception. As such, if text proper is not perceptible or misinterpreted by a computer, it can result in an apparent reduction or absence of ideation quality. Such an irregularity may be difficult to detect when there are thousands of essays.

5.4. Constraints

We experienced difficulties acquiring additional data to further the development of our heuristic. While our theoretical and methodological underpinnings were sound, the lack of data meant that the cosine similarity ranges we presented are specific to the two datasets. Even so, we do not think that it is possible or practical for one size to fit all. Creativity is not a domain general capacity [11,38]. Moreover, creativity tasks are knowledge- and skill-dependent [8,12]. So, while we seek to computationally appraise ideation quality, we also recognise that it is part of a complex system of interacting personal, social, and

cultural factors. This was why we felt that it is important to involve domain experts in the calibration.

5.5. Implications

The results were congruent with expectations concerning the ideation quality of doctoral and Master's students. If this had been a quantitative research study, this paper would have little to offer in terms of original contribution to knowledge. This is not our value proposition. Instead, our value proposition is a heuristic, incorporating human expertise, that we developed to computationally appraise ideation quality. This resulted in artefacts that are congruent with expectations concerning the ideation quality of doctoral and Master's students at a macro level. Furthermore, we employed free, off-the-shelf tools that can be run offline on a laptop with reproducible results, as opposed to paid online services/products with inner workings that change regularly, such as ChatGPT. To the best of our knowledge, the computational appraisal of ideation quality is something novel. As ideation in this study was conceptualised as an ongoing future-ready habit directed at innovation, as opposed to a learning outcome that lends itself to summative assessments, it is imperative that the process can be guided or at least be provided with a point of reference using pragmatic methods. Academics cannot be expected to appraise the ideation quality of their students' essay drafts full time, and consistently so. Instead of going about the process blindly, a point of reference that can be quickly acquired, such as an ideation score, may allow students to give measured consideration to their ideation. As such, our original contribution to knowledge is not the results of the study in itself, but the illumination of a little-explored problem and a corresponding pragmatic solution.

6. Conclusions

Measurements or assessments in education serve four broad purposes: (a) monitoring system performance, (b) holding schools or individuals accountable for student learning, (c) setting priorities by signalling to teachers or parents which competencies are valued, and (d) supporting instructional improvement (according to [39], cited in [40]). These purposes, which apply to assessments in general as well as to the context of developing future-ready learners, are by no means trivial.

To develop our students into future-ready individuals, we must also be ready to re-evaluate familiar problem spaces through a future-ready lens. In doing so ourselves, we identified a little-explored gap. We should also not shy away from adopting new, unconventional approaches and technology.

Author Contributions: Conceptualisation, D.N. and D.K.; methodology, D.N. and D.K.; formal analysis, D.K. and D.N.; resources, D.N.; writing—original draft preparation, D.K.; writing—review and editing, D.N. and D.K.; visualisation, D.K.; project administration, D.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Institute of Education Singapore, grant number "Project RS 10/18 DN RS-SAA".

