



Article

Examining Computer-Supported 3D Event Recreation for Enhancing Cognitive Load, Memorability, and Engagement

Ruochen Cao ^{1,*}, James Walsh ¹, Andrew Cunningham ¹, Mark Kohler ², Ross T. Smith ¹ and Bruce H. Thomas ¹

¹ IVE: Australian Research Centre for Interactive and Virtual Environments, University of South Australia, Adelaide, SA 5095, Australia; james.walsh@unisa.edu.au (J.W.); andrew.cunningham@unisa.edu.au (A.C.); ross.t.smith@unisa.edu.au (R.T.S.); bruce.thomas@unisa.edu.au (B.H.T.)

² School of Psychology, Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, SA 5008, Australia; mark.kohler@adelaide.edu.au

* Correspondence: ruochen.cao@mymail.unisa.edu.au

Received: 2 May 2020; Accepted: 29 June 2020; Published: 6 July 2020



Abstract: Three-dimensional rendering technologies have long been utilized for explanatory purposes in scientific visualization and related areas. Their applications to wider fields, however, have often been limited. In this paper, we explore the use of 3D model and animation techniques, combined with narrative techniques, for recreating event-based information to aid understanding. An empirical experiment was conducted which examined the effectiveness of 3D model images and 3D animation videos compared to reading narratives in textual form. The results indicated that both forms of 3D graphical techniques positively supported users in terms of cognitive load, recall, and engagement over reading text.

Keywords: 3D model; 3D animation; narrative; cognition

1. Introduction

While textual description is still the most common way to communicate facts (as narrative) in law enforcement, static text documents have limitations that may hinder the efficiency of people that must rely on those methods to understand court cases. For example, text alone cannot effectively recreate *how* a case with many spatial and temporal elements unfolded. In a law enforcement and court scenario, the general public may be required to consume these cases as a member of jury in a court of law. Other methods may significantly improve recall and engagement in such scenarios, factors that are crucial to reaching a fair verdict. In this paper, we examine the use of computer-generated 3D content videos for representing narratives of many cases. This work is part of a larger project with the Data to Decisions Cooperative Research Centre (<https://www.d2dcrc.com.au/>) and the Australian Federal Police that aims to explore the impact of introducing new technologies for law enforcement and justice. We are motivated by discussions with federal police officers and state and federal prosecutors.

There is evidence to suggest that videos using 3D rendered models and animation techniques can improve the understanding of particular types of narratives when compared to textual prose alone. For example, it has been shown that “multimedia” information (containing both visual and audio narration) can be easier to process than text and static images alone [1–3]. Furthermore, 3D models and animations can help to convey specific types of narratives that rely on spatial and temporal information by visualizing the passing of time and relative positions. The use of 3D techniques enables the visualization of phenomena that are typically invisible in the real-world (e.g., cell cultures [4]) or past events that are difficult to recreate in real-world (e.g., a flight crash [5]). Consequently,

3D techniques have been applied in recreating real-world events and narratives (examples of which can be found in [6,7]). However, there are few empirical studies examining the benefits that 3D models and animations can provide over text or static images in recreating and communicating court cases or similar real-world narratives. Furthermore, 3D techniques have the potential to introduce issues of bias and information inaccuracy [8] that will affect an audience's understanding, recall, and attitudes [9]. This research gap and related issues suggest that a more clear understanding is needed for applying 3D techniques to recreate real-world narratives.

In this work, we examine the effects of using 3D models and animations on cognitive load, recall, and engagement when conveying case narratives. Three modalities were compared: textual prose, static images of 3D models, and 3D animation videos. As the baseline modality, the textual presentation uses only prose to describe the narrative. The static images 3D models replicates a "Powerpoint" presentation, featuring static images of 3D model renders. The 3D animation video condition is for examining the value of applying 3D models with animation for telling narratives. The contribution of this work is to provide the first user study examining the effect of 3D models and animations for presenting court cases, demonstrating their potential for benefiting users' recall, cognitive load, and engagement. It should be mentioned that the technologies and benefits studied in our work are not limited to law enforcement, and we envision their application to domains that involve event-based narratives such as medical cases, briefings for defence, and public communication of major news events.

Following the review of related work, we present our experiment in detail, including its purpose, conditions, hypothesis, design, process, and results. We also provide our insights regarding the results. After an overview of the limitations of this study, we conclude the paper and discuss future work.

2. Related Work

Three-dimensional graphics such as 3D animations have been postulated to be effective in delivering information (e.g., [10,11]); however, some researchers have raised concerns about the possible issues of using these 3D techniques, such as the possibility of distorting the information [12] and an effect of extraneous cognitive load [13]. These concerns suggest that there would be value in user studies to examine those technologies and provide effective guidance for their applications. The design of such studies, however, is an open question. In our work, we focused on three main metrics: cognitive load, memorability (recall), and engagement. In this section we introduce those metrics in detail. Examining several modalities for conveying narratives requires multiple different narratives to be delivered while ensure that the narratives are equivalent in regards to complexity. As such, we also provide an overview a number of linguistic complexity metrics that can be used to compare narratives.

2.1. Cognitive Load, Memorability, and Engagement

Cognitive Load Theory (CLT) was initially explored in the early 1980s. Cognitive load has been described as containing three types [14]: intrinsic (introduced by the inherent difficulty of the learning contents), extraneous (affected by the way that the learning tasks are organized and presented), and germane load (dedicated to the processing, construction, and automation of schemas). While each type of cognitive load focuses on a unique affecting aspect of learning process and would be beneficial to be measured separately, the majority of the current methods do not allow distinguishing the specific types of measured cognitive load. Therefore, it remains important to analyze the results in the context of the associated learning performance [15]. Cognitive load measurement techniques have been classified into three categories: subjective (self-report) measures [16,17], performance-based measures [18,19], and physiological measures [15,20]. Each method has its advantages and disadvantages, and therefore should be utilized based on the specific use-case. While subjective measures have been highly applied and proved to be reliable and easy to use [21,22], they are sensitive to a user's self-description of their perceived mental burden by providing a numerical value. Performance-based techniques normally

include a dual set of tasks: a primary task and a secondary task. In this case, the performance of the secondary task works by reflecting the cognitive load level of the primary task. Because of the interference between tasks, the use of performance-based techniques needs careful design. Physiological methods allow researchers to look at data in a continuous manner and describe the relations between cognitive load and learning process at a specific instance of time instead of an overall measurement [15]. However, the measures of physiological data are normally very intrusive and cumbersome to deploy.

