



Toward Defining and Assessing Creativity in Sandbox Games

Seyedahmad Rahimi, Justice T. Walker, Lin Lin-Lipsmeyer & Jinnie Shin

To cite this article: Seyedahmad Rahimi, Justice T. Walker, Lin Lin-Lipsmeyer & Jinnie Shin (2024) Toward Defining and Assessing Creativity in Sandbox Games, Creativity Research Journal, 36:2, 194-212, DOI: [10.1080/10400419.2022.2156477](https://doi.org/10.1080/10400419.2022.2156477)

To link to this article: <https://doi.org/10.1080/10400419.2022.2156477>



Published online: 03 Jan 2023.



Submit your article to this journal [↗](#)



Article views: 1294



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 20 View citing articles [↗](#)



Toward Defining and Assessing Creativity in Sandbox Games

Seyedahmad Rahimi^a, Justice T. Walker^b, Lin Lin-Lipsmeyer^c, and Jinnie Shin^d

^aSchool of Teaching and Learning, University of Florida, Gainesville, FL, USA; ^bCollege of Education, the University of Texas, Teacher Education, El Paso, TX, USA; ^cDepartment of Teaching and Learning, Southern Methodist University, Dallas, TX, USA; ^dSchool of Human Development and Organizational Studies in Education, University of Florida, Gainesville, FL, USA

ABSTRACT

Digital sandbox games such as *Minecraft* can be used to assess and support creativity. Doing so, however, requires an understanding of what is deemed creative in this game context. One approach is to understand how Minecrafters describe creativity in their communities, and how much those descriptions overlap with the established definitions of creativity in the literature. Based on prior work, we used a modified rubric for assessing a creative product in this study. We investigated the evaluation of a total of 52 *Minecraft* builds (i.e., structures that are created by the Minecrafters) from five YouTube videos. In addition to qualitatively coding the video transcripts, we also used natural language processing (NLP) for automatic coding of our data at scale. Our qualitative analysis of the YouTube transcripts revealed evidence for all but one dimension (i.e., humor) of the creativity rubric we used in our dataset. In addition, two new dimensions of creativity emerged in our qualitative coding process (i.e., realism and effort). Finally, the NLP model that we developed achieved about 80% accuracy in distinguishing the sentences that showed evidence of creativity. This study provides insights for research and educational practices of creativity in *Minecraft* or similar game or simulation environments.

ARTICLE HISTORY

Received May 31, 2022

Introduction

Creativity has been listed as one of the essential skills for success in life in the 21st-century (Glaveanu et al., 2020; Gray, 2016). New and unexpected problems in the ever-changing world that we live in require creative individuals who can provide novel solutions (Resnick, 2014, 2018). For instance, since the beginning of the COVID-19 pandemic, creativity has been quintessential to the design of effective vaccines (e.g., Tan et al., 2021). Moreover, creativity has been integral to sustaining economic infrastructures. For example, restaurants, schools, and businesses have had to come up with creative means to remain open. From an educational perspective, creativity is seen as the highest level of learning in Bloom's taxonomy; higher than remembering, understanding, applying, analyzing, and evaluating (Bloom, 1956; Krathwohl, 2002). Creativity is valuable to society. Attention has been placed on understanding the various ways creativity manifests in everyday life, the so-called little-c creativity (Kaufman & Beghetto, 2009; Richards, 2010). As such, we focused on this aspect of creativity in this study.

To support creativity, one must first accurately assess it. However, it is challenging to assess creativity (Shute & Wang, 2016). For instance, similar to beauty, creativity

can be in the eyes of the beholder. That is, it could be perceived as subjective. Some researchers have even suggested we need to drop the idea of objectively assessing creativity (e.g., Amabile, 1996) – we discuss indicators for a creative product in detail later. Moreover, traditional assessments of creativity might suffer from low ecological validity (i.e., the extent to which an assessment predicts performance in real-world settings). Thus, new environments and assessment techniques are needed to assess creativity. One promising medium to assess and enhance creativity is digital games (e.g., Rahimi & Shute, 2021a; Shute & Rahimi, 2021). Digital games, in their ubiquity, reach a broad audience around the world (Gee, 2008). However, not all digital games are successful at enhancing players' various skills and knowledge. Shute and Ke (2012) indicate seven characteristics of well-designed games that differentiate them from poorly-designed games. Those characteristics are interactive problem solving, specific goals and rules, adaptive challenges, control, ongoing feedback, the uncertainty that evokes suspense and player engagement, and sensory stimuli (i.e., a mixture of graphics, sounds, and narrative or storyline used to excite the senses). Well-designed digital games may enable people to practice or learn to be creative (Rahimi & Shute, 2021b) since

through the often-repeated practices, people can obtain creative skills that may transfer to other parts of life (Green & Kaufman, 2015).

When it comes to creativity, some digital games can improve creativity (Rahimi & Shute, 2021a; Bowman et al., 2015). There are several reasons why that is the case: (1) well-designed digital games facilitate flow (Csikszentmihalyi, 1997), and flow can facilitate creativity (e.g., Cseh, Phillips, & Pearson, 2015); (2) some digital games encourage players to create new things (e.g., sandbox games) (Gee, 2005); (3) some other games encourage players to solve problems in creative ways (e.g., puzzle games that allow multiple solutions to problems) (e.g., Inchamnan, Wyeth, & Johnson, 2013); and (4) it is intrinsically motivating to play digital games (e.g., Przybylski, Rigby, & Ryan, 2010; Rahimi et al., 2021) and intrinsic motivation is key for being creative (Amabile, 1982, 1983). Now, the question is whether there is enough empirical evidence for the positive effect of digital games on creativity in the literature.

The literature on digital games and creativity includes three categories of studies: correlational, observational or perceptual, and experimental (see Rahimi & Shute, 2021a for a systematic review). The correlational studies (e.g., Hamlen, 2009, 2013; Jackson et al., 2012) showed mixed results. For instance, Jackson et al. (2012) found a positive relationship between playing digital games and participants' creativity (i.e., the more the participants played digital games the higher their creativity scores were). In Hamlen (2009, 2013) found no relationship between the amount of time spent playing digital games and participants' creativity. These mixed results might be due to the inclusion of all digital games in these studies. That is, the researchers asked participants about their gaming habits with little or no attention to the type of games that the participants had played. Observational or perceptual studies (e.g., Inchamnan et al., 2013), however, aimed to identify the games that have more potential for improving participants' creativity. For example, comparing three puzzle digital games (i.e., *Braid*, *I-Fluid*, and *Portal 2*), Inchamnan et al. (2013) reported that all three games have the potential—not equally though—for improving creativity. Other observational studies (e.g., Cipollone, Schifter, & Moffat, 2014; Karsenti & Bugmann, 2017) also showed promising findings for *Minecraft* as a game with a high potential for improving creativity. However, these claims needed to be investigated in more controlled experimental studies. The last category of studies, experimental and quasi-experimental studies (e.g., Blanco-Herrera, Gentile, & Rokkum, 2019; Moffat & Shabalina, 2017) compared the digital games that could potentially improve participants' creativity (e.g., *Minecraft*,

SimCity, and *Portal 2*) with digital games with low or no potential to improve creativity (e.g., a shooting or racing game). Those studies showed positive results for the games that could improve creativity such as *Minecraft* and *Portal 2*.

In general, the literature on digital games and creativity indicates that not all digital games can improve creativity. Some genres such as sandbox games (e.g., *Minecraft*), puzzle games (e.g., *Portal 2*), and strategy games (e.g., *SimCity*) showed to be suitable for improving creativity. These games allow the players to add to the game, build constructions, and become game developers (Gee, 2005).

Most of the studies discussed above, consider digital games as a black box. That is, the effect of the digital game at hand was examined using a pre- and posttest of creativity. This indicates a gap that the current study is trying to address – i.e., examining the creativity that the participants produced in a digital game rather than using a traditional creativity test.

To assess creativity using digital games new assessment techniques are needed. For instance, stealth assessment (Shute, 2011) uses log files generated in an immersive environment (e.g., digital games) to make inferences about learners' levels of competency on the skill at hand (e.g., creativity). Researchers have started to use stealth assessment to objectively (i.e., automatically) assess creativity (e.g., Rafner et al. 2022 Shute & Rahimi, 2021) in various digital games. However, before a stealth assessment is designed and embedded in a digital game environment, it is essential to find appropriate indicators (i.e., observables) in the immersive environment (e.g., a digital game) to provide evidence for the competency of interest (e.g., creativity). The immersive, digital-game environment that we are focusing on in this study is called *Minecraft*.

Minecraft is one of the most popular games of all time (Gilbert, 2018). In *Minecraft*, players can use various tools to create new things, modify the game space, and choose how to engage within the game. *Minecraft's* open-ended gameplay style has made this game suitable for players to express creativity (Bowman et al., 2015). According to the literature, one's creativity levels (specifically divergent thinking skills – discussed later) are better shown through open-ended tasks or problems with multiple or even an endless number of solutions (e.g., in *Minecraft* creating a build is an open-ended problem) than close-ended tasks or problems (e.g., Fan et al., 2022; Kwon et al., 2006). In this space, players can engage in exploring the environment, generating content within the game (e.g., building structures), or interacting with other players within the space. This game has shown promising results in improving people's levels of

creativity (e.g., Blanco-Herrera et al., 2019; see Rahimi & Shute, 2021a for a systematic review on the effects of digital games on creativity). *Minecraft* includes five modes of gameplay: creative, survival, adventure, spectator, and hardcore. In the creative mode, players usually create various structures (e.g., a recreation of real-world buildings, fantasy worlds, and mini games inside *Minecraft*). These structures have been showcased online via various platforms such as YouTube which consistently ranks *Minecraft* as one of the most popular terms searched on the platform globally (Hardwick, 2020). Others have fruitfully leveraged YouTube to examine *Minecraft* productions (e.g., Walker et al., 2019.) and to find out how those productions are represented and constructed at scale. Therefore, YouTube content about *Minecraft* has a lot of potentials for creativity researchers to investigate how creativity is manifested in this digital game.