Institutional Review Board Statement: The study was approved by the Nanyang Technological University (NTU-IRB Reference: IRB-2021-01-016).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data cannot be made publicly available. It comprises the essays submitted by a relatively small number of doctoral and Master's students in educational leadership courses and cohorts we explicitly identified. Many of these students are now in educational leadership positions. Contextual information in an essay can allow its author to be identified. Please contact the corresponding author for more information.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ng, D.; Wong, C.P.; Liu, S. *Future-Ready Learners: Learning, Lifework, Living, and Habits of Practices*; NIE Working Paper Series No. 20; National Institute of Education: Singapore, 2020. Available online: <https://repository.nie.edu.sg/handle/10497/22423> (accessed on 6 December 2023).
- Brown, T. Design thinking. *Harv. Bus. Rev.* **2008**, *86*, 84–92. [[PubMed](#)]
- Burkus, D. *The Myths of Creativity: The Truth about How Innovative Companies and People Generate Great Ideas*; Jossey-Bass: San Francisco, CA, USA, 2014.
- Hennessey, B.A.; Amabile, T.M. Creativity. *Annu. Rev. Psychol.* **2010**, *61*, 569–598. [[CrossRef](#)] [[PubMed](#)]
- Kaufman, J.C.; Plucker, J.A.; Baer, J. *Essentials of Creativity Assessment*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
- Plucker, J.A.; Beghetto, R.A.; Dow, G.T. Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educ. Psychol.* **2004**, *39*, 83–96. [[CrossRef](#)]
- Reiter-Palmon, R. Creative cognition at the individual and team levels: What happens before and after idea generation. In *The Nature of Human Creativity*; Sternberg, R.J., Kaufman, J.C., Eds.; Cambridge University Press: Cambridge, UK, 2018; pp. 184–208. [[CrossRef](#)]
- Smith, G.F. Towards a logic of innovation. In *The International Handbook on Innovation*; Shavinina, L.V., Ed.; Elsevier Science: Oxford, 2003; pp. 347–365. [[CrossRef](#)]
- Sternberg, R.J.; Pretz, J.E.; Kaufman, J.C. Types of innovations. In *The International Handbook on Innovation*; Shavinina, L.V., Ed.; Elsevier Science: Oxford, UK, 2003; pp. 158–169.
- Shah, J.J.; Smith, S.M.; Vargas-Hernandez, N. Metrics for measuring ideation effectiveness. *Des. Stud.* **2003**, *24*, 111–134. [[CrossRef](#)]
- Boden, M.A. *The Creative Mind: Myths & Mechanisms*; BasicBooks: New York, USA, 1991.
- Mumford, M.D.; Martin, R.; Elliott, S.; McIntosh, T. Creative thinking in the real world processing in context. In *The Nature of Human Creativity*; Sternberg, R.J., Kaufman, J.C., Eds.; Cambridge University Press: Cambridge, UK, 2018; pp. 147–165. [[CrossRef](#)]
- Couger, J.D. *Creative Problem Solving and Opportunity Finding*; Boyd & Fraser Publishing Company: Hinsdale, IL, USA, 1995.
- Brown, T. *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*; Harper Business: New York, NY, USA, 2009.
- Osborn, A.F. *Applied Imagination: Principles and Procedures of Creative Problem-Solving*, 3rd ed.; Creative Education Foundation: Buffalo, NY, USA, 1993.
- Isaken, S.G.; Dorval, K.B.; Treffinger, D.J. *Creative Approaches to Problem Solving: A Framework for Innovation and Change*, 3rd ed.; Sage Publications: Los Angeles, CA, USA, 2011.
- Goldsmith, R.E.; Foxall, G.R. The measurement of innovativeness. In *The International Handbook on Innovation*; Shavinina, L.V., Ed.; Elsevier Science: Amsterdam, The Netherlands, 2003; pp. 321–330. [[CrossRef](#)]
- Sass, L.A. Schizophrenia, modernism, and the “creative imagination”: On creativity and psychopathology. *Creat. Res. J.* **2001**, *13*, 55–74. [[CrossRef](#)]
- Mumford, M.D.; Mobley, M.I.; Reiter-Palmon, R.; Uhlman, C.E.; Doares, L.M. Process analytic models of creative capacities. *Creat. Res. J.* **1991**, *4*, 91–122. [[CrossRef](#)]
- Reinig, B.A.; Briggs, R.O.; Nunamaker, J.F. On the measurement of ideation quality. *J. Manag. Inf. Syst.* **2007**, *23*, 143–161. [[CrossRef](#)]
- Kudrowitz, B.M.; Wallace, D.R. Assessing the quality of ideas from prolific, early-stage product ideation. *J. Eng. Des.* **2013**, *24*, 120–139. [[CrossRef](#)]
- Kim, K.-H. The two Torrance creativity tests: The Torrance tests of creative thinking and thinking creatively in action and movement. In *Creativity: A Handbook for Teachers*; Tan, A.-G., Ed.; World Scientific: Singapore, 2007; pp. 117–141. [[CrossRef](#)]
- Drouin, J. Close- and distant-reading modernism: Network analysis, text mining, and teaching *The Little Review*. *J. Mod. Period. Stud.* **2014**, *5*, 110–135. [[CrossRef](#)]
- Goldstone, A. The doxa of reading. *Publ. Mod. Lang. Assoc. Am.* **2017**, *132*, 636–642. [[CrossRef](#)]
- van de Ven, I. Creative reading in the information age: Paradoxes of close and distant reading. *J. Creat. Behav.* **2019**, *53*, 156–164. [[CrossRef](#)]
- Drucker, J. Why distant reading isn't. *Publ. Mod. Lang. Assoc. Am.* **2017**, *132*, 628–635. [[CrossRef](#)]
- Hayles, N.K. *How We Think: Digital Media and Contemporary Technogenesis*; The University of Chicago Press: Chicago, IL, USA, 2012.
- Bode, K. The equivalence of “close” and “distant” reading; or, toward a new object for data-rich literary history. *Mod. Lang. Q.* **2017**, *78*, 77–106. [[CrossRef](#)]
- Csikszentmihalyi, M. Society, culture, and person: A systems view of creativity. In *The Nature of Creativity: Contemporary Psychological Perspectives*; Sternberg, R.J., Ed.; Cambridge University Press: Cambridge, UK, 1988; pp. 325–339.
- Altshuller, G.S. *Creativity as an Exact Science: The Theory of the Solution of Inventive Problems*; Williams, A., Translator; Gordon & Breach Publishers: New York, NY, USA, 1984.
- Sawyer, R.K. An interdisciplinary study of group creativity. In *The Nature of Human Creativity*; Sternberg, R.J., Kaufman, J.C., Eds.; Cambridge University Press: Cambridge, UK, 2018; pp. 280–290. [[CrossRef](#)]
- Liu, B. *Sentiment Analysis and Opinion Mining*; Synthesis Lectures on Human Language Technologies; Springer: Cham, Switzerland, 2012; ISBN 978-3-031-02145-9.
- Harris, Z.S. Distributional structure. *Word* **1954**, *10*, 146–162. [[CrossRef](#)]