Memorability, defined as “a capability of maintaining and retrieving information” [23], has been suggested to be important for viewing visualizations, as a major purpose of communicating information is to enable users to remember and discuss the relevant knowledge at a later time [24]. There have been many experiments measuring the memorability in visualization fields [25,26]. Popular measurements include *free recall* and *cued recall*, both of which require participants to learn an item or a list of items first and describe the item(s) either freely or with the help of cues. One method is to simply ask participants questions regarding what they learned, such as “How many items have you seen?”, and measure their recall based on the accuracy of the answers [27]. Considering the complexity of real-world narratives and the purpose of communicating information in real life, we believe reciting the whole narrative is too difficult and not necessary. Instead, remembering some key information such as time, location, and involved people is essential. As such, we examine another method named Signal Detection Theory (SDT) [28] to approach recall. In an SDT model, participants only need to indicate if the provided information they currently see was previously presented before. In this case, key information is provided, such as the time when the case happened, the important events, and people who are involved. The value of d' is used as an indicator of memorability, which is calculated by the value of four types of response: *Hit*, *False Alarm (FA)*, *Correct Rejection*, and *Miss*. *Hit* refers to a Yes response given to old information (facts that were seen before); *FA* refers to a Yes response to new information (facts that were not seen before); *Correct Rejection* indicates a No response to new information; *Miss* is a No response given to old information.

User engagement is used for describing the level of being captivated by technologies and the motivation to use it. The engagement level is indicated to affect a user’s experience of viewing visualizations. For example, users may investigate and more closely examine a visualization when they are engaged [29]. A high-level engagement can also cause users to underestimate the passage of time [30]. There are many categories that have been classified for indicating people’s engagement level from various aspects [31–34], such as cognitive involvement, affective involvement, and focused attention. The most common measuring method is a subjective measurement [34]. By using a series of statements such as “I consider this task a success” and Likert scales, researchers can gain an understanding of the participants’ engagement level.

2.2. Linguistic Complexity Metrics

Linguistic complexity is a significant topic in linguistic research [35], where *complexity* has three main meanings:

“1. *Structural complexity*, a formal property of texts and linguistic systems having to do with the number of their elements and their relational patterns; 2. *Cognitive complexity*, having to do with the processing costs associated with linguistic structures; 3. *Developmental complexity*, i.e., the order in which linguistic structures emerge and are mastered in second (and, possibly, first) language acquisition.” [36]

As illustrated, *cognitive complexity* focuses on the investigation of people’s language processing progress, while *structural complexity* focuses on the complexity of a whole linguistic system, and *developmental complexity* is for assessing people’s language learning progress. Therefore, we only focus on *cognitive complexity* in our work. There are numerous complexity metrics that have been developed (e.g., Mean Clause Per Utterance; MCU [37], Developmental Sentence Score; DSS [38],

The Index of Productive Syntax; IPSyn [39]). However, some have proven more successful than others in predicting cognitive performance, with popular metrics in current literature including Mean Length of Utterance (MLU) [40,41] and Idea Density (ID) [42,43]. MLU focuses on the average number of words in a sentence in a piece of text and ID calculated the average number of ideas (i.e., the relations expressed by verbs, adjectives, adverbs, prepositions, and conjunctions) packed into the overall text. As we believe the information density is more important for illustrating the complexity of narratives, in this work, we applied the ID method.

The following is an example of how the ID is calculated for the sentence: “There was a flashing red light under the back of her bicycle seat.” There were seven *ideas*: 1. Was (v), a flashing red light; 2. Flashing (adj), light; 3. Red (adj), light; 4. Under (prep), her bicycle seat; 5. back (adv), her bicycle seat 6. Of (prep), back, her bicycle seat. 7. her (possessive pronoun), bicycle seat. This sentence has seven ideas across 13 words. As such its ID score is 7 ideas/13 words = 0.54.

3. Experiment

In order to test the effectiveness of 3D models and animations, we conducted a study comparing text, still images of 3D models (shown in videos applying slides transition effect), and 3D animation videos. We collected three narratives with similar complexity (based on their ID, see Section 2.2) and created 3D computer graphics animations to represent those narratives. The study was a within-subjects experimental design, with each participant experiencing all three techniques with different narratives. The same level of complexity ensured that different narratives did not affect the comparison of the three techniques for each participant. The order of narratives and techniques between participants was counterbalanced to counter effects of learning and fatigue. We measured three metrics: (1) cognitive load, (2) recall, and (3) engagement. In addition to these metrics, we recorded the participant’s preferences for exploring purpose. The measure of user preference is a standard measure in HCI research. In this section, we will describe the experiment in detail.

3.1. Conditions

There are three presentation conditions in this experiment for the consumption of the narratives: (1) text condition, (2) 3D model images condition, and (3) 3D animation videos condition. The difference between conditions only involved method by which the narrative was conveyed. The narrative aspects, such as narrative structures and narrations, were the same for each condition.

Text condition: In the text condition, each participant consumed the text-only version of the narrative. The text was taken from the summary of the court cases we collected (as described in Section 3.3.2) and printed on A4 sheets of paper.

3D model images condition: In the 3D model images condition, participants consumed narratives by watching a slideshow of still images of 3D models with a voiceover reading the text. The 3D model images were created by taking keyframes from a 3D animation video. These keyframes were made into a slide show video by applying a slide transition effect. *The selection of keyframes will be introduced in Section 3.3.2.* To ensure consistency, we applied the same slide show effect and audio track (music and voice) in the 3D animations videos for presenting images. Each video in this condition was shown for approximately same time length as the equivalent 3D animation version/scene (within a few seconds). Some example images are in Figure 1a. In order to simplify terminology for the remainder of this paper, we will call the videos presenting 3D model images “3D model image videos”.

3D animation videos condition: In 3D animation video condition, participants watched 3D graphics and animation videos of the cases with a voice over reading the text. The 3D animations were created based on the text version. All the information contained in the text version of the cases was presented in videos by animation, voice-over, or annotations. Some example keyframes from the video are depicted in Figure 1b.

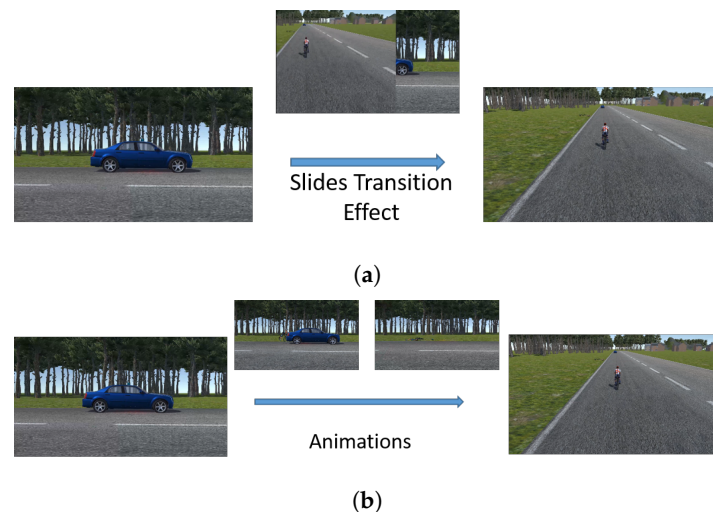


Figure 1. Example of two of the conditions; (a) 3D model images with slides transition effect, and (b) 3D animation videos.

