Title	A systematic review of digital storytelling in language learning in adolescents and adults
Author(s)	Nikki Zhi Li Lim, Azrifah Zakaria and Vahid Aryadoust

A Systematic Review of Digital Storytelling in Language Learning in Adolescents and Adults

1 **Abstract**

2 Digital storytelling (DST) is a novel approach that uses modern computer technology

3 to amplify language learning and teaching. The present study aims to review how the

4 published DST research utilizes visuals and audio to influence the learning environment

5 and engage adolescent and adult language learners. This was measured through their

6 improvement in the four main language skills: reading, writing, listening, and speaking.

7 A total 71 journal papers were identified using the Scopus database. The papers studied

8 both first and second language learning and were coded in full-text screening for the

9 research topics and methods adopted, theories or frameworks adopted, outcomes across

10 the language skills, and reliability investigation of the studies. The results showed a

11 range of research used in the studies, with 39.7% of the total studies using a mixed

12 number of methods. The theories adopted in these studies were limited to components

13 of DST, age group, and the type of study. Most studies neither tested nor mentioned the

14 use of the three theoretical variables mentioned above. Notably, a majority of the

15 studies reported positive outcomes when DST was used in the learning environment.

16 However, not all claims were supported with evidence. Lastly, only a handful of the

17 studies reviewed reported reliability, highlighting a lack of verification of the precision

18 of the measurement instruments used. Implications of these findings and

19 recommendations for designing DST and language learning research in the future will

20 be discussed.

21 *Keywords:* digital storytelling; reading; writing; speaking; listening

22

23

25

26

27

28

29

30

A Systematic Review of Digital Storytelling in Language Learning in Adolescents and Adults

Storytelling is the social and cultural activity of sharing stories, sometimes with improvisation, theatrics, or embellishment. New forms of media are creating new ways for people to record, express, and consume stories, most notably digital storytelling (DST).

DST is generally defined by researchers in the field of language teaching in terms of an extension of traditional storytelling and an integration of several forms of media. For example, Shin and Park (2008, p. 418) defined DST as "a form of storytelling that is conducted using digital technology as the medium or method of expression". Similarly, Hett (2012) proposed that DST is a new generation storytelling that fuses images, video, music, and audio through computers. Likewise, Yamaç and Ulusoy (2016) suggested that all digital stories are a combination of digital graphics, audios, videos, and music to showcase information, and that they have a certain theme and viewpoint as in the traditional stories. DST consists of a variety of formats such as text web pages, nonlinear interactive website, digital song, digital video (Chung, 2006), online game or virtual reality (VR) world (Shin & Park, 2008). In this present study, we have taken a more expansive approach to DST and defined it as a way of telling stories using digital media including webpages, song, online game, or even virtual and augmented reality (VR & AR).

DST has become a new form of communicating and expressing ideas and stories in language learning. DST can also help to make learning more interactive and fun for learners, especially younger learners. According to Xu et al. (2011), students can learn by doing when making digital stories. Similarly, Barrett (2005) argues that digital storytelling can assist in the convergence of four student-centred learning strategies, namely, student engagement, reflection for deep learning, project-based learning, and the effective integration of technology into instruction.

DST has been increasingly popular, and researchers have conducted studies to evaluate the effectiveness of DST on language learning by examining its significance in speaking (Alley-Young, 2017), listening (Tanrikulu, 2020a), reading (Sukovic, 2014), and writing (Girmen & Kaya, 2019). For example, Liu et al. (2019) investigated the effects of DST elementary students' reading skills in a two-year study. Using a

Chinese storytelling application, students in Liu et al.'s (2019) study shared and reviewed one another's stories and combined them to create a picture book. The results showed that students' reading skills improved over the two years as they consistently assessed and edited their stories. In another study, Lim and Noor (2019) studied the integration of Web 2.0 tools to enhance writing skills. The results were further broken down to language components like grammar and vocabulary, and it was determined that students made fewer mistakes in their writing throughout the study.