Approaches in creativity research

Creativity needs to be evaluated in a certain context. However, the degree of this contextualization varies. That is, creativity studies could be categorized as more vs. less contextualized (Sternberg & Lubart, 1996). Less contextualized studies examine creativity from a psychometrics perspective (e.g., focusing on creativity assessment leading to the creation of traditional creativity tests; Runco & Acar, 2012). For instance, the Torrance Test of Creative Thinking (TTCT; Torrance, 1974) has been widely used based on divergent thinking operationalization (Guilford, 1956). Guilford (1956) operationalized divergent thinking using its four facets: *flexibility* (e.g., the number of categories or themes used when solving a problem or the ability to come up with relevant ideas from different categories or themes); *fluency* (the ability to produce a large number of relevant ideas); *originality* (the ability to produce statistically rare ideas); and *elaboration* (the ability to implement and expand on an idea in detail and high quality). An example test item of TTCT includes turning many empty circles into something meaningful (e.g., a ball, or an apple). Then, for example, a person with more creative potential can turn more circles to a meaningful object than a person with low creative potential (i.e., the person with more creative potential scores higher on the *fluency* facet).

On the other hand, the more contextualized studies of creativity focus on the social psychology of creativity. For instance, the componential theory of creativity by Amabile (1983) arose from her investigations into the roles of motivation (both intrinsic and extrinsic), affective states, and the environment (e.g., workplace) on

creativity. Similarly, Rhodes (1961) introduced a four-P model of creativity suggesting that creativity can be understood relative to the: (1) person (e.g., personality, intellect, habits, traits), (2) process (e.g., perception, learning, thinking, and communication), (3) press (i.e., the environment that one lives in or interacts with), and (4) product (e.g., a communicated thought to others in words, paint, clay, metal, stone, fabric, or other materials).

All these approaches to studying creativity resulted in the development of creativity theories and models that can help us understand what creativity is, what it is not, and what possibly can help enhance creativity in various contexts – especially in educational settings. In this study, we focus on the approach of examining a product in a certain context. That is, we aim to identify how creative products by *Minecraft* players (i.e., a build) are explained by applying uncontextualized, psychometric methods (using a rubric based on the literature) in a highly contextualized setting (*Minecraft*). Next, we discuss the indicators of a creative product in the literature.

Creativity Indicators in a Product

A creative product is what remains after the creative process is done (Besemer & O'Quin, 1986). One way to assess a person's creativity is to judge what they created (Besemer & O'Quin, 1986). Many aspects of a creative product have been discussed in the literature (Amabile, 1983; Besemer & O'Quin, 1986; Besemer & Treffinger, 1981; Sternberg & Lubart, 1996). For instance, Jackson and Messick (1965) talked about *novelty* (unusualness), *appropriateness*, and *transformation* (i.e., new forms in a product rather than recreation of the existing products) – we discuss these three aspects in more depth later. Besemer and Treffinger (1981) reviewed more than 90 studies and found 125 criteria for a creative product. The authors then categorized those 125 criteria into 14 categories and three larger constructs. Besemer and O'Quin (1986) expands on the mentioned literature review (i.e., Besemer & Treffinger, 1981), included the 14 elements in their theoretical creativity assessment model called the Creative Product Analysis Matrix (CPAM). Those 14 elements were then grouped under three larger constructs: *novelty* (i.e., unusualness or originality), *resolution* (i.e., functionality, usefulness, appropriateness), and *elaboration and synthesis* (i.e., the details of the finished product). Adapted from Rahimi (2020), Figure 1 shows the aspects that we included for a creative product in this study. In this section, we discuss the origins of these facets in the creativity literature.

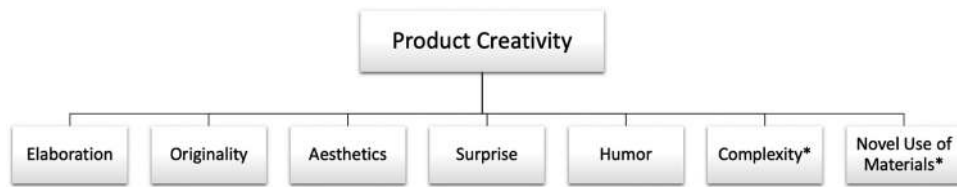


Figure 1. The initial facets for creativity in minecraft builds (adapted from Rahimi, 2020). Note. *These two categories were added to the codes from Rahimi (2020) before we started coding.

Of the seven creative facets included in Figure 1, *Elaboration* (i.e., referring to the amount of details in a product), and novelty or *originality* (i.e., the rareness of the product) come from the divergent thinking literature (Guilford, 1956). There is almost a complete consensus among the researchers that originality is essential for a creative product (Runco & Jaeger, 2012). That is, if a product is not novel, rare, or unusual, it is commonplace, and thus, not creative (Runco & Jaeger, 2012). Other creativity researchers (e.g., Amabile, 1996; Besemer & O'Quin, 1986) have also included elaboration and originality in their assessment of creative products. *Aesthetics*, however, has been subject to mixed points of view among the researchers (e.g., Amabile, 1982; Cropley & Cropley, 2011; Kharkhurin, 2014; Runco, 2003). For instance, Amabile (1982), in her consensual assessment of creativity, does not include aesthetics as one aspect of her creativity measure. Instead, she collects data on the aesthetic aspect of creativity to differentiate creativity from aesthetics. However, Runco (2003) indicates that a creative product should go beyond just being original, and it should also have aesthetic appeal. Similarly, Cropley and Cropley (2011) indicate that creativity exceeds ordinary beauty through novelty, unusualness, and appropriateness. In other words, a product is creative when it is novel, appropriate, and aesthetically beautiful. Moreover, Cropley and Cropley (2011) point out that aesthetics is not just about artistic beauty. Rather, according to Emmanuel Kant, truth and order are the essences of beauty regardless of the domain (e.g., a well-drawn painting, a well-implemented plan, and a well-written physics formula are beautiful as they show truth and order). Finally, Kharkhurin (2014) included aesthetics in the four-criterion construct of creativity and defined an aesthetic reaction as the ability to appreciate respond appreciatively to beauty. Creative products should be able to generate such reactions (Kharkhurin, 2014). Therefore, we decided to include aesthetics as one of the aspects of our creativity rubric.

Surprise refers to the reaction of an evaluator when she or he sees a creative product. According to Jackson and Messick (1965), a creative product should be *unusual*,

appropriate, and *transformational*. These three characteristics of a creative product generate corresponding responses of *surprise*, *satisfaction*, and *stimulation* in a person who sees or evaluates that product. Seeing something out of the ordinary or unusual catches our eyes and can come across as something surprising. Depending on the level of creativity, specifically, the unusualness of a product, the surprise reaction can occur multiple times. For instance, a viewer might discover various unusual aspects of a product one after another. Usually, utterances such as *wow!* *Amazing*, *impressive*, and *incredible* can be identified as evidence for a surprise reaction which indicates a form of witnessed unusualness present in a creative product. Besemer and O'Quin (1986) had a similar take on surprise by putting the transformative and attractiveness of a creative product as aspects that could evoke a surprise reaction in the viewer.

Humor can manifest itself in creative individuals by what they say, do, or create (Treadwell, 1970). Wilson, Guilford, and Christensen (1953) referred to cleverness as one aspect of originality which can tap into humor as well (i.e., a humorous idea is clever; e.g., Forthmann, Holling, Çelik, Storme, & Lubart, 2017). Usually, humor shows itself in verbal creativity (Nusbaum et al., 2016). Nusbaum et al. (2016) defined humor as the ability to generate funny ideas on the spot. Similarly, Ruch and Heintz (2019) pointed out that humor can be in various materials (e.g., jokes, cartoons, memes, or videos) and in various domains (e.g., comprehension, appreciation, or production). The humor that we included in our rubric could be seen as an appreciation of the humor in the products (*Minecraft* builds) by the evaluators (YouTube evaluators). Therefore, humor can also be shown in producing a funny product such as a build in *Minecraft*. Seeing a humorous feature in a creative product usually generates laughter, a smile, and, in general, a positive affective state (Humke & Schaefer, 1996). However, as mentioned earlier, the context of a creative task can dictate whether a sense of humor is appropriate. For instance, cartoon captioning (Treadwell, 1970) leads individuals to be humorous. However, designing a building in the real world might not make an architect use his or her sense of humor. Therefore, we do not see humor as a necessary aspect of a creative product. Since the context of our study is a digital game (thus, fun rather than serious),

we decided to include humor in our rubric and to explore whether humor would play a role in *Minecraft* builds. Note that we did not include the type of humor that was produced by the YouTube evaluators just to attract more viewers or make the video fun (e.g., when the evaluator made a joke that was not related to the build being described). We only included the humor that was related to the build being evaluated (e.g., when the evaluator said something like “this build looks funny!”).

In addition to the above facets of a creative product discussed based on a prior study (Rahimi, 2020), that is, *elaboration*, *originality*, *aesthetics*, *surprise*, and *humor*, we added two new facets: *complexity* and *novel use of materials*. *Complexity* refers to the degree of using unique, different elements to create one entity (Guilford, 1950). For example, using glass, metal, wood, and other materials to build a building makes that building a complex one in terms of its materials. *Novel use of materials* (or objects) refers to using a material in a way that was not used before or using that material in a way that is not the usual usage of that material. Amabile (1982) included both complexity and novel use of materials in her assessment of creativity. The results of her validation study indicated that products that were novel, complex, and included some form of novel use of materials were rated as highly creative by the independent raters (Amabile, 1982). Relying on the research by Amabile and considering the context of our study (creating builds in *Minecraft*), we decided to include these two facets in the list of our previous facets before starting the qualitative coding process.

Before we move on to describe the current study, we discuss one aspect of a creative product that we did not include in Figure 1: appropriateness: the extent to which a creative product is logical, applicable, and useful. This aspect is one of the most agreed upon aspects of a creative product in the literature (e.g., Amabile, 1983; Plucker & Makel, 2010; Runco & Jaeger, 2012; Sternberg & Lubart, 1996). We decided to omit appropriateness from our rubric given that we relied on builds that were already selected and thus deemed appropriate by the YouTube *Minecraft*-build evaluators who qualify as experts for this specific type of content. Therefore, we can conclude that all the builds were creative; thus, screened for appropriateness first. In the future, if any researcher wants to assess any *Minecraft* builds, they need to first evaluate the builds for appropriateness.