3.2. Hypotheses

The core contribution of this work was to test whether 3D techniques introduce benefits for conveying narratives. We propose 3D model image videos and 3D animation videos can benefit participants' recall, cognitive load, and engagement.

Hypothesis 1 (H1). *The use of 3D model images can reduce the cognitive load when participants consume narratives as compared to text.*

Hypothesis 1a (H1a). *The use of 3D animations can reduce the cognitive load when participants consume narratives as compared to static 3D model images.*

Hypothesis 2 (H2). *The use of 3D images can improve a participant's recall of narratives as compared to text.*

Hypothesis 2a (H2a). *The use of 3D animations can improve a participant's recall of narratives when compared to 3D model images.*

Hypothesis 3 (H3). *The use of 3D images can keep participants more engaged when compared to text.*

Hypothesis 3a (H3a). *The use of 3D animations allow participants to become more engaged when compared to 3D model images.*

3.3. 3D Content Creation

In this section, we will introduce the process that we followed for creating the 3D content in the user study, including the problems we faced and the important decisions we made. We do not claim that our solutions are the only workable methods for creating 3D visual content for conveying narratives; however, it does provide a usable framework.

3.3.1. Experimental Materials Selection

Three narratives were constructed from the summaries of real-world court cases from the website of Australasian Legal Information Institute (<http://www8.austlii.edu.au>). All three narratives have a similar complexity rating (ID score). Table 1 presents the information of experimental materials, including ID scores, length of narratives as textual prose, and the duration of the resulting 3D animation videos and 3D model static image videos.

The first narrative involved one count of attempted murder [44]. The second narrative was constructed from a murder case [45]. The third narrative was from a case about a car accident causing death [46]. An example for each narrative is as follows:

Narrative one: Walsh attempted to get away from the situation by overtaking the appellant and accelerating at speeds in excess of 150 km per hour. The appellant pursued Walsh and then overtook him a second time. As Walsh again trying to overtake the appellant in order to get away from him, he saw the appellant hanging out of the driver's side window pointing a gun at him.

Narrative two: The appellant then went to the kitchen and opened a drawer to obtain a knife. He returned to the bathroom with the knife, cornered the victim and made several stabbing motions towards him. He then grabbed the victim by his T-shirt, pulled the victim towards himself and then stabbed him once to the chest.

Narrative three: On a straight, level section of the road the appellant's vehicle struck the back tyre of Ms Heraghty's bicycle. The left side of the front bumper bar of the vehicle being driven by the appellant struck the rear wheel of the deceased's bike causing her to be thrown back over the left side of the car's bonnet, hitting and shattering the left side of the windscreen, with her helmet hitting the roof of the car. She died at the scene of the collision.

Table 1. Experimental Materials Information.

Forms	Narratives	Narrative 1	Narrative 2	Narrative 3
ID Score		ID = 0.46 ^a	ID = 0.45 ^a	ID = 0.45 ^a
Text		337 words	580 words	306 words
3D Model Image Videos		192 s	241 s	202 s
3D Animation Videos		185 s	274 s	202 s

^a ID scores were calculated by Computerized Propositional Idea Density Rater (CPIDR), Version 5.1.4637.21009 compiled 11/09/2012.

3.3.2. Design Decisions

To ensure reproducibility and robustness, we determined on following a structured design approach to create 3D animations from textual descriptions of court cases. The following describes the design decision in logical ordering:

Narrative Structures Creation: Generally, the aim when presenting a narrative is to convey the narrative in a logical way. To achieve this, there is a history of narrative methods and patterns to follow. An early decision when presenting a narrative is to determine a suitable structure to convey that narrative. Through discussions with a narrative expert, we determined to apply the classical *Three-Act Structure* [47,48] model for all the narratives. Three-Act Structure divides a story into three acts: (1) Setup (exposition of the story's background including characters, relationships, and so forth), (2) Confrontation (where the main part of the narrative occurs), and (3) Resolution (final confrontation, denouement, and normally climax occurs.) We chose the *Three-Act Structure* for two reasons: (1) it has been well developed in many mediums such as novels, video games, and movies, providing a strong background knowledge support; and (2) fictional crime storytelling has extensively utilized the Three-Act Structure, with works such as *The Maltese Falcon* (John Huston, 1941), *Scarface* (Brian De Palma, 1983), and *The Departed* (Martin Scorsese, 2006). The re-constructed text scripts were applied in the text condition as learning materials and the two 3D conditions as scripts for creating 3D content. As such, the narrative structure was the same between conditions. It should be noted that

the Three-Act Structure is not the only workable structure for illustrating court cases, with the ideal structure(s) varying between cases.

Events Recreation Much like creating a cinematic movie, the recreation of court cases involves creating characters, actions, and environments. The creation of these objects should be strictly based on their text descriptions. However, the summaries of court cases do not always present all the details for recreation. While visualizing the provided information, such as the gender of characters and location of the case, we aimed to not present any extraneous information that was not included in the text description (such as facial expression). Characters models were made using Autodesk Character Generator (<https://charactergenerator.autodesk.com/>). Human animations were captured using an OptiTrack Motion Capture system (<https://tracklab.com.au/>). The environment recreation was completed in Unity 2017.3.1f1 (<https://unity3d.com/>). Additionally, for consistency, a voice-over was made directly from text description using an online text to speech (TTS) service (<http://www.fromtexttospeech.com/>), and the final videos were edited by Adobe Premiere Pro CC 12.1.2 (<https://www.adobe.com/au/>).

Viewpoint and 3D Model Images Selection The creation of our 3D animations were performed under instruction from a narrative expert with experience in cinema. After obtaining the court cases, the narrative expert provided a storyboard with drawn keyframes and introductions for the first animation video, and instructional suggestions for the rest two animation videos. Some examples of drawn storyboards are shown in Figure 2. We then created the animations based on the storyboard. At the same time, keyframes of drawn storyboard were captured for the 3D model images condition. More keyframes introducing key information of the cases such as characters, time, and location were added in the 3D model image videos.