However, even as DST progressively made its way into language education (Robin, 2005), the effects derived from using DST as a pedagogical tool could not be fully evaluated (Signes, 2008). Boydell et al.'s (2012) review of DST in research highlighted that there are challenges in adapting creative activities like DST for research uses, particularly because the methods of research or analysis may be subjective or unreliable. As research on language learning, particularly on second language acquisition, is still in its early stages, it is important that researchers and educators do not hastily apply pedagogical tools based on the findings (Tarone et al., 1976). Chamot (2004), for example, stated that language learning strategies are largely unobservable as different learners' have different ways of learning. The studies in which DST have been used examined a variety of constructs associated with language learning such as motivation (Radaideh et al., 2020), creativity (Rubino et al., 2018) and cognitive skills (Foley et al., 2013). However, a closer examination of these studies shows that some of the claims regarding the utility of DST in enhancing these constructs of language leaners' is not well supported by evidence. For example, Batsila and Tsihouridis (2016) concluded that learners were more motivated in learning when learners claimed the DST software used was easy to use but did not provide any documentation of the evidence supporting the claims made. Therefore, a review of the theoretical contributions and methodological procedures in DST research in the field of language learning is essential for understanding the practical value of DST and ensuring the reliability of research instruments used. However, to date, there has not been any reviews conducted to investigate the quality of research on DST in terms of the methods and frameworks adopted in language learning.

The present study aims to provide a critical review on the design of studies from multiple perspectives such as the components and theory of DST, type of study, age group and DST software used in each study (refer to the Methodology section for

further details on the coding scheme). We formulated the research questions (RQs) addressed in this study following the Sample, Phenomenon of Interest, Design, Evaluation, Research type (SPIDER) theoretical framework by Cooke et al. (2012). The SPIDER tool is used to come up with effective search strategies for qualitative, quantitative, and mixed methods research.

1. What are the research topics and methods that have been adopted in previous DST research?
2. What are the theories or frameworks of DST that have been adopted in previous DST research?
3. How were the studies designed in terms of the reliability of the instruments used?
4. How do the outcomes of using DST vary across the four language skills and variables in the coding scheme?

Methodology

Dataset

The Scopus database was used to identify publications that investigated the application of storytelling in the teaching and learning of listening and speaking skills of young and adult learners. Papers from conferences, workshops and other events were not included in the data. The different stages of the systematic review conducted are visualized using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement. The PRISMA Statement was designed for transparent and complete analyses of systematic reviews (Sarkis-Onofre et al., 2021). Figure 1 shows the PRISMA Statement diagram of the stages of research.

Studies on DST have emerged and been indexed in Scopus since 1997. As of September 2020, the quantity of papers has increased significantly, resulting in a total of 1,605 published papers. In the past five years (2016 to 2020), 141, 165, 203, 202, and 192 manuscripts that had some link with DST were published on each consecutive year. The peak publication years were 2018, 2019, 2020. Giving consideration that the publications were not representative of the total publications in 2020 as the search period in Scopus was limited to September 2020, we can see an increasing trend in the study of DST, indicating researchers' peak in interest of the topic.

The 1,605 publications from Scopus were further screened by limiting the search protocol for DST in language learning. We used the Boolean (AND/OR) search protocol "Reading" OR "writing" OR "speaking" OR "listening" as primary search terms along with "digital storytelling". We discovered that 1,374 publications had no relation to the respective language skills and, thus, excluded them from the search. The remaining 231 publications were then grouped into multiple datasets, relating DST to the four main language skills: (i) reading, which consisted of 62 publications including 27 (43.5%) journal articles (ii) writing consisting of 122 publications, including 77 (63.1%) journal articles, (iii) speaking consisting of 22 publications, of which 12 (54.5%) were journal articles, and (iv) listening which consisted of 25 publications inclusive of 10 (40.0%) journal articles (see Appendix A for the search code.)

Next, we browsed through each dataset, focusing on the title and abstracts of the publications, and highlighted the publications which focused on the relationship between digital storytelling and the four respective language skills in terms of language learning. A set of inclusion and exclusion criteria for the analysis were also designed and defined from three main perspectives: target language, publication source, and research design. Table 1 presents the detailed inclusion and exclusion criteria.