Current Study

Among the YouTube videos about *Minecraft* are some videos (e.g., top 10 constructions in *Minecraft*) created to showcase and evaluate top *Minecraft* builds created by other players that could be classified as creative. These

videos usually have a voice-over describing why those 10 builds, for example, are among the top *Minecraft* constructions. The evaluators' descriptions could potentially include the language used to describe creativity discussed in the literature. In addition to coding the text-based information (i.e., the narrations) for our creativity dimensions, which is a time-consuming process, we aimed to develop and validate an Artificial Intelligence (AI)-driven natural language processing (NLP) analysis framework. The framework was designed to automate the coding process for future research at scale. In this study, we will address the following research questions:

- (1) *To what extent are the evaluators' comments about the Minecraft builds consistent with the model of creativity which is based on the literature (see Figure 1)?*
- (2) *What new facets of creativity emerge, if any, from the comments narrated by the YouTube evaluators?*
- (3) *To what extent can NLP be used to analyze text-based narrative data?*

The goal of this study is to extend our understanding of the ways creativity manifests in online, sandbox digital games such as *Minecraft*; ultimately enhancing the assessments and future interventions of creativity in similar digital environments. Recently, Shaw (2022) conducted a study using *Minecraft* builds. Shaw scored each *Minecraft* build on three variables: overall creativity, novelty, and usefulness. Although our rubric has commonalities with Shaw's work (e.g., novelty which we refer to as originality, the current study expands and differs from this work in three ways: (1) we explored how creativity might manifest itself in *Minecraft* using the YouTube community that might be specific to this game only; (2) we included more variables in our rubric which can help us gain a more granular assessment of the builds; and (3) we used a Natural Language Processing (NLP) approach to automatize the coding of text data extracted from YouTube.

Various studies demonstrated validity evidence in capturing creativity dimensions from text using NLP approaches (e.g., Beaty & Johnson, 2021; Toubia et al., 2021; Zedelius, Mills, & Schooler, 2019). Toubia et al. (2021) compared different distributional semantics approaches to quantify the semantic path of the narrative stories. Applying the same methods, Beaty & Johnson (2021) evaluated the performance accuracy of the BERT model (Bidirectional Encoder Representation from Transformers; Devlin et al., 2019) to automatically score the given text based on their creativity scoring. The results showed that close to 70% of the total variance in

creativity scoring provided by the human rater, could be explained by the system predicted score. The study specifically focused on one type of creativity dimension of the connection of divergent ideas in the text.

The current study varies from the existing literature on using NLP in creativity research in three ways. First, we focus on evaluating the evidence of creativity from the discourse where the speaker is evaluating how an artifact is developed and showcasing varying types of creativity aspects. Second, our study evaluated various subcomponents of creativity (e.g., elaboration, originality, surprise). Unlike the previous literature where creativity is often scored and identified as a sole dimension, our study enriches the evidence of automatically identifying varying components of the latent construct (creativity) using the natural language processing approaches. Third, the study adopts NLP approaches to understand a relatively fewer number of corpus (text) which limits the use of more complex AI models (e.g., deep-neural models), which have showcased success in previous literature. Still, our study demonstrates how a corpus of small sample sizes could be evaluated and adopted for systematic and automated scoring using machine learning approaches.

Method

The current study uses a combination of deductive and inductive qualitative coding (Ravitch & Carl, 2019) and natural language processing (NLP) analyses to examine the transcripts of a set of evaluative *Minecraft* videos hosted on YouTube. We specifically looked for patterns of common descriptions of creative builds in the video transcripts and categorized them into the seven categories shown in Figure 1. As we will discuss in more detail later, we started coding the transcripts using the facets shown in Figure 1. We also looked for new emerging facets in our dataset. After identifying new facets, we recoded the first sub-set of our data.

Selection of videos

To find the most appropriate YouTube videos, we first consulted with two *Minecraft* experts (with more than three years of experience in *Minecraft*) to identify main search keywords, main *Minecraft* YouTubers, *Minecraft* players, and our video selection criteria. The search terms that we came up with included “Minecraft builds” “top Minecraft buildings” “best Minecraft constructions” “top ten” “top five” “Minecraft” and “creative.” The screening and inclusion/exclusion criteria for the *Minecraft* YouTube videos involved iterative steps to pare down a sample curated through targeted search terms, content, formats, and expert insights. Also, we used the YouTube search box

to conduct our search process. These steps are summarized in Figure 2 below and detailed in the next section. Our search terms yielded thousands of results on YouTube. We developed the *inclusion/exclusion criteria* to only include videos that had narrations in English and were evaluative of *Minecraft* builds. Based on our *inclusion/exclusion criteria*, we included the *Minecraft* “creative mode” and excluded the other modes such as the “survival mode.” In survival mode, players must collect resources, build structures, fight some enemies while surviving in their world. In creative mode, players can create structures, explore the environment in *Minecraft* with unlimited resources. We also excluded videos about competitions. We did not filter our search to a specific geographical place.

Using the inclusion/exclusion criteria, each author created a list of videos with the YouTube links they found. After individual screening (i.e., viewing the videos and applying the inclusion/exclusion criteria) and consulting with our *Minecraft* experts, we collectively selected a sample of 150 videos. Then, three authors went through an additional round of screening to determine the duplicates and apply the inclusion/exclusion criteria until a 100% consensus was reached. The final data set included a collection of 36 videos (about 360 unique builds). Finally, the researchers came together to define the creativity facets (shown in Figure 1) using the context of this study (see Table 1). Next, we briefly explain how we extracted the YouTube videos’ transcripts.”

Extracting the transcripts

Using YouTube’s publicly accessible application program interface (API), we wrote a code using RStudio (RStudio Team, 2016) to read the YouTube URLs from a Google spreadsheet, extract each video’s transcript, and generate a dataset with the transcripts for all the videos (see Appendix A for the code). The variables in that dataset included: video ID, caption (as they would appear on the screen), time stamp, and duration for each caption. Our final dataset included 14,860 lines of caption, 96,606 words, 6.5 words per line on average ($SD = 2.38$; $Min = 1$; $Max = 20$). These lines of caption came from the 36 videos and each video included 2,918 words on average ($SD = 1,973.04$; $Min = 512$; $Max = 9,364$). The videos were about 931 seconds on average ($SD = 607.63$; $Min = 362$; $Max = 3,233$).

Qualitative coding process

For this study, we used a combination of deductive and inductive coding processes (Creswell & Creswell, 2022). We included five videos with a total of 52 builds with 1,536 lines of caption, 10,451 words, 6.8 words per line on average

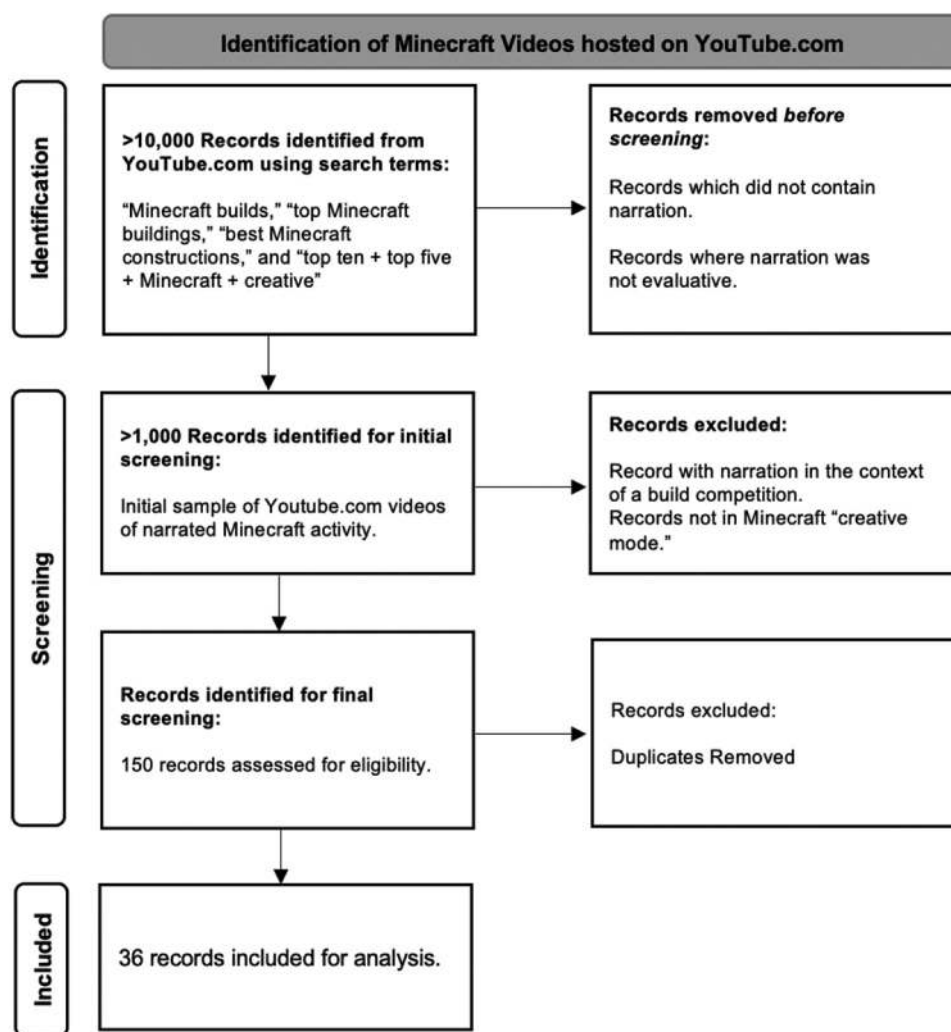


Figure 2. Videos identification, screening, and inclusion/exclusion process.

Table 1. Frequencies of product creativity codes (1536 lines of caption; 52 builds).