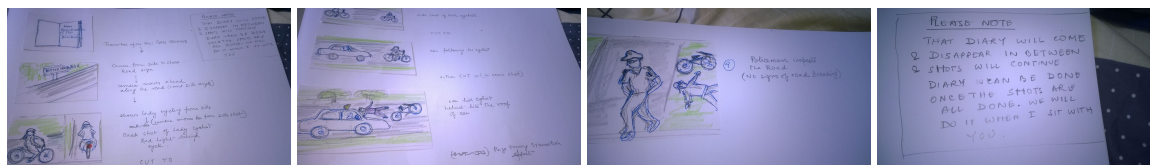


Figure 2. Example of a storyboard developed for the user study, created by a professional storyboarder. Each scene is illustrated with accompanying notes.

3.4. Experiment Design

We applied a within-subjects experimental design, where each participant experienced all three conditions with different narratives. After each of the conditions, participants were immediately given questionnaires about cognitive load, recall, and engagement. After all three conditions, participants were asked to rank the three narrative methods based on their experience. The experiment for each participant was approximately 50 min; however, the actual length depended on the time that each unique participant spent on consuming the narratives and finishing the questionnaire. The narratives' content and formats are two independent variables. The five dependent variables of the experiment were: (1) cognitive load score, (2) recall score, (3) engagement score, (4) preference of presentation for consuming narratives.

3.4.1. Participants

We recruited 36 participants in our experiment. All subjects gave their consent for inclusion before they participated in the study. The project was approved by the University of South Australia's Human Research Ethics Committee (protocol number: 201172). Participants were aged from 19 to 57 years old ($M = 29.82$, $SD = 8.84$) with 24 males and 12 females. There were 22 students, 11 university personnel, and 3 members of the general public with various professional and academic backgrounds. As the narratives were collected from the real court cases, we screened out potential participants by asking the following question: "Have you ever had any involvement in a court cases? (e.g., charged or convicted, participated in jury duty, you work as a lawyer/judge, you work as a police officer)".

This excluded participants with possible previous knowledge of the cases that we applied. We also screened out participants with English as a second language to ensure they were capable of correctly comprehending the court cases. Participants were compensated with a \$20 AUD gift card for their time.

3.4.2. Experiment Apparatus

For the text condition in the experiment, three narratives were printed on A4 sheets of paper, with 14 pt Calibri font. A computer with an Intel Core i7-7700, nVidia 1080 TI Graphics Card, and 64GB RAM with a 24-inch monitor with a 1920×1080 resolution was employed for showing 3D animation videos (*3D animation videos condition*) and 3D model image videos (*3D model images condition*). In order to avoid the influence of environment, the same empty room was employed for all participants throughout all conditions: there was only one participant and one researcher in the room for the duration of the experiment.

3.4.3. Measurements

As discussed in Section 2.1, numerous cognitive load measures are potentially applicable in our study. Based on the following considerations, we used subjective measures in our experiment: (1) Our study involved the test of recall and engagement, which is sensitive to distractions and interference possibly introduced by a secondary task; (2) Applying physiological methods requires complex setups which was considered as not necessary for this initial exploration. Paas [17] is one of the most utilized subjective methods for measuring overall cognitive load (also referred as mental effort), the level of which reflects how hard an individual tries during learning processes. It should be mentioned that, while a few subjective surveys of cognitive load measuring the three distinct load types have been presented recently (e.g., [49]), as an initial exploration, we designed our study for only investigating the difference of overall cognitive load between different presentation techniques. Therefore we still used the Paas evaluation that has widely been proven to be reliable [21,22].

For the measurement of recall, we used SDT (as described in Section 2.1). Fifteen pieces of existing “old” information were selected from each narrative, and fifteen pieces of fabricated new information were created. We selected old information from our narratives, including all the key information such as time, location, and knowledge of victims and appellants. While it is unavoidable that 3D animations include more dynamic information such as the manner of motion, all the old information was selected for equivalence in all three conditions; no information regarding motion was selected. New information was selected from other similar narratives.

The level of engagement affects viewers in many ways. In our work, we selected to measure two dimensions of engagement—*Focused Attention* and *Perceived Usability*—as we believe these two dimensions could potentially affect a visualization’s efficiency when it is used in information communication. *Focused Attention* is for measuring the level that viewers could concentrate their attention on the narratives. *Perceived Usability* measures whether viewers feel the presented narratives are meaningful. We measured perceived usability through a negative attribute by measuring viewers troubles of following the narrative, where questions such as *I could not get the information that I need to find in this narrative* and *I found this narrative confusing to understand* were applied.

3.4.4. Task and Procedure

The experiment had the participants perform three tasks, consuming three narratives each in one of the three display conditions. Each task was performed in two phases: Narrative Consumption and Questionnaire phase.

Narrative Consumption phase: At the beginning of a task, participants experienced the consumption phase, where they were asked to consume one of the narratives. A printed or video scenario was provided depending on the condition. Participants were encouraged to take as much time as they required to finish consuming the cases to fully understand the information in narratives. For the text condition, participants were free to re-read the text. For the 3D model image and 3D animation videos

conditions, participants were allowed to pause, fast forward, rewind, and replay the videos. All their operations were recorded for future analysis. When the participants felt they understood the cases and were ready for the questionnaire, they told the experimenter to start the next phase.

Questionnaire phase: After indicating that they believed they understood the cases, participants were then provided with questionnaire testing their cognitive load, recall, and engagement level regarding the cases that they consumed in the last phase. Participants were also encouraged to provide any comments that they had on the condition. After finishing the questionnaire phase, participants then immediately started the next task.

3.5. Results

For the data analysis in this work, we used IBM SPSS (<https://www.ibm.com/au-en/analytics/spss-statistics-software>). The interpretation of the results followed the work of Andy Field [50]. We analyzed the collected data in terms of both the particular narrative and presentation format. In order to validate that the different narratives did not affect the results, we compared the data between narratives. In order to test whether static images of 3D models and 3D animations can improve the effectiveness of presenting narratives, we compared the collected data between different conditions.

3.5.1. Cognitive Load

As the Paas scores were not normally distributed, we applied Friedman's ANOVA on Paas scores. The results showed that the Paas scores did not significantly differ between different narratives, $\chi^2(2) = 3.47$, $p = 0.18$. This demonstrates that the different narrative stories did not have a significant influence on cognitive load in our experiment.

The same Friedman's ANOVA was applied to assess cognitive load across presentation conditions. The results showed that there was significant difference between conditions, $\chi^2(2) = 16.14$, $p < 0.001$. Wilcoxon post hoc tests were then employed to further analyze the data. The Paas scores did not significantly differ between text condition and 3D model images condition, $T = 216.00$, $r = -0.14$, $p = 0.25$. The 3D animation videos significantly decreased participants' cognitive load compared to text, $T = 77.50$, $r = -0.41$, $p < 0.001$, and 3D model images, $T = 56.50$, $r = -0.34$, $p = 0.004$. As such, H1 was not supported, but H1a was supported. The means of the Paas scores for the different conditions are shown in Figure 3.