Table 1

Inclusion and Exclusion Criteria

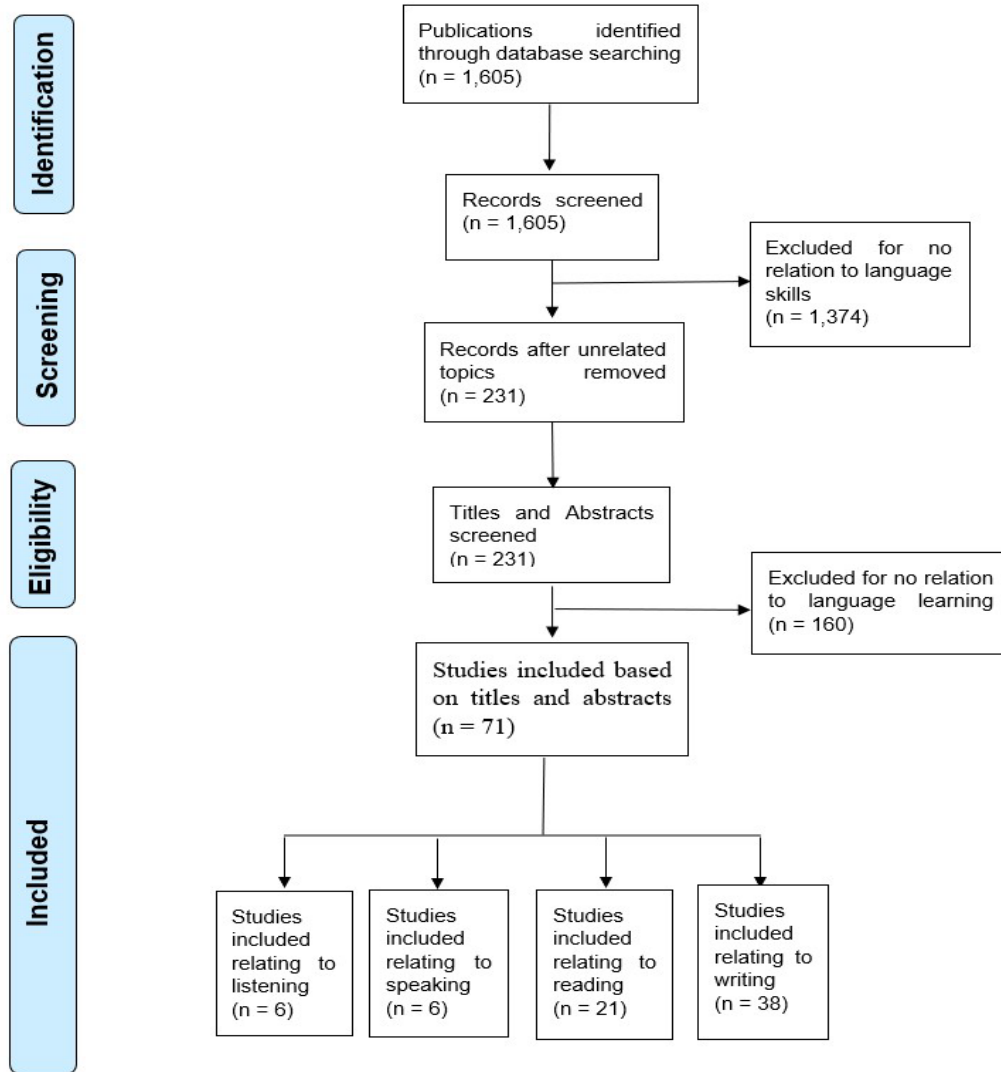| Inclusion criteria: The paper…	Exclusion criteria: The paper…
1. investigated English as second or foreign language learning or teaching.	1. investigated foreign languages other than English.
2. was a journal article or a review paper.	2. was part of a conference proceeding.
3. investigated the effects of DST on learning or teaching any of the four language skills.	3. investigated skills or topics unrelated to the teaching or learning of the four language skills.

Through the screening process, 13 papers were found to have investigated more than one language skill and were hence repeated in the 71 publications identified. These papers were not excluded in the total publications to avoid any confusion during the analysis. After reading the content of the publications in each dataset, we identified the publications that examined listening (n=6), speaking (n=6), reading (n=21) and writing (n=38) skills relating to DST (see Appendix B for the complete lists of publications and the skills examined in each).