Codes	Freq.	Definition: Any references by the YouTube-video evaluator to ...	Examples
Surprise	116	The unexpected quality of the build.	"Wow, this is really unexpected!!" "an unknown user brought the towering structure to life in impressive form."
Elaboration	82	The level of detail in the builds.	"Each of the seven million blocks and 90,000 lights come together perfectly to create this unique, modern take."
Aesthetics	79	The quality of a build (e.g., color, design, scale, shape, feature, etc.).	"Wow, this is beautiful, look at the colors!" "Such a symmetrical shape"
Realism**	49	The resemblance or replication of a real building or place or how the build generated a call-back or remembrance.	"Wow, that is just like the Taj Mahal" "They recreated the Eiffel tower" "Wow, this reminds me of Paris' streets"
Originality	38	The rareness or uniqueness of a build.	"I have not seen something like this before." "This is a unique built."
Effort**	29	The dedication and time that the Minecrafter took or the difficulty that they had to go through to create the build.	"They pour their heart and soul into this" "and he did all this in just two weeks."
Complexity*	22	How a build was created by different elements (e.g., variations of shapes, colors, materials, items) that had an emergent quality.	"This is a complex design as the builders combined multiple materials" "Each of the seven million blocks come together perfectly to create this unique, modern take."
Novel use of materials*	13	The use of a material or an object in an unusual way.	"a blue stained-glass ceiling that mimics the ocean surface"
Humor	0	The humorous aspect of a build (e.g., when phrases such as "wow that build looks funny" or "this X [referring to an aspect of a build] is hilarious" were used).	n/a

Note. The table is sequenced based on the frequency of codes. One code could occur multiple times within the transcript of one video. * Indicates new codes that we added based on the literature and the context of this study before starting our coding process. ** indicates the codes that emerged during our coding process.

($SD = 2.99$; $Min = 1$; $Max = 20$). Each of these five videos included 2,090 words on average ($SD = 1,174.57$; $Min = 512$; $Max = 9,364$). Four of the videos included 10 builds and one video included 12 builds. On average, each build included 204 words ($SD = 120.96$; $Min = 107.50$; $Max = 411$). Three authors independently coded the transcripts and met weekly to compare and discuss the similarities and differences of our codes. We started with the existing codebook and applied the codes in Figure 1 (i.e., *Elaboration*, *Aesthetics*, *Originality*, *Humor*, *Complexity*, and *Novel use of materials*) to a subset of transcripts deductively for the unit of analysis. When we met weekly, we presented, compared, reasoned, and debated until we reached 100% consensus of our code application. Additionally, we discussed emergent themes and generated representative codes and definitions. As such, the process was also inductive. The consensus was important because it allowed us to conceptualize and apply codes uniformly. We then used the newly agreed-upon understanding of the codes to update our prior coding, and apply them to the next set of transcripts, until we completed the entire 52 builds. In the process, two new codes also emerged (i.e., *Realism* and *Effort*). We added these two new codes to the codebook (see Table 1). We made sure to recode the transcripts from the beginning to see if the two new codes could be applied to the transcripts we coded before. Through this iterative process, we coded the 52 builds both deductively and inductively. Finally, to get a sense of code distribution for each build and across the entire sample, we generated descriptive statistics in the form of frequency counts per code.

Quantitative coding process: an NLP approach

For our NLP analysis framework,¹ we adopted a word-occurrence-based text vectorization approach (e.g., term-frequency inverse-document-frequency [TF-IDF]) and a supervised learning algorithm (Kowsari et al., 2019). The analysis framework classifies the transcript based on a total of nine specific facets of creativity (the seven codes shown in Figure 1 and the two new codes that emerged in the qualitative coding process; see Table 1). To identify and restructure the data based on the unit of analysis (sentence level), we used regular expressions to capture the appropriate grammatical structure expected in sentence opening (e.g., when a punctuation, such as a period, was observed and the next letter was capitalized; for instance: “. The best build . . .”). Additionally, we used the sentence tokenization by Python 3.6’s Natural Language Toolkit (NLTK; Bird, Klein, & Loper, 2009) on a combined transcript which disregarded the timing of the discourse. The analytic process consisted of three major components.

First, we defined the unit of analysis for our NLP model from the video transcripts at a sentence level and organized

the expert-identified labels (done qualitatively) representing the underlying facets of creativity accordingly. A total of nine categories were selected, which included surprise, elaboration, aesthetics, realism, originality, effort, complexity, and novel use of material. The humor category was removed from the analysis as it had no occurrence in the transcript data. Additionally, our NLP models were not developed to distinguish and differentiate the builds within the videos.

Second, we preprocessed the dataset using the commonly adopted data-driven text processing approaches – tokenization and lemmatization (Jivani, 2011; Webster & Kit, 1992). Lemmatization attempts to map the variational form of words (e.g., goes, going) into the basic dictionary form of words (e.g., go) to reduce the noise in the text data. Specifically, we first removed the unnecessary phrases and words from the transcript, such as “-[Narrator],” and replaced the abbreviated expressions into its expanded forms, such as “don’t,” “haven’t,” “could’ve.” We then performed lowercasing and removing stop words (e.g., “the,” “a”) from the transcript prior to the vectorization by NLTK using Python 3.6. The supervised learning algorithms evaluate the linear and non-linear connections between the independent (i.e., sentence-level video transcript) and dependent (i.e., creativity facet) variables using the training sample and attempt to generalize the learned associations using the testing sample.

Third, we vectorized the pre-processed sentences from the transcripts using the term-frequency inverse-document-frequency (TF-IDF) vectorization approach. The TF-IDF vectorization approach computes the occurrence frequency ($tf_{i,s}$) for each unique word (w_i) for each sentence in the transcript, $s \in D$. Then, the relative occurrence frequency of w_i (or term-frequency $tf_{i,s}$) is compared to the inverse proportion of the word occurring in the entire transcript (df_i) as shown in equation 1. This way, the final vectorization outcome provides higher weights to the vocabularies that exclusively, but frequently, appear in the sentence.

$$w_{i,s} = tf_{i,s} \times \log(|N|/df_i) \quad (1)$$

Then, we structured two prediction learning tasks for our NLP model. The first task was to evaluate the algorithms’ capacity to distinguish the sentences that include (= 1) or do not include (= 0) evidence of creativity (i.e., the identification of sentences that could be coded for one or more than one creativity facets shown in Table 1). The second task, similarly, attempted to understand whether our algorithm could successfully label the given sentence in a video transcript to one or more corresponding categories of creativity facets. Specifically, we adopted two specific types of supervised learning algorithms for the classification tasks – the

logistic regression (LR) algorithm and the support vector classifier (SVC) algorithm with the one-vs-all strategy (Galar, Fernández, Barrenechea, Bustince, & Herrera, 2011; Rifkin & Klautau, n.d.). The one-vs-rest strategy focuses on constructing a classification model for each class (i.e., creativity facet) which is fitted against the rest of the classes (see Figure 3).

In Figure 3, we provided a conceptual representation of a scenario where each sentence (a dot) is transformed into a vector space using TF-IDF. For our algorithm to successfully identify the evidence of creativity facets (as shown in Figure 3 on the left), the algorithm uses the one-vs-all strategy (Figure 3 on the right). The one-vs-all strategy identifies one target (=1; e.g., elaboration) at a time and treats the rest of the categories as 0s (e.g., Aesthetic, Novel use of materials).

Last, we evaluated the performance of the classification models using commonly adopted classification evaluation metrics, such as accuracy and F1-score. The F1-score provides a harmonic mean of precision and recall, where a value close to 1 indicates high-performance accuracy. To ensure the generalizability and the interpretability of the performance results, cross-validation (CV) was conducted. Specifically, the logistic regression algorithm was evaluated using a 5-fold stratified cross-validation approach. The stratified CV selects the folds (i.e., part of the data) so that each class (i.e., 1 = creativity evidence, 0 = no evidence) is equally represented across each test fold. Similarly, the SVC algorithm was evaluated using multiple training and testing folds (i.e., each fold consists of the labeled sentences from one video). Iteratively, we used one of the folds to train the rest of the folds to test the classification performance of the model. This way, we could evaluate whether our supervised algorithms could transfer their classification decisions across multiple videos with different narrators.

Results

We coded five videos consisting of 52 builds (our unit of analysis) with 1536 lines of caption (i.e., each line of text or time-coded chunks of text that appears on screen as the video is playing) using open coding (Ravitch & Carl, 2019). The codes' appearances are listed in Table 1. The most frequent codes (creativity facets) were Surprise, Elaboration, and Aesthetics. We did not notice any explicit statements to be coded for humor in the 52 builds that we coded. Therefore, we can conclude that the YouTube narrators' comments were consistent with our rubric of creativity except for humor.

In the process of coding, two new codes emerged: (1) *Effort* referring to how much effort it seemed a build took to be created. For example, "almost a full year of work was put into bringing this creation to life" was coded as evidence of effort. We found 29 incidents for effort in the ten builds we coded; and (2) *Realism* which refers to how real the build was in the evaluator's view. It is hard to create real-looking constructions using blocks in *Minecraft* because the scales should be accurately estimated, and it takes a lot of detail. Therefore, when a building resembles a real building, the evaluators found it as doing something creative. We found 49 incidents for realism in the builds we coded. These findings indicate that in a new context, new aspects of creativity might matter that they may not matter in other contexts.

Qualitative review of two cases

To provide more granular detail to our descriptive analysis and creativity codes in the context, we present two illustrative example cases of how a YouTube-video evaluator appraised two *Minecraft* builds constructed by others. In the first example, the YouTube evaluator describes

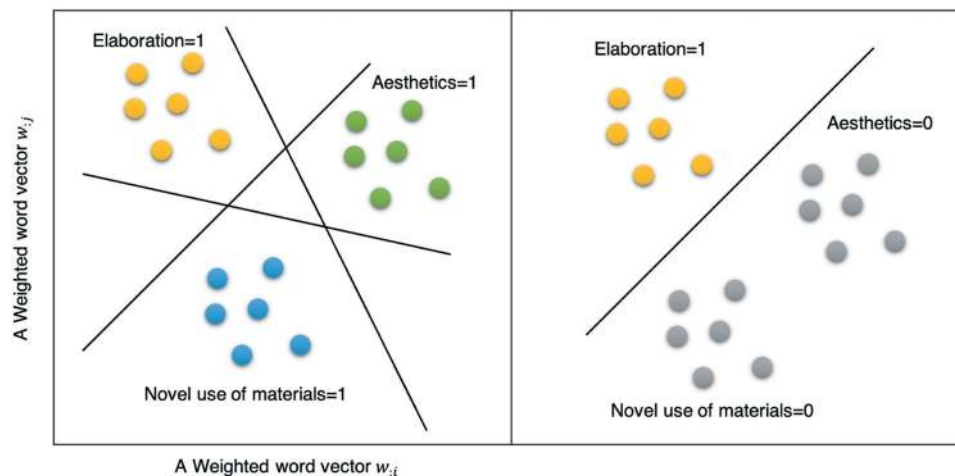


Figure 3. A conceptual representation of the one-vs-all classification approach.

a building that is meant to be a representation of the mythical tower of Babel (shown below in Figure 4(a-f)) – which is based on a religious origin story that is not depicted in illustrated human history. We raise this point to highlight the fact that this representation is born purely out of imagination since it is not a replica or actual representation of an existing structure.