34. Vylomova, E.; Rimell, L.; Cohn, T.; Baldwin, T. Take and took, gaggle and goose, book and read: Evaluating the utility of vector differences for lexical relation learning. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Association for Computational Linguistics: Berlin, Germany, 2016; pp. 1671–1682. [CrossRef]
35. Lankshear, C.; Knobel, M. *New Literacies: Everyday Practices and Social Learning*, 3rd ed.; Open University Press: Hong Kong, China, 2011.
36. Law, N.; Woo, D.; de la Torre, J.; Wong, G. *A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2*; UNESCO Institute for Statistics: Montreal, Quebec, Canada, 2018. Available online: <https://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf> (accessed on 6 December 2023).
37. Pangrazio, L.; Sefton-Green, J. Digital Rights, Digital Citizenship and Digital Literacy: What's the Difference? *J. New Approaches Educ. Res.* **2021**, *10*, 15–27. [CrossRef]
38. Baer, J. The trouble with “creativity”. In *The Nature of Human Creativity*; Sternberg, R.J., Kaufman, J.C., Eds.; Cambridge University Press: Cambridge, UK, 2018; pp. 16–31. [CrossRef]
39. Schwartz, H.L.; Hamilton, L.S.; Stecher, B.M.; Steele, J.L. *Expanded Measures of School Performance*; Technical Report TR-968-SANF; RAND Corporation: Santa Monica, CA, USA, 2011. Available online: https://www.rand.org/pubs/technical_reports/TR968.html (accessed on 6 December 2023).
40. Soland, J.; Hamilton, L.S.; Stecher, B.M. *Measuring 21st Century Competencies: Guidance for Educators*; Global Cities Education Network Report EP-50463; RAND Corporation: Santa Monica, CA, USA, 2013. Available online: https://www.rand.org/pubs/external_publications/EP50463.html (accessed on 6 December 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.