155

Figure 1

Visualization of stages of research (PRISMA)

158

160

161

Coding Scheme

A coding scheme was adapted from Plonsky and Oswald (2012) to examine substantive

factors in data collection, data analysis, and learning outcomes. We read through the 71

publications which included 13 repeated papers for the various language skills and

identified relevant content and data that were split into five categories: study

descriptors, study context, research design, measures, and research outcome (Plonsky

& Oswald, 2012). For each category, we developed a preliminary list of variables to code for, which was refined during the coding process. A total of 30 variables were recorded in the final coding scheme (refer to Appendix C). We highlight several variables below that require further elaborations:

- **Age group** – To categorize the age of participants, we adapted the age categorization from Erikson's Psychosocial Development Theory (Erikson, 1963). The theory divides human development into eight stages: infant to 18 months, 18 months to three years, three to five years, five to 13 years, 13 to 21 years, 21 to 39 years, 40 to 65 years, and 65 years and older. Furthermore, it establishes a relationship between learners' developmental needs at different stages in life and their education (Batra, 2013).
- **Type of study** – We separated the journals into quantitative and qualitative studies following the definitions by Mackey and Gass (2005). *Commentary studies* was added as a third category during the coding process after discovering a few studies (n=16; 27.6%) focused on the authors' opinions on the topic of DST.
- **Components of DST** – Another group of variables was adapted from the Centre for Digital Storytelling (CDS), which identifies seven distinctive elements of DST: point of view, dramatic question, emotional content, gift of voice, power of soundtrack, economy, and pacing (Robin, 2008). According to CDS, this framework is key in making DST effective (Lim & Noor, 2019). Table 2 provides a brief definition of the seven key elements of DST.

Table 2

Definitions of the elements of DST

Element	Definition
Point of view	What is the main point of the story and what is the perspective of the author?
A dramatic question	A key question that keeps viewer's attention and will be answered by the end of the story.

Element	Definition
Emotional content	Serious issues that come alive in a personal and powerful way and connects the story to the audience.
Gift of voice	A way to personalize the story to help the audience understand the context.
Power of soundtrack	Music or other sounds that support and embellish the storyline.
Economy	Using just enough content to tell the story without overloading the viewer.
Pacing	The rhythm of the story and how slowly or quickly it progresses.

194

195 • **Miscellaneous constructs –** During the coding process, we discovered several
196 additional features or constructs that were mentioned in the studies. This
197 included constructs such as motivation, creativity, cognitive skill, social skill,
198 communication, self-efficacy, and transliteracy skills. Several papers had
199 indicated these concepts in the study and therefore we included them in the final
200 coding scheme.

201

202 Adapting the preceding variables in our coding scheme allowed us to investigate
203 the replicability of DST research, which would be facilitated if the design of the original
204 studies were clear and authors provided sufficient information concerning the
205 theoretical framework, variables, features of the sample (e.g., age group, first/second
206 language, and so forth), instruments, and data analysis.

207 **Data Analysis**

208 To address RQ 1, we grouped variables as descriptive factors (e.g., language level,
209 gender, research method) to investigate the frequency and percentage of
210 implementation of DST. For RQ 2, components of DST, age group and type of study
211 were labelled as theoretical variables to identify the frequency and percentage of their
212 application. RQ 3 was evaluated by identifying the frequency of methods used for
213 reliability assessment. RQ4 was investigated by assessing and comparing the datasets
214 of the four language skills on the effect size and outcomes stated in the publications.

Threats to Validity and Reliability

216   There might have been potential bias when selecting and assessing the publications

217   during the screening process. To minimize the effects of selecting bias, the inclusion

218   and exclusion criteria, as well as the PRISMA statement was used to distinctly identify

219   the publications suitable for review. This reduces the ambiguity and the possibility of

220   poor reproducibility, which potentially decreases random error. In addition, while

221   developing the datasets, only one coder was reviewing and coding the articles, although

222   the relevance of the articles was later checked by two independent coders. We suggest

223   that a second coder should be invited to ensure that no pertinent publications were

224   excluded. This can also reduce reviewer bias when selecting studies.

225   **Results**

226   Overall, reviewing the publications showed a variety of results. Tobin and Blanton

227   (2014), Pardo (2014), Shelby-Caffey et al. (2014), Xu et al. (2011) were the most highly

228   cited publications in research on each of these language skills. Most of the studies listed

229   were conducted for students and language learners. However, Andayani (2019) and

230   Shelby-Caffey et al. (2014) focused on the effectiveness of English student teachers

231   preparing digital stories for their students, which presents a different perspective on

232   how DST can be used effectively to successfully teach languages to learners. We also

233   found that many publications (e.g., Hava, 2019, Foley et al., 2013; Lee, 2014; Pardo,

234   2014) that examined the effectiveness DST in learners' engagement and motivation in

235   language learning. Two studies (Evmenova & Regan, 2019; Strassman & O'Dell, 2012)

236   focused on the relationship between DST and language learning among learners with

237   disabilities.