The evaluator's appraisal of the structure offers a good account of how creativity manifests in *Minecraft*, as they assert:

Number 10, the Tower of Babel. The Tower of Babel may be a structure rooted in the myth of the origin of man, specifically the diversification of civilizations, but an unknown user brought the towering structure to life in impressive form [*Surprise*]. There are many incarnations of the tower in *Minecraft*, but the minute details put into this circular structure make it one of the greatest [*Elaboration* and *Originality*]. Each of the seven million blocks and 90,000 lights come together perfectly to create this unique, modern take [*Elaboration, Originality, Aesthetics, and Complexity*]. A spiral staircase winds up each of the hundred floors, ending at a glass bulb that allows visitors to view the surrounding world [*Elaboration* and *Aesthetics*].

Several constructs of creativity are coded for the above quote. They include *Elaboration, Originality, Aesthetics, Surprise*, and *Complexity*. In this summary evaluation of the structure, we observed aspects of *Surprise* as the narrator characterizes the creation as “impressive” in form – an appraisal that is not comparative in any way and yet values the structure for being developed beyond what is expected or average. The YouTube evaluator goes on to explain that while others have produced other renditions or incarnations of the mythical structure – the detailed minutiae make it stand above the others as the “greatest.” The evaluator

also provides justification for the appraisal by highlighting the number of blocks and lights, respectively on the order of millions and tens of thousands – which they argue makes this structure original or unique. We observed these characteristics to be consistent with *Elaboration* and *Originality* – because the evaluator attributes the appraisal to the number of elements (i.e., blocks and lights) composing the structure and also because of the clear comparative references that suggest this structure is unlike any others. The evaluator reiterates these ideas in their final remarks which are consistent with *Elaboration* and *Originality*. We also observed aspects of *Aesthetics* in the evaluator's narration – as the YouTube evaluator characterized the builder's creation as “perfectly” assembled and then goes on to recognize the design elements of the structure including how the staircase is arranged along with other features to enable functionality (i.e., “... a glass bulb that allows visitors to view the surrounding world”).

When examining the remarks more closely, we also observed ideas consistent with the *Complexity* facet of creativity – insomuch as the evaluator's arrival at the structure being “[great]” and of “[uniquely modern]” is the result of an amalgamation of elements, or as the YouTube evaluator notes, a “[coming] together to create [a] unique, modern take.” These results suggest that evaluator comments provide nuanced insights consistent with facets of the creativity construct as reflected in the redundant summary appraisals of the structure's elaborative and original characteristics. Importantly, we also observed in this illustrative case example something more. As a sandbox game, *Minecraft* uniquely offers expansive opportunities for users to digitally create – often beyond what is otherwise possible in the real world. As a result, emergent properties are readily actualizable and recognized (in this case by the YouTube

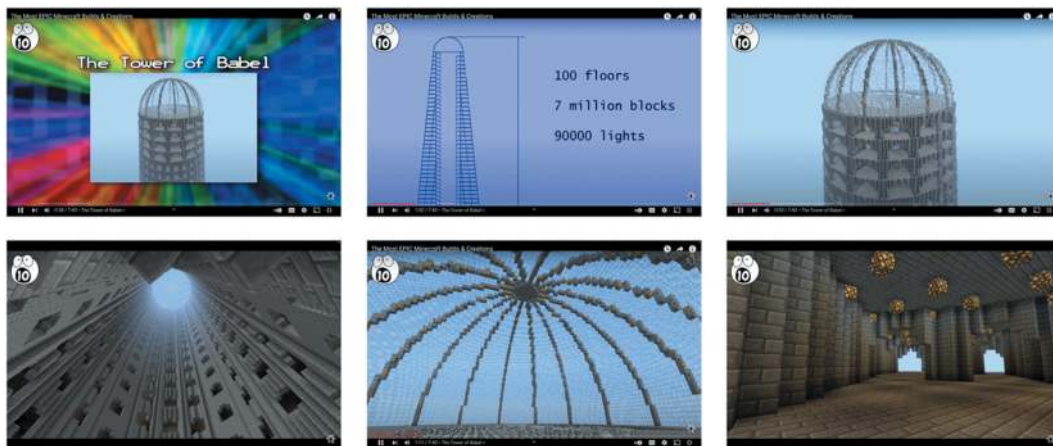


Figure 4. a-f (top to bottom, left to right). Images of tower of babel build reviewed on minecraft and appraised consistent with creativity subcodes.

Note. The YouTube evaluator describes and highlights elements that make the mythical tower a highly regarded construction in *Minecraft* – with classical elements of creativity (e.g., *Elaboration, Originality, Aesthetics*, and *Surprise*) as well as *Complexity*; a construct extending aspect.

evaluator) in the context of creativity as reflected in the evaluator's recognition that the whole of this structure is made better through the sum of its parts. This is the *Complexity* facet of creativity – defined as an emergent property derived from the coalescing of individual elements.

In our second illustrative case example, we highlight a *Minecraft* construction that is not a mythical representation of an object. Rather, it is an actual structure and documented in human history: the ancient Mesopotamian city of Babylon (shown below in Figure 5(a-f)).

Like the previous example, we observed *Elaboration*, *Aesthetics*, *Surprise*, and *Complexity* facets of creativity that we introduced in Figure 1. These attributes are evident in the YouTube evaluator's appraisal of the build as they explain:

Number eight, Babylonian City. Built by a team of players SwiftSampson and his girlfriend, this throwback to the Babylonian empire is rife with incredible details, such as the specks of color in the pillars to give the sandstone structures a little vibrancy [*Elaboration* and *Aesthetics*]. Nooks and crannies hide scattered plant life while every city light is purposefully placed to bring the metropolis to life come nightfall [*Elaboration*, *Aesthetics*, and *Complexity*]. Flanked by a nearby tower, the beauty and intricacies of the Babylonian call-back can either be viewed and admired from above or within the carved-out streets [*Elaboration*, *Aesthetics*, and *Surprise*].

As the YouTube evaluator assesses the Minecrafter's build, they place significant attention on the amount of detail reflected in the construction as the narrator characterizes the detail as “rife” and goes on to explain everything from the more granular (e.g., “nooks and crannies” and “intricacies”) attributes of the build to

more broad terms by emphasizing larger physical structures (e.g., “the pillars,” “sandstone structures,” “plant life,” “carved out streets,” etc.). These details are consistent with the elaboration aspect of our creativity construct since the evaluator almost itemizes the features of the *Minecraft* build. Concomitant with these remarks, we also observed evidence of aesthetics as the evaluator consistently appraised the features for their pleasing nature or design. For example, the evaluator included vibrant structural appeal, design perfection in light placement, and the build's overall beauty in their narration. The YouTube evaluator also suggests these attributes make it a construction to be admired – an appraisal that is consistent with the *Surprise* facet. If the evaluator did not use the word “admired” and only had used “viewed” we would not code this sentence for *Surprise*. Finally, we observed an instance of *Complexity* as the evaluator asserts that the collection of features identified come together to “bring the metropolis to life” at nightfall. This final point highlights an emergent quality that is the result of how all the elements in this build come together and are distinct from any individual part.

These illustrative case examples provide more detailed and contextualized evidence of how creativity manifests in *Minecraft*. These cases also provide evidence of how *Minecraft* provides a unique lens through which to examine creativity – even beyond classical constructs. We address these observations in relation to our research questions and the literature in the Discussion.

NLP and classification model

Our preliminary analysis using NLP and machine learning classification algorithms revealed insights about the

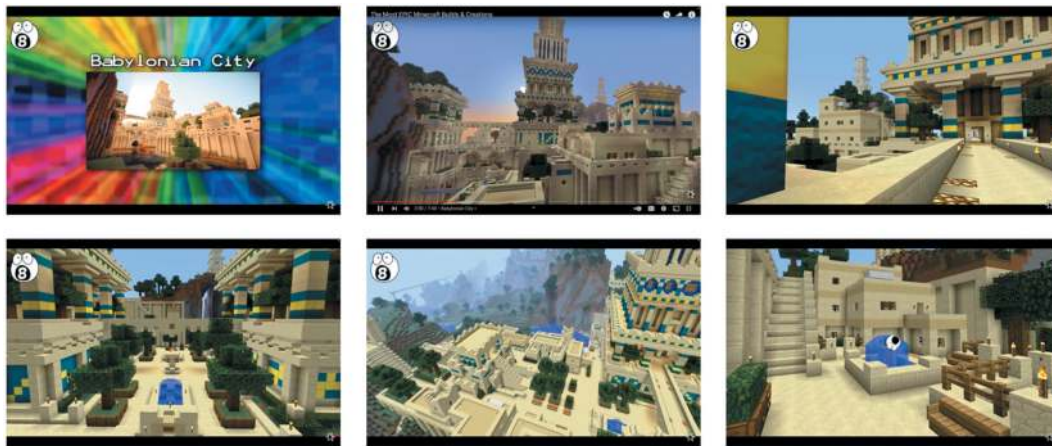


Figure 5. a-f (top to bottom, left to right). images of Babylonian city reviewed on minecraft and appraised consistent with creativity subcodes.

Note. The YouTube evaluator describes and highlights elements that make a Babylonian city replica a significant construction in *Minecraft* – with creativity aspects: *Elaboration*, *Aesthetics*, and *Surprise*.

possibility of automated coding of the evaluations provided by the YouTube evaluators. Table 2 provides the final performance evaluation results of the logistic regression and linear kernel support vector classification algorithms. The findings showed that our logistic regression algorithm could achieve close to 80% accuracy, on average based on the 5-fold stratified cross-validation results, to distinguish the sentences that should be coded for one of the facets of creativity ($\text{acc} = 0.75$, $\text{F1} = 0.70$) from the ones that were not relevant.

The results indicate that our logistic regression algorithm could perform relatively well in detecting and distinguishing the sentences that are related to at least one facet of creativity. The support vector classification algorithms, on the other hand, indicated relatively reduced performance accuracy in classifying sentences based on the specific creativity facets. Specifically, among the five facets, classifying the sentences with Elaboration and Complexity evidence achieved the highest performance ($\text{acc} = 0.82$, $\text{F1} = 0.40$).