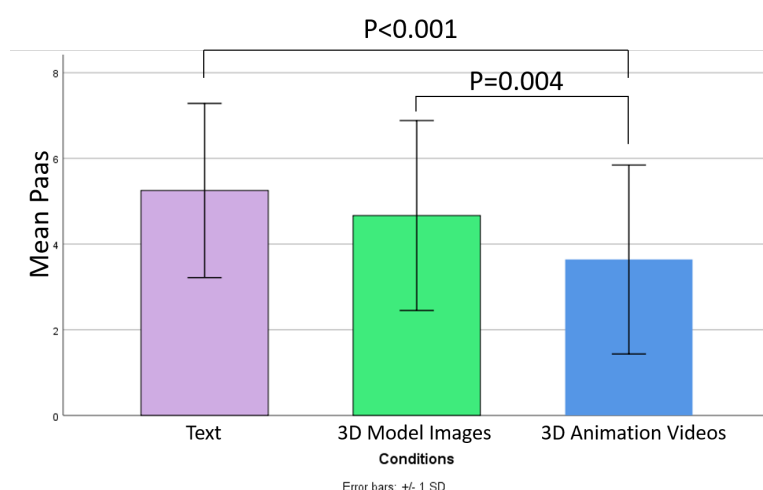


Figure 3. Means of Paas scores for each condition: Text condition (**left**), 3D model images condition (**middle**), and 3D animation videos condition (**right**). The 3D animation videos condition was associated with a significantly lower Paas score (indicating lower cognitive load; better performance) as compare to the text and 3D model image condition, however, there was no significant difference between the text condition and 3D model images condition.

3.5.2. Recall

After counting the number of Hit, FA, Correct Rejection, and Miss according to SDT, we then calculated the value of d' for each participant for each condition with the R *psycho* package [51]. As the d' values for all the participants were normally distributed, a repeated-measures ANOVA was applied.

We first assess the data across narratives. Mauchly's test indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 5.29$, $p = 0.07$, therefore sphericity assumed is reported. The results showed that the d' value was not significantly affected by different narratives in our experiment, $F(2.00, 70.00) = 0.725$, $p = 0.49$, $\omega^2 = -0.004$.

We analyzed the data across the presentation conditions. Mauchly's test indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 0.419$, $p = 0.811$, therefore sphericity assumed is reported. The results showed that the recall of the narratives was significantly affected by the presentation methods, $F(2.00, 70.00) = 15.11$, $p < 0.001$, $\omega^2 = 0.13$. Wilcoxon post hoc tests were then used to further analyze the data. The recall scores of text were significantly smaller than that of 3D model images ($p < 0.001$, $r = 0.42$) and 3D animation videos ($p < 0.001$, $r = 0.46$). This demonstrates the use of 3D graphics (both models and animations) improves a user's recall of a narrative, supporting H2. However, there is no significant difference between the recall scores of 3D model images and 3D animation videos ($p = 0.619$), rejecting H2a. The means of d' for each condition are shown in Figure 4.

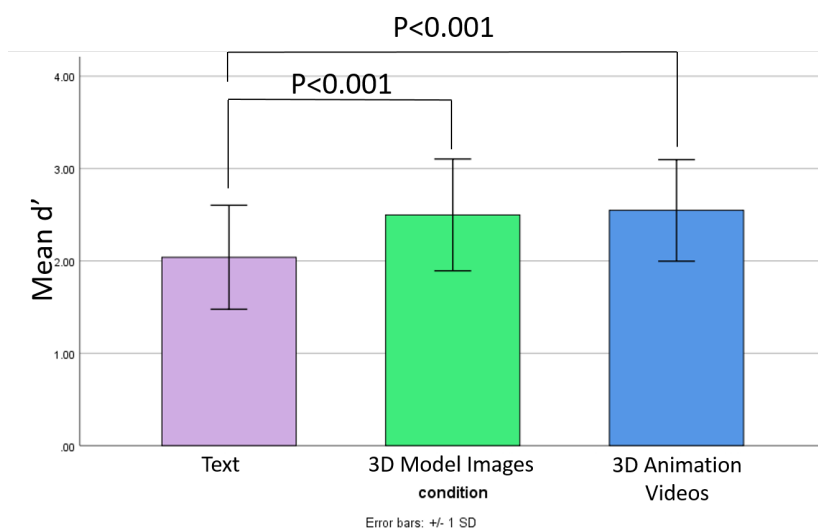


Figure 4. Means of d' (see Sections 2.1 and 3.4.3 for a detailed introduction of d') for each condition: Text condition (left), 3D model images condition (middle), and 3D animation videos condition (right): the 3D model images condition and 3D animation videos condition were associated with a significantly higher d' score (indicating better recall performance) as compared to the text condition, however, there was no significant difference between 3D model images conditions and 3D animation videos condition.

3.5.3. Engagement

As previously introduced, the engagement level was measured by two dimensions: *Focused Attention* and *Perceived Usability*. We explored engagement scores of both dimensions and found that they were not normally distributed. As such, Friedman's ANOVA was applied. We calculated and submitted the overall scores of participants' ratings for each of the two engagement dimensions to Friedman's ANOVA. Significant findings were found in both of our engagement items, both H3 and H3a were supported. Engagement scale means are shown in Figure 5.

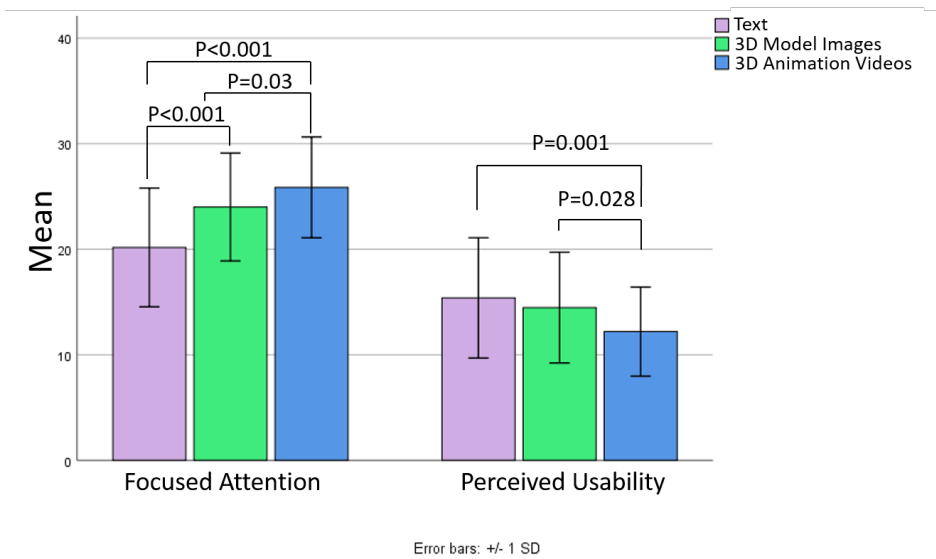


Figure 5. Means of different engagement scale ratings for each condition. In focused attention, the 3D model images condition was associated with a significantly higher score (reflecting better performance) as compared to the text condition. The 3D animation videos condition was associated with a significantly higher score (reflecting better performance) as compared to the text and 3D model images condition. In perceived usability, the 3D animation videos was associated with a significantly lower score (reflecting better performance) as compared to the text and 3D model condition, however, there was no significant difference between the text and 3D model condition.