The F1-score, especially, decreased noticeably from the overall classification results in our first model ($\text{acc} = 0.75$, $\text{F1} = 0.70$). The worst performing classifications were observed in the *Realism* and *Surprise* facets.

Error analysis

Figure 6 and Table 3 provide the results of the misclassified sentences in our logistic regression analysis. Table 3 provides a list of mislabeled sentences. The error analysis of our logistic regression algorithm showed that the majority of the misclassifications, on average, occurred with the sentences that are labeled as evidence of the *Surprise* facet (Figure 5 and Table 3).

A qualitative review of the mislabeled sentences revealed several potential challenges which may have prevented the algorithm from successfully performing the classification task. The results indicated that many misclassifications occurred due to the contextuality or the interconnectedness between the sentences in the video transcript. To avoid this issue and improve performance accuracy, future studies can focus on adopting “contextualized” and neural vectorization approaches, such as BERT (Bidirectional Encoder Representation from Transformers; Devlin et al., 2019) and other types

of sentence embeddings. This way, the meaning and the representation of the sentence will not solely depend on the occurrence of the words in the sentence. However, this will require larger sample sizes for adequate training and testing. We encourage future studies to explore the different vectorization approaches with more complex machine learning models to overcome the issues in the contextuality and interconnectedness of the sentences.

Consider the sentence “But so much amazing work has gone into this tree” (see Table 3, line 9) as an example. This sentence was not qualitatively labeled for any of the creativity facets by the researchers because the next sentence in the database expanded this sentence. Therefore, we did not want to code each sentence as a new piece of evidence for creativity as the two sentences could be seen as one. Another example could be the sentence “Impressive mega builds on *Minecraft*, that’s what” (see Table 3, line 6) which is a sentence that includes the word *impressive* and was coded by our NLP model for evidence of creativity. However, this sentence is a general statement about *Minecraft* and not a specific statement evaluating a build (our unit of analysis). It is difficult for the NLP model to identify such irrelevant sentences.

Discussion

In this study, we used a modified rubric for assessing a creative product (builds in *Minecraft*) and investigated if we could find evidence for this rubric in how *Minecraft* builds are being evaluated on YouTube. We coded the captions related to 52 builds (1536 lines of caption). We found evidence for most of the codes or creativity facets we used except for humor. We also found two new facets: *Effort* and *Realism*. These facets are significant because they highlight how platform-specific features shape how creativity manifests itself. In other words, creativity facets can in many ways be generalized, however, digital games and their diverse context may manifest creativity in different terms. Therefore, it is important to reevaluate creativity assessments considering the context of interest.

After we developed the two new facets based on our data, we investigated to see if these facets have any roots in the creativity literature. Researchers included these

Table 2. Final model classification performance results.

Algorithm	Creativity Dimension (Label)	Accuracy (SD)	F1 (SD)
LR	Creativity evidence vs. None	0.75 (0.05)	0.70 (0.04)
One-vs-Rest	Elaboration and Complexity	0.82 (0.07)	0.40 (0.01)
+	Novel use of materials and Originality	0.92 (0.01)	0.15 (0.01)
Linear Kernel	Aesthetics	0.89 (0.06)	0.34 (0.17)
SVC	Surprise	0.76 (0.04)	0.21 (0.15)
	Realism/Recreation	0.92 (0.02)	0.22 (0.01)

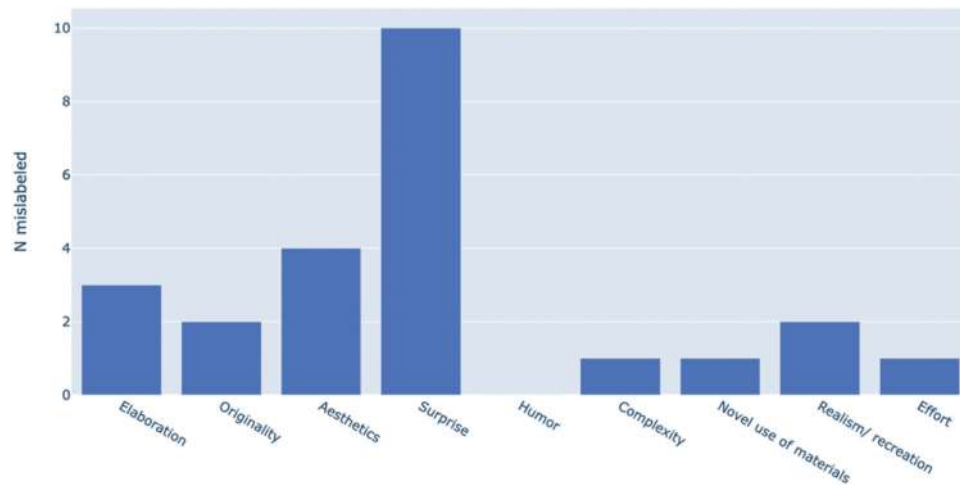


Figure 6. An average occurrence of misclassification across the nine creativity dimensions.

Table 3. Misclassification Error Analysis Results.

Line	Sentence	True (Y)	Pred ($\hat{Y}$)
1	"Number one, the entire country of Denmark."	1	0
2	"Just how big is it you ask?"	1	0
3	"So I think what I am trying to say is the man is an absolute madman."	1	0
4	"But the center of the base that ruined section uh all came from mambo's mind"	1	0
5	"Number aside, the masterpiece is really expansive."	1	0
6	"Impressive mega builds on Minecraft, that's what."	0	1
7	"There's something mystical about the combination of a desert, an oasis, and a towering havoc."	0	1
8	"Yep, there's that too and so much more."	0	1
9	"But so much amazing work has gone into this tree."	0	1
10	"This is Coldplay playing paradise and I just absolutely love the idea."	0	1

two facets in different contexts. For Amabile (1982, 1983) included *Effort* as one of the aspects that creativity raters had to evaluate for a creative product. She referred to *Effort* as "effort evident." What we called *Realism* is referred to as "imitative recreation" by Ward and Sonneborn (2011). Ward and Sonneborn (2011) investigated how creativity manifests in a virtual world called *Second Life*. One of the questions they asked was related to *Realism*: can *Realism* in *Second Life* (i.e., the recreation of real-world's landmarks) be seen as evidence for creativity? One could argue that recreation of the real-world structures could not be considered as creative or original. However, Warn and Sonneborn argue that as realistic paintings of famous painters can be considered creative, so too can the recreation of real-world buildings or landmarks. Similarly, Wright, Boria, and Breidenbach (2002), and Bowman et al. (2015) consider using game tools to create real-world buildings or places with high accuracy as creative. We add to this argument by saying that the accurate recreation of a real-world landmark in *Minecraft* using those simple blocks and materials can lead to the viewer's surprise reaction (i.e., creating something which looks just like the real structure using some simple blocks) could be considered creative. Moreover, the recreation of a real-world

structure in a digital world needs creative thinking as the builder needs to figure out ways that are not similar to how the actual building was built in the real world. For example, the recreation of the Taj Mahal in the real world might not be considered as creative, however, the recreation of the Taj Mahal in *Minecraft* (if no one else had done so) can be considered as creative. One of the surprising, original, and thus, creative recreations in *Minecraft* is the recreation of the entire country of Denmark (Farokhmanesh, 2014). Ward and Sonneborn (2011) indicate that an accurate virtual recreation can give the viewers an aesthetically pleasing experience. Moreover, Flint, Hall, Stewart, and Hagan (2018) pointed out that even though the accuracy of the recreations in *Minecraft* is reduced (due to the pixelated look of the game), the viewers can still produce personal and affective responses when they encounter a well-designed recreation. Therefore, *Realism* could be one way of expressing creativity in *Minecraft*.

Moreover, our NLP analysis with classification algorithms could successfully identify the syntactic and semantic differences among the creativity dimensions. However, the analysis was conducted with a relatively small set of sample sentences for both training and evaluation, which may limit the generalizability of the findings. Hence, we

encourage a future focus on expanding the labeled dataset to evaluate the capacities of our algorithms for automated coding. Moreover, we found that classifying the sentences for Realism and Surprise was not successful compared to Elaboration and Complexity. This finding highlights the methodological limits inherent in using NLP and similar approaches that use data sets that are dynamic or evolving by nature. Because contexts within which data are produced have evolved, because datasets change – for instance, because there are new data, or because the construct used to conceptualize an idea has shifted – it requires we be upfront about both the methodological utility and predictive limits NLP and similar approaches have in understanding phenomena. For example, the nature of the space (*Minecraft*), the nature of the data sources (language/sample), and the nature of the construct (creativity) are dynamic which leads to a complex situation for using NLP. In our case (i.e., assessing and defining creativity in *Minecraft*), it means there must be ongoing recognition or appreciation for how data sets evolve and why context matters in describing phenomena. Like most conceptualizations in the learning sciences – appreciating this constraint reinforces sociocultural notions that what we know about learning processes and ideas shift and will thus always be context-based – shaped by practices, ideas, and cultural values of the time. The same notions exist for creativity sciences (Glaveanu et al., 2020). We thus attribute this finding to the nature of the data set we collected, and the potentially transient nature of creativity constructs in sandbox platforms like *Minecraft*.

Finally, when we coded our data in some cases, we had to code a sentence for more than one code (e.g., coding a sentence for both Surprise and Elaboration when the evaluator mentioned “the details of this build are impressive”). Although this might look like a disadvantage psychometrically (e.g., the creativity dimensions may not be distinguishable sometimes), the co-occurrence of codes for one sentence happens a lot in qualitative research. Moreover, the literature on divergent thinking shows strong correlations among the sub-facets of divergent thinking (Runco & Acar, 2012) which means these dimensions should have a meaningful relationship if they are measuring one construct (i.e., divergent thinking). However, if they correlate too much (e.g., fluency and flexibility) we will have a problem as they almost become one. This problem then needs to be dealt with mathematically (e.g., averaging the two dimensions). Therefore, we do not see the co-occurrence of codes in one sentence as a disadvantage, rather it can be related to the complex nature of creativity (Csikszentmihalyi, 1997). Next, we discuss the limitations and future research.

Limitation and future research

This study faced several limitations that could be addressed in the future. First, our initial plan was to hand-code the transcripts of the 52 builds and then use NLP to code the rest of our dataset. However, since we used the AI-generated transcripts from YouTube, we ended up with some transcripts that were difficult to use (i.e., without proper punctuations, errors in transcripts, etc.). We expect to see more accurate AI-generated transcription soon on platforms such as YouTube, and in turn, using approaches such as NLP would help us conduct more in-depth, and at scale analysis of similar data see (Appendix B).

Moreover, we included a few videos in our study (due to the issue mentioned above) which can be improved in future research. However, the unit of analysis in this study was the builds in each video totaling 52 builds with more than 200 words per build on average which is a reasonable number. Also, our goal was to find evidence for how a creative product in *Minecraft* is being described by this game’s players or evaluators. Despite the low number of videos we included, we found ample evidence related to most of the aspects of the creativity facets. Also, we were able to identify two new creativity facets. Moreover, the NLP model we developed could identify the sentences that could be coded for creativity facets with high accuracy. Therefore, we do not think that the number of videos included in our study threatens our findings.

In the future, we could investigate the relationships among the codes and find out which codes co-occur frequently with which codes, or which ones are the precondition for which codes. For instance, we could answer questions such as “is Effort necessary for a build to be realistic?”. Finally, the eight facets we introduced in this study (except for humor; see Table 1) could be used to assess and foster creativity in *Minecraft* especially when the purpose of using *Minecraft* is to build creative structures. Before doing so, we propose conducting some validation studies using the eight facets of creativity we introduced in this study. That is, having a number of individuals complete an external, relevant measure of creativity (i.e., the figural Torrance Test of Creativity; Torrance, 1974) and then ask them to create a creative structure in *Minecraft*. Then, using the facets we provided here score the creativity of each build. If what we proposed here is related to creativity, the participants’ scores on the external measure of creativity should have a significant, positive relationship with the *Minecraft*-creativity-assessment scores.

Conclusion

In this study, we used and modified an existing operationalization of creative products (Rahimi, 2020) to examine how *Minecraft* experts evaluate this game's builds on YouTube. We found evidence for most of our creativity facets. Moreover, we found two new facets that expanded our operationalization of creativity. *Minecraft* is one of the most popular games in the world with a lot of affordances for creativity. Creativity researchers can use this game to assess and foster creativity. The findings of our study can facilitate the research on creativity in *Minecraft* and similar platforms. The more we understand how creativity manifests itself in various environments, the more accurately we can assess it and foster it.

Note

1. See <https://github.com/jinnieshinufl/Creativity-mining-minecraft> to access the Python code we used.

Acknowledgments

We would like to thank the Learning Analytics in STEM Education Research (LASER) Institute for providing a wonderful networking opportunity, which enabled the authors to learn and collaborate on this project. We also would like to thank Mohammadali Agharazidermani and Dr. Eric Kildebeck of UT Dallas, the *Minecraft* experts who helped us identify the videos we needed for this study.

Disclosure statement

We wish to confirm that there are no known conflicts of interest associated with this manuscript.

ORCID

Seyedahmad Rahimi  <http://orcid.org/0000-0001-9266-758X>

Justice T. Walker  <http://orcid.org/0000-0002-4356-0396>

Lin Lin-Lipsmeyer  <http://orcid.org/0000-0002-9630-9744>

References

- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43(5), 997–1013. doi:10.1037/0022-3514.43.5.997
- Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45(2), 357. doi:10.1037/0022-3514.45.2.357
- Amabile, T. M. (1996). *Creativity In context: Update to the social psychology of creativity* (New ed.). New York: Routledge.
- Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2), 757–780. doi:10.3758/s13428-020-01453-w
- Besemer, S., & O'Quin, K. (1986). Analyzing creative products: Refinement and test of a judging instrument. *The Journal of Creative Behavior*, 20(2), 115–126. doi:10.1002/j.2162-6057.1986.tb00426.x
- Besemer, S. P., & Treffinger, D. J. (1981). Analysis of creative products: Review and synthesis. *The Journal of Creative Behavior*, 15(3), 158–178. doi:10.1002/j.2162-6057.1981.tb00287.x
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. Sebastopol: O'Reilly Media Inc. 3.7.
- Blanco-Herrera, J. A., Gentile, D. A., & Rokkum, J. N. (2019). Video games can increase creativity, but with caveats. *Creativity Research Journal*, 31(2), 119–131. doi:10.1080/10400419.2019.1594524
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives, handbook 1: Cognitive domain* (2nd ed.). Boston: Addison-Wesley Longman Ltd.
- Bowman, N. D., Kowert, R., & Ferguson, C. J. (2015). The impact of video game play on human (and Orc) creativity. In G. P., Green, J. C., Kaufman (Ed.), *Video games and creativity*. (pp. 39–60). Cambridge, Massachusetts: Academic Press / Elsevier. doi:10.1016/B978-0-12-801462-2.00002-3
- Cipollone, M., Schifter, C. C., & Moffat, R. A. (2014). *Minecraft* as a creative tool: A case study. *International Journal of Game-Based Learning*, 4(2), 1–14. doi:10.4018/ijgbl.2014040101
- Creswell, J. W., & Creswell, J. D. (2022). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: SAGE Publications.
- Cropley, D. H., & Cropley, A. J. (2011). Aesthetics and creativity. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of Creativity* (pp. 24–28). Cambridge, Massachusetts: Academic Press/Elsevier. doi:10.1016/B978-0-12-375038-9.00004-2
- Cseh, G. M., Phillips, L. H., & Pearson, D. G. (2015). Flow, affect and visual creativity. *Cognition and Emotion*, 29(2), 281–291. doi:10.1080/02699931.2014.913553
- Csikszentmihalyi, M. (1997). *Creativity: Flow and the psychology of discovery and invention*. New York, NY: HarperCollins Publishers.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Vol1, Minneapolis, Minnesota: Association for Computational Linguistics. 4171–4186. (arXiv:1810.04805). arXiv. doi: 10.48550/arXiv.1810.04805.
- Fan, Y., Lane, H. C., & Delialioğlu, Ö. (2022). Open-ended tasks promote creativity in *Minecraft*. *Educational Technology & Society*, 25(2), 105–116.
- Farokhmanesh, M. (2014, April 24). The entire country of Denmark has been recreated in *Minecraft*. *Polygon*. <https://www.polygon.com/2014/4/24/5650592/the-entire-country-of-denmark-has-been-recreated-in-minecraft>

- Flint, T., Hall, L., Stewart, F., & Hagan, D. (2018). Virtualizing the real: A virtual reality contemporary sculpture park for children. *Digital Creativity*, 29(2–3), 191–207. doi:10.1080/14626268.2018.1511601
- Forthmann, B., Holling, H., Çelik, P., Storme, M., & Lubart, T. (2017). Typing speed as a confounding variable and the measurement of quality in divergent thinking. *Creativity Research Journal*, 29(3), 257–269. doi:10.1080/10400419.2017.1360059
- Galar, M., Fernández, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44(8), 1761–1776. doi:10.1016/j.patcog.2011.01.017
- Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, 2(1), 5–16. doi:10.2304/elea.2005.2.1.5
- Gee, J. P. (2008). Learning theory, video games, and popular culture. In K. Drotner & S. Livingston (Eds.), *The international handbook of children, media, and culture*. (pp.196–212). London: SAGE Publication Inc.
- Gilbert, B. (2018). “*Minecraft*” is still one of the biggest games in the world, with over 91 million people playing monthly. Business Insider. Retrieved June 20, 2021, from <https://www.businessinsider.com/minecraft-has-74-million-monthly-players-2018-1>
- Glaveanu, V. P., Hanchett Hanson, M., Baer, J., Barbot, B., Clapp, E. P., Corazza, G. E., ... Sternberg, R. J. (2020). Advancing creativity theory and research: A socio-cultural manifesto. *The Journal of Creative Behavior*, 54(3), 741–745. doi:10.1002/jocb.395
- Gray, A. (2016). The 10 skills you need to thrive in the fourth industrial revolution. *World Economic Forum*. Retrieved June 23, 2021, from <https://www.weforum.org/agenda/2016/01/the-10-skills-you-need-to-thrive-in-the-fourth-industrial-revolution/>
- Green, G., & Kaufman, J. C. (Eds.). (2015). *Video games and creativity*. Cambridge, Massachusetts: Academic Press.
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454. doi:10.1037/h0063487
- Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53(4), 267. doi:10.1037/h0040755
- Hamlen, K. R. (2009). Relationships between computer and video game play and creativity among upper elementary school students. *Journal of Educational Computing Research*, 40(1), 1–21. doi:10.2190/EC.40.1.a
- Hamlen, K. R. (2013). Trends in children’s video game play: Practical but not creative thinking. *Journal of Educational Computing Research*, 49(3), 277–291. doi:10.2190/EC.49.3.a
- Hardwick, J. (2020, January 1). *Top YouTube Searches* (2021). SEO Blog by Ahrefs. <https://ahrefs.com/blog/top-youtube-searches/>
- Humke, C., & Schaefer, C. E. (1996). Sense of humor and creativity. *Perceptual and Motor Skills*, 82(2), 544–546. doi:10.2466/pms.1996.82.2.544
- Inchamnan, W., Wyeth, P., & Johnson, D. (2013). Does activity in computer game play have an impact on creative behaviour? *2013 IEEE International Games Innovation Conference (IGIC)*. Vancouver, BC, Canada, 77–84. doi:10.1109/IGIC.2013.6659169
- Jackson, P. W., & Messick, S. (1965). The person, the product, and the response: Conceptual problems in the assessment of creativity. *Journal of Personality*, 33(3), 309–329. doi:10.1111/j.1467-6494.1965.tb01389.x
- Jackson, L. A., Witt, E. A., Games, A. I., Fitzgerald, H. E., von Eye, A., & Zhao, Y. (2012). Information technology use and creativity: Findings from the children and technology project. *Computers in Human Behavior*, 28(2), 370–376. doi:10.1016/j.chb.2011.10.006
- Jivani, A. G. (2011). A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl*, 2(6), 1930–1938.
- Karsenti, T., & Bugmann, J. (2017). Exploring the educational potential of Minecraft: The case of 118 elementary-school students. *International Association for Development of the Information Society*, 175–179.
- Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four-c model of creativity. *Review of General Psychology*, 13(1), 1–12. doi:10.1037/a0013688
- Kharkhurin, A. V. (2014). Creativity.4in1: Four-criterion construct of creativity. *Creativity Research Journal*, 26(3), 338–352. doi:10.1080/10400419.2014.929424
- Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4) Article 4, 150. doi:10.3390/info10040150
- Krathwohl, D. R. (2002). A revision of Bloom’s taxonomy: An overview. *Theory Into Practice*, 41(4), 212–218. doi:10.1207/s15430421tip4104_2
- Kwon, O. N., Park, J. H., & Park, J. S. (2006). Cultivating divergent thinking in mathematics through an open-ended approach. *Asia Pacific Education Review*, 7(1), 51–61. doi:10.1007/BF03036784
- Moffat, D., & Shabalina, O. (2017). Some immediate effects of games on players’ creativity. *Proceedings of the European Conference on Games-Based Learning*. Oct. 2017, Graz, Austria. 442–447. <https://login.proxy.lib.fsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=126280775&site=eds-live&scope=site>
- Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2016). Ha ha? Assessing individual differences in humor production ability. *Psychology of Aesthetics, Creativity, and the Arts*, 11(2), 231. doi:10.1037/aca0000086
- Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 48–73). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511763205.005
- Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, 14(2), 154–166. doi:10.1037/a0019440
- Rafner J., Biskjær M. M., Zana B., Langsford S., Bergenholtz C., Rahimi S., Carugati A., Noy L. and Sherson J. (2022). Digital Games for Creativity Assessment: Strengths, Weaknesses and Opportunities. *Creativity Research Journal*, 34(1), 28–54. doi:10.1080/10400419.2021.1971447
- Rahimi, S. (2020) *Inspire, Instruct, or Both? Game-based and support of creativity*. [Doctoral Dissertation]. Florida State University.
- Rahimi, S., Shute, V. J. (2021a). The effects of video games on creativity: A systematic review. In Russ, S. W., Hoffmann, J. D., Kaufman, J. C. (Eds.), *Handbook of lifespan development of creativity*. (pp. 368–391). Cambridge: Cambridge University Press. doi:10.1017/9781108755726.021.