Focused Attention: Our analysis revealed that different narratives did not significantly affect focused attention in our experiment, $\chi^2(2) = 0.191$, $p = 0.909$. Significant difference between presentation conditions was observed, $\chi^2(2) = 33.279$, $p < 0.001$. Wilcoxon post hoc tests were used to follow up this finding. Participants' ratings of focused attention for 3D model images were significantly higher than that of text, $T = 492.00$, $r = 0.45$, $p < 0.001$. The 3D animation videos significantly improved the focused attention of narratives compared to text, $T = 629.00$, $r = 0.55$, $p < 0.001$, and 3D model images, $T = 362$, $r = 0.26$, $p = 0.03$.

Perceived Usability: As perceived usability was measured through a negative attribute, we reverse coded the rating for this item. Analysis of scores given for this item showed that there was significant difference between narratives, $\chi^2(2) = 7.86$, $p = 0.02$. We further analyzed this data by applying Wilcoxon Post Hoc tests. The results showed that while the perceived usability scores did not significantly differ between narrative one and narrative three ($p = 0.30$), and narrative two and narrative three ($p = 0.16$), there was significant difference between narrative one and two, $T = 505.00$, $r = 0.37$, $p = 0.002$.

The results of analyses also indicated the significant difference between presentation conditions, $\chi^2(2) = 12.477$, $p = 0.002$. According to the results from Wilcoxon post hoc tests, 3D model images did not significantly improve perceived usability of story narratives, $T = 197.50$, $r = -0.08$, $p = 0.47$. However, 3D animation videos had a significantly higher level of perceived usability compare to text, $T = 82.50$, $r = -0.38$, $p = 0.001$, and 3D model images, $T = 169.5$, $r = -0.26$, $p = 0.028$.

3.5.4. Users Preference

For measuring participants' preference of the three methods, we asked them to rate methods using a three-point Likert scale ranging from dislike (1) to like (3). A Wilcoxon Signed-ranks test indicated that participants preferred the 3D animation videos (*Median* = 3), as compared to text (*Median* = 2), $p < 0.001$, and 3D model image videos (*Median* = 2), $p = 0.002$. However, there was no significant difference between text and 3D model image videos ($p = 0.417$).

3.6. Discussion

The purpose of this study is to examine the use of 3D models and animations for representing and communicating law enforcement cases. The results of our experiment demonstrate the value of using 3D models and animations for supporting work in legal scenarios. Firstly, supporting results from [52,53], our experiment shows that the use of 3D models and animations can improve viewers' recall. Additionally, we found that applying 3D animations further decreases viewers' cognitive load for consuming narratives. However, 3D animations did not improve people's memorability compared to static 3D content. Based on participants' comments, a possible reason is that, compared to 3D model image videos, the 3D animations did not provide any additional visual cues that can further benefit viewers for recall.

Secondly, 3D content keeps viewers engaged. The results indicate that viewers can better keep focusing on the presented cases when 3D graphics are applied, and the use of 3D animations makes participants feel the presented cases more meaningful and less confusing. We believe this affordance of 3D graphics is beneficial for communicating narratives, as it means the use of 3D techniques can potentially keep users interested. These 3D techniques could simplify users' work in the communication of information in narratives, which could also benefit cognitive load and recall; thus furthering the user's working efficiency. Further investigation is required.

During our discussions with participants and analysis of the results and feedback, the potential for 3D visual content (i.e., models and animations) to introduce misleading information was raised. Specifically, communications in some real-world domains (law enforcement in our work) should be based solely on the known facts. However with use of 3D visual content, there is a high possibility to introduce misleading information and pose a potential distraction. For example, we did not apply any facial animations because they are not indicated in the narratives. However, in one of the 3D animation videos, some participants claimed that "the appellant was very calm when being arrested because he looks like (calm)". This is a significant issue in communicating factual information. As such, the misleading aspects of 3D models and animations should be carefully studied. Important research includes "what kinds of visual content could bring misleading information", "to what extent different misleading information affects the objectivity of communication", and "how to avoid the misleading information or at least reduce the negative effects".

To highlight, our work demonstrates the possibility of applying 3D models and animations to support people's consumption of court cases. However, we are not aggrandizing the applicability of these 3D techniques. We claim that the application of these techniques in real-world scenarios (e.g., law enforcement) is limited by issues such as bias and misleading as introduced above. In addition, currently the creation of 3D visual products costs great longer time than that of textual scripts. Therefore, in this work, we only suggest 3D models and animations can be potential options for people to consider in their work. More importantly, by exposing the value of 3D models and animations, we hope our work can motivate future research to overcome the problems and lower the barrier to entry these 3D techniques, and consequently, popularize these 3D techniques in real work environment.

4. Limitations

The novelty of our work is to evaluate the benefits of 3D content and animations for conveying narratives, focusing on the cognitive aspects. The first limitation comes from the selection of our participants; feedback and cognitive performance may vary across genders, ages, cultures and education backgrounds, however we did not consider these factors in the work presented in this paper. Further, our investigation was conducted on a low-level of cognitive function (i.e., recall, cognitive load, and engagement). There are more complex cognition-related measures that were not measured, such as neurocognitive mechanism [15]. This we identify as future work.

For the user study, we created 3D models and animations presenting three narratives, and compared them with text narratives, in order to test if 3D content can benefit in cognitive

load, recall, and engagement. We acknowledge that there are many other ways of visualizing the presented narratives. In this paper, we present our methods based on our experience and discussions with narrative experts. We claim that there would be bias that is inescapable with user-authored content. Our goal was to examine the power of 3D models and animations for presenting narratives compared to text. Since the 3D model images and 3D animation videos were made following the same process and applying the same resources, we believe the comparison in our experiment was fair. Additionally, while we tried to include all the information of the textual narratives in their 3D conditions, we did not develop a method to test if any non-evaluated information was missing when creating 3D model/animation conditions.