- Rahimi, S., Shute, V. J. (2021b). First inspire, then instruct to improve students' creativity. *Computers & Education*, 174, 1–27. doi:10.1016/j.compedu.2021.104312.
- Rahimi, S., Shute, V. J., Kuba, R., Dai, C.-P., Yang, X., Smith, G., Alonso Fernández, C. (2021) The use and effects of incentive systems on learning and performance in educational games. *Computers & Education*, 165, 1–17. doi:10.1016/j.compedu.2021.104135.
- Ravitch, S. M., & Carl, N. M. (2019). *Qualitative research: Bridging the conceptual, theoretical, and methodological*. India: SAGE Publications.
- Resnick, M. (2014). Give P's a chance: Projects, peers, passion, play. In G. Kynigos & C. Futschek (Eds.), *Proceedings of the 3rd international constructionism conference 2014*. Vienna, Austria. (pp. 13–20). Austrian Computer Society.
- Resnick, M. (2018). *Lifelong kindergarten: Cultivating creativity through projects, passion, peers, and play* (Reprint ed.). Cambridge, Massachusetts: MIT Press.
- Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, 42(7), 305–310. JSTOR
- Richards, R. (2010). Everyday creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 189–215). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511763205.013
- Rifkin, R., & Klautau, A. (n.d.). In defense of one-vs-all classification. *The Journal of Machine Learning Research*, 5, 101–141.
- RStudio Team. (2016). *RStudio: Integrated Development Environment for R* (1.0.136). RStudio, Inc. Retrieved July 25 2021. <http://www.rstudio.com/>
- Ruch, W., & Heintz, S. (2019). Humor production and creativity: Overview and recommendations. In S. R. Luria, J. Baer, & J. C. Kaufman (Eds.), *Creativity and Humor* (pp. 1–42). Cambridge, Massachusetts: Academic Press. doi:10.1016/B978-0-12-813802-1.00001-6
- Runco, M. A. (2003). Education for creative potential. *Scandinavian Journal of Educational Research*, 47(3), 317–324. doi:10.1080/00313830308598
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, 24(1), 66–75. doi:10.1080/10400419.2012.652929
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, 24(1), 92–96. doi:10.1080/10400419.2012.650092
- Shaw, A. (2022). Creative minecrafters: Cognitive and personality determinants of creativity, novelty, and usefulness in minecraft. *Psychology of Aesthetics, Creativity, and the Arts*. doi:10.1037/aca0000456
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, North Carolina: Information Age Publishers.
- Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: foundations, innovations, and perspectives* (pp. 43–58). New York, NY: Springer. doi:10.1007/978-1-4614-3546-4_4
- Shute V. J. & Rahimi S. (2021). Stealth assessment of creativity in a physics video game. *Computers in Human Behavior*, 116. 10.1016/j.chb.2020.106647
- Shute, V. J., & Wang, L. (2016). Assessing and supporting hard-to-measure constructs in video games. In A. A. Rupp & J. P. Leighton (Eds.), *The handbook of cognition and assessment* (pp. 535–562). Hoboken, New Jersey: John Wiley & Sons, Inc. doi:10.1002/9781118956588.ch22
- Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist*, 51(7), 677. doi:10.1037/0003-066X.51.7.677
- Tan, X., Letendre, J. H., Collins, J. J., & Wong, W. W. (2021). Synthetic biology in the clinic: Engineering vaccines, diagnostics, and therapeutics. *Cell*, 184(4), 881–898. doi:10.1016/j.cell.2021.01.017
- Torrance, E. P. (1974). *Torrance tests of creative thinking: Norms-technical manual*. Bensenville, Illinois: Scholastic Testing Service.
- Toubia, O., Berger, J., & Eliashberg, J. (2021). How quantifying the shape of stories predicts their success. *Proceedings of the National Academy of Sciences*, 118(26), e2011695118. doi:10.1073/pnas.2011695118
- Treadwell, Y. (1970). Humor and creativity. *Psychological Reports*, 26(1), 55–58. doi:10.2466/pr0.1970.26.1.55
- Walker, J., Slater, S., & Kafai, Y. (2019). A scaled analysis of how Minecraft gamers leverage YouTube comment boxes to participate and collaborate. In (Eds.), *A wide lens: combining embodied, enactive, extended, and embedded learning in collaborative settings* (Vol. 1, pp. 440–447). International Society of the Learning Sciences. eds. In K. Lund, G. P. Nicolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, M. Baker. <https://repository.isls.org/handle/1/1601>
- Ward, T. B., & Sonneborn, M. S. (2011). Creative expression in virtual worlds: Imitation, imagination, and individualized collaboration. *Psychology of Popular Media Culture*, 1(S), 32. doi:10.1037/2160-4134.1.S.32
- Webster, J. J., & Kit, C. (1992). Tokenization as the initial phase in NLP. *COLING 1992 Volume 4: The 14th International Conference on Computational Linguistics*. Nantes, France. vol 1992. COLING. <https://aclanthology.org/C92-4173>
- Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin*, 50(5), 362. doi:10.1037/h0060857
- Wright, T., Boria, E., & Breidenbach, P. (2002). Game studies—Creative player actions in FPS online video games: Playing counter-strike. *The International Journal of Computer Game Research*, 2(2). <http://www.gamestudies.org/0202/wright/>
- Zedelius, C. M., Mills, C., & Schooler, J. W. (2019). Beyond subjective judgments: Predicting evaluations of creative writing from computational linguistic features. *Behavior Research Methods*, 51(2), 879–894. doi:10.3758/s13428-018-1137-1

Appendix A: R code to extract video captions

```
## Install packages
## Author: Seyedahmad Rahimi
install.packages("youtubecaption")
install.packages("googlesheets4")

## Call package
library(youtubecaption)
library(googlesheets4)
library(tidyverse)

#to bypass the authentication
gs4_deauth()

#to read the Google sheet
#this address only contains the all-together links
videos <- read_sheet("CREATE A GOOGLE SPREADSHEET WITH YOUTUBE LINKS STORED IN ONE VARIABLE")

## Get caption
#to create a dataframe called captions
captions <- data.frame()

#to go through the videos, get the URLs, and extract the captions
for(i in videos$`YouTube Link`){
  captions <- rbind(captions, get_caption(i))
}

#to write out the final scv file
write.csv(captions, file = "captions.csv")
```

Appendix B: Python code for the NLP

```

9  import pandas as pd
10 import numpy as np
11 import nltk
12 from collections import Counter
13 #nltk.download('punkt')
14 #nltk.download('wordnet')
15 nltk.download('omw-1.4')
16 def data_import(data='Book2'):
17     dat = pd.read_csv(data+'.csv')
18     dat = dat[dat.vid.apply(lambda x: len(str(x))>4)]
19
20     def clean(string):
21         string = str(string)
22         string = string.replace('- [Narrator]', '')
23         string = string.replace('\n', ' ')
24         string = string.replace('-', ' ')
25         string = string.replace("don't", "do not")
26         string = string.replace("can't", "cannot")
27         string = string.replace("'ll", " will")
28         string = string.replace("she's", "she is")
29         string = string.replace("he's", "he is")
30         string = string.replace("'re", " are")
31         string = string.replace("it's", "it is")
32         string = string.replace("I'm", "I am")
33         string = string.strip()
34         return string
35
36     dat['text'] = dat.text.apply(clean)
37     #dat['start_sent'] = dat.text.apply(nltk.word_tokenize).apply(lambda x: x[-1] in [',', '?'])
38     #dat['start_sent'] = dat.start_sent.replace({False:'', True:'//end/'})
39     #dat['text'] = dat.text + ' ' + dat.start_sent
40
41     #df = pd.DataFrame(
42     #    ''.join(dat.text.values)
43     #    .split('//end/'), columns=['text']
44     # )
45
46     #df_short = dat[dat.start_sent == '//end/']
47     #df_short['sentence'] = df.text[:-1].tolist()
48
49     df_short = dat
50     df_short[['vid',
51               'Elaboration', 'Originality', 'Aesthetics ',
52               'Surprise', 'Humor', 'Complexity',
53               'Novel use of materials', 'Realism/ recreation',
54               'text']]
55     return df_short
56
57
58 df_short = data_import(data='Book1')
59 df_test = data_import(data='Book2')

```