Finally, we should note that users' feedback and performance is always subject to the time and frequency of exposure in the technology. Although 3D computer-generated content is not novel for the general public, to use it for the consumption and comprehension of narratives is new for our participants, and therefore participants may show different results after greater exposure. Therefore, we assert that participants' first-time exposure to the technologies should be accounted for the results.

5. Conclusions and Future Work

In this paper, we present an empirical experiment evaluating the effectiveness of 3D models and 3D animations for recreating and communicating real-world narratives. We selected three real-world narratives with similar complexity, and created 3D models and animations for each of the narratives. Three measurements were applied. The first is the Paas scale that was utilized for measuring participant's cognitive load when consuming cases. The results showed that 3D animation technology can help to decrease viewers cognitive load. The second is the SDT, which was used to test viewers' memorability of information in narratives. According the the SDT results, we found that both 3D graphic conditions improve viewers' recall compared to text. Finally, the measurement of engagement was indicated by tests regarding focused attention and perceived usability. The results indicated that viewers can be better engaged when 3D graphical techniques were applied. Our results indicated the power of 3D technology for communicating narratives.

There are several interesting future research directions from this work. Firstly, we adopted three narratives with the same level of complexity (ID scores) in order to compare three presentation techniques. In the future, it is also important to test these techniques using narratives with different level of complexity. This could help to trace the effectiveness of presentation techniques and reveal their best suited level of complexity. Secondly, the evaluation and creation of visual information presentation should be based on the specific types of narratives. We gained an initial understanding of the effects of 3D content on viewers' cognitive functions in this work. In the future, we will focus on specific types of real-world narratives (e.g., a specific type of court cases, historical events, and news events). Thirdly, after demonstrating the power of 3D rendering techniques for communicating narratives, we are also interested in testing if this power is still true when abstract data visualizations are involved (such videos have been well accepted as "data videos"). Finally, we are keen to explore the use of Virtual Reality for storytelling given its potential to present information in a more immersive manner.

Author Contributions: Conceptualization, R.C., J.W., A.C. and B.H.T.; Data curation, R.C.; Formal analysis, R.C. and M.K.; Investigation, R.C.; Methodology, R.C. and M.K.; Project administration, B.H.T.; Supervision, J.W., A.C., R.T.S. and B.H.T.; Validation, J.W., A.C. and B.H.T.; Visualization, R.T.S.; Writing—original draft, R.C.; Writing—review & editing, R.C., J.W., A.C., R.T.S. and B.H.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the Data to Decisions Cooperative Research Centre whose activities are funded by the Australian Commonwealth Government's Cooperative Research Centres Programme.

Acknowledgments: We would like to thank the members of the Wearable Computer lab for their support during this investigation.

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

References

1. Mitra, S.; McEligot, A.J. Can Multimedia Tools Promote Big Data Learning and Knowledge in a Diverse Undergraduate Student Population? *Calif. J. Health Promot.* **2018**, *16*, 54. [[CrossRef](#)] [[PubMed](#)]
2. Mayer, R.E.; Heiser, J.; Lonn, S. Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *J. Educ. Psychol.* **2001**, *93*, 187. [[CrossRef](#)]
3. Kallinikou, E.; Nicolaidou, I. Digital Storytelling to Enhance Adults' Speaking Skills in Learning Foreign Languages: A Case Study. *Multimodal Technol. Interact.* **2019**, *3*, 59. [[CrossRef](#)]
4. Duval, K.; Grover, H.; Han, L.H.; Mou, Y.; Pegoraro, A.F.; Fredberg, J.; Chen, Z. Modeling physiological events in 2D vs. 3D cell culture. *Physiology* **2017**, *32*, 266–277. [[CrossRef](#)] [[PubMed](#)]
5. Marcortte, P. Animated Evidence-Delta 191 Crash Re-Created through Computer Simulations at Trial. *ABAJ* **1989**, *75*, 52.
6. Cao, R.; Dey, S.; Cunningham, A.; Walsh, J.; Smith, R.T.; Zucco, J.E.; Thomas, B.H. Examining the use of narrative constructs in data videos. *Vis. Inform.* **2019**, *4*, 8–22. [[CrossRef](#)]
7. Cao, R.; Walsh, J.; Cunningham, A.; Thomas, B.H. Examining the affordances for multi-dimensional data videos. In Proceedings of the 2017 International Symposium on Big Data Visual Analytics (BDVA), Adelaide, Australia, 7–10 November 2017.
8. Schofield, D. Playing with evidence: Using video games in the courtroom. *Entertain. Comput.* **2011**, *2*, 47–58. [[CrossRef](#)]
9. Schofield, D. Courting the Visual Image: The Ability of Digital Graphics and Interfaces to Alter the Memory and Behaviour of the Viewer. In *International Conference on Human-Computer Interaction*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 325–344.
10. Schofield, D. The future of evidence: New applications of digital technologies, forensic science: Classroom to courtroom. In Proceedings of the 18th International Symposium of the Forensic Sciences, Fremantle, Australia, 2–7 April 2006; pp. 2–8.
11. Ebert, L.C.; Nguyen, T.T.; Breitbeck, R.; Braun, M.; Thali, M.J.; Ross, S. The forensic holodeck: An immersive display for forensic crime scene reconstructions. *Forensic Sci. Med. Pathol.* **2014**, *10*, 623–626. [[CrossRef](#)]
12. Kumar, N.; Benbasat, I. The effect of relationship encoding, task type, and complexity on information representation: An empirical evaluation of 2D and 3D line graphs. *MIS Q.* **2004**, *28*, 255–281. [[CrossRef](#)]
13. Keller, T.; Gerjets, P.; Scheiter, K.; Garsoffky, B. Information Visualizations for Supporting Knowledge Acquisition-The Impact of Dimensionality and Color Coding. In Proceedings of the Annual Meeting of the Cognitive Science Society, Chicago, IL, USA, 4–7 August 2004; Volume 26.
14. Sweller, J.; Van Merriënboer, J.J.; Paas, F.G. Cognitive architecture and instructional design. *Educ. Psychol. Rev.* **1998**, *10*, 251–296. [[CrossRef](#)]
15. Antonenko, P.; Paas, F.; Grabner, R.; Van Gog, T. Using electroencephalography to measure cognitive load. *Educ. Psychol. Rev.* **2010**, *22*, 425–438. [[CrossRef](#)]
16. Marcus, N.; Cooper, M.; Sweller, J. Understanding instructions. *J. Educ. Psychol.* **1996**, *88*, 49. [[CrossRef](#)]
17. Paas, F.G. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* **1992**, *84*, 429. [[CrossRef](#)]
18. Brunken, R.; Plass, J.L.; Leutner, D. Direct measurement of cognitive load in multimedia learning. *Educ. Psychol.* **2003**, *38*, 53–61. [[CrossRef](#)]
19. Chandler, P.; Sweller, J. Cognitive load while learning to use a computer program. *Appl. Cogn. Psychol.* **1996**, *10*, 151–170. [[CrossRef](#)]
20. Iqbal, S.T.; Zheng, X.S.; Bailey, B.P. Task-evoked pupillary response to mental workload in human-computer interaction. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2004; pp. 1477–1480.
21. Paas, F.; Tuovinen, J.E.; Tabbers, H.; Van Gerven, P.W. Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* **2003**, *38*, 63–71. [[CrossRef](#)]
22. Tuovinen, J.E.; Paas, F. Exploring multidimensional approaches to the efficiency of instructional conditions. *Instr. Sci.* **2004**, *32*, 133–152. [[CrossRef](#)]
23. Brown, J.; Lewis, V.; Monk, A. Memorability, word frequency and negative recognition. *Q. J. Exp. Psychol.* **1977**, *29*, 461–473. [[CrossRef](#)]

24. Saket, B.; Endert, A.; Stasko, J. Beyond usability and performance: A review of user experience-focused evaluations in visualization. In Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization, Baltimore, MD, USA, 24 October 2016; pp. 133–142.
25. Borgo, R.; Abdul-Rahman, A.; Mohamed, F.; Grant, P.W.; Reppa, I.; Floridi, L.; Chen, M. An empirical study on using visual embellishments in visualization. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 2759–2768. [[CrossRef](#)]
26. Jianu, R.; Rusu, A.; Hu, Y.; Taggart, D. How to display group information on node-link diagrams: An evaluation. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1530–1541. [[CrossRef](#)]
27. Ademoye, O.A.; Ghinea, G. Information recall task impact in olfaction-enhanced multimedia. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2013**, *9*, 17. [[CrossRef](#)]
28. Abdi, H. Signal detection theory (SDT). In *Encyclopedia of Measurement and Statistics*; Sage: Thousand Oaks, CA, USA, 2007; pp. 886–889.
29. Haroz, S.; Kosara, R.; Franconeri, S.L. Isotype visualization: Working memory, performance, and engagement with pictographs. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 1191–1200.
30. Baldauf, D.; Burgard, E.; Wittmann, M. Time perception as a workload measure in simulated car driving. *Appl. Ergon.* **2009**, *40*, 929–935. [[CrossRef](#)] [[PubMed](#)]
31. Barnum, C.M.; Palmer, L.A. More than a feeling: Understanding the desirability factor in user experience. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2010; pp. 4703–4716.
32. Boy, J.; Detienne, F.; Fekete, J.D. Storytelling in information visualizations: Does it engage users to explore data? In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 1449–1458.
33. Saket, B.; Scheidegger, C.; Kobourov, S. Comparing Node-Link and Node-Link-Group Visualizations From An Enjoyment Perspective. In *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2016; Volume 35, pp. 41–50.
34. Amini, F.; Riche, N.H.; Lee, B.; Leboe-McGowan, J.; Irani, P. Hooked on data videos: Assessing the effect of animation and pictographs on viewer engagement. In Proceedings of the 2018 International Conference on Advanced Visual Interfaces, Grosseto, Italy, 29 May–1 June 2018; p. 21.
35. Bulté, B.; Housen, A. *Defining and Operationalising L2 Complexity*; John Benjamins: Amsterdam, The Netherlands; Philadelphia, PA, USA, 2012.
36. Pallotti, G. A simple view of linguistic complexity. *Second Lang. Res.* **2015**, *31*, 117–134. [[CrossRef](#)]
37. Cheung, H.; Kemper, S. Competing complexity metrics and adults' production of complex sentences. *Appl. Psycholinguist.* **1992**, *13*, 53–76. [[CrossRef](#)]
38. Lee, L.L. *Developmental Sentence Analysis: A Grammatical Assessment Procedure for Speech and Language Clinicians*; Northwestern University Press: Evanston, IL, USA, 1974.
39. Scarborough, H.S. Index of productive syntax. *Appl. Psycholinguist.* **1990**, *11*, 1–22. [[CrossRef](#)]
40. Kemper, S.; Sumner, A. The structure of verbal abilities in young and older adults. *Psychol. Aging* **2001**, *16*, 312. [[CrossRef](#)]
41. Kynette, D.; Kemper, S. Aging and the loss of grammatical forms: A cross-sectional study of language performance. *Lang. Commun.* **1986**, *6*, 65–72. [[CrossRef](#)]
42. Brown, C.; Snodgrass, T.; Kemper, S.J.; Herman, R.; Covington, M.A. Automatic measurement of propositional idea density from part-of-speech tagging. *Behav. Res. Methods* **2008**, *40*, 540–545. [[CrossRef](#)]
43. Snowdon, D.A.; Kemper, S.J.; Mortimer, J.A.; Greiner, L.H.; Wekstein, D.R.; Markesbery, W.R. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Jama* **1996**, *275*, 528–532. [[CrossRef](#)]
44. Criminal Case: R v ANDERSON [2017] SASCFC 125 (21 September 2017). Available online: <http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/sa/SASCFC/2017/125.html> (accessed on 3 July 2020).
45. Criminal Case: R v CORLETT [2017] SASCFC 112 (4 September 2017). Available online: <http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/sa/SASCFC/2017/112.html> (accessed on 3 July 2020).
46. Criminal Case: R v FARRER [2017] SASCFC 3 (10 February 2017). Available online: <http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/sa/SASCFC/2017/3.html> (accessed on 3 July 2020).

47. Trottier, D. *The Screenwriter's Bible: A Complete Guide to Writing, Formatting, and Selling Your Script*; Silman-James Press: Los Angeles, CA, USA, 1998; Volume 5.
48. McKee, R. Story: Substance. In *Structure, Style, and the Principles of Screenwriting*, Methuen; Methuen Publishing: North Yorkshire, UK, 1999.
49. Eysink, T.H.; de Jong, T.; Berthold, K.; Kolloffel, B.; Opfermann, M.; Wouters, P. Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *Am. Educ. Res. J.* **2009**, *46*, 1107–1149. [[CrossRef](#)]
50. Field, A. *Discovering Statistics Using IBM SPSS Statistics*; Sage: Thousand Oaks, CA, USA, 2013.
51. Makowski. The psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *J. Open Source Softw.* **2018**, *3*, 470. [[CrossRef](#)]
52. Krieger, R. Sophisticated computer graphics come of age and evidence will never be the same. *J. Am. Bar Assoc.* **1992**, *78*, 621–632.
53. Yadav, A.; Phillips, M.M.; Lundeborg, M.A.; Koehler, M.J.; Hilden, K.; Dirkin, K.H. If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *J. Comput. High. Educ.* **2011**, *23*, 15–37. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).