238

239   **Descriptive Features**

240   Table 3 shows the frequency and percentage of different types of descriptive features

241   adopted in the dataset of DST publications. For first and second language, the writing

242   dataset had the most journals studying the language skill for both levels (n = 4; 6.9% &

243   n = 10; 17.2%). The main target language was English (n = 17; 29.3%), with the

244   majority of the studies examining writing (n = 9; 15.5%). Notwithstanding that most of

245   the studies (n = 41; 70.7%) did not report on the gender of the participants, it can be

246   noted that those that did report gender were found to have a mix of participants in their

studies (n = 12; 20.7%). The research methods applied included observation (n = 3; 5.2%), interview (n = 2; 3.4%), journal (n = 3; 5.2%), review assignment (n = 2; 3.4%), test score (n = 7; 12.1%), mix (n = 23; 39.7%), and not mentioned (n = 18; 31.0%). Notably, a mix of research methods used was the most prevalent (n = 23; 39.7%) among the studies. Among the methods of data analysis for research, t-test was the dominant model used, especially among the writing (n = 7; 12.1%) and reading (n = 3; 5.2%) datasets. This is followed by a use of mixed data analysis methods in the writing (n = 4; 6.9%), reading (n = 2; 3.4%) and speaking (n = 1; 1.7%) datasets. There were various types of software used in the studies on DST. However, it was noted that apart from the speaking dataset, the majority of studies did not mention the software used for each of the language skills: listening (n = 2; 3.4%), reading (n = 11; 19.0%), writing (n = 17; 29.3%), mix (n = 3; 5.2%). Storybird (n = 3; 5.2%) and Microsoft Photo Story 3 (n = 3; 5.2%) were the most popular software used, with the other software having only one paper (1.7%) reporting the use of it.

Table 3

Descriptive variables for DST dataset

Variable		Number of journals (%)				
		Listening Dataset	Speaking Dataset	Reading Dataset	Writing Dataset	Mix Dataset
Language level	First language	0 (0.0%)	1 (1.7%)	1 (1.7%)	4 (6.9%)	1 (1.7%)
	Second language	1 (1.7%)	1 (1.7%)	5 (8.6%)	10 (17.2%)	3 (5.2%)
	Not mentioned	2 (3.4%)	0 (0.0%)	9 (15.5%)	17 (29.3%)	3 (5.2%)
Target language	English	0 (0.0%)	2 (3.4%)	4 (6.9%)	9 (15.5%)	2 (3.4%)
	Non-English	1 (1.7%)	0 (0.0%)	2 (3.4%)	8 (13.8%)	1 (1.7%)
	Not mention-ed	2 (3.4%)	0 (0.0%)	9 (15.5%)	14 (45.2%)	4 (6.9%)
Gender	Female	0 (0.0%)	0 (0.0%)	2 (3.4%)	1 (1.7%)	0 (0.0%)

Variable		Number of journals (%)				
		Listening Dataset	Speaking Dataset	Reading Dataset	Writing Dataset	Mix Dataset
	Male	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	1 (1.7%)
	Mix	1 (1.7%)	1 (1.7%)	4 (6.9%)	6 (10.3%)	0 (0.0%)
	Not mentioned	2 (3.4%)	1 (1.7%)	9 (15.5%)	23 (39.7%)	6 (10.3%)
Research method	Observation	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (3.4%)	1 (1.7%)
	Interview	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (6.5%)	0 (0.0%)
	Journal	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (5.2%)	0 (0.0%)
	Review assignment	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (3.4%)	0 (0.0%)
	Test score	0 (0.0%)	0 (0.0%)	3 (5.2%)	4 (6.9%)	0 (0.0%)
	Mix	2 (3.4%)	1 (1.7%)	5 (8.6%)	10 (17.2%)	5 (8.6%)
	Not mention-ed	1 (1.7%)	1 (1.7%)	7 (12.1%)	8 (13.8%)	1 (1.7%)
Method of data analysis	t-test	0 (0.0%)	0 (0.0%)	3 (5.2%)	7 (12.1%)	0 (0.0%)
	ANOVA	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	1 (1.7%)
	ANCOVA	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)	0 (0.0%)
	MANOVA	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	MANCOVA	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)	0 (0.0%)
	Data triangulation	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Thematic analysis	0 (0.0%)	0 (0.0%)	1 (1.7%)	2 (3.4%)	1 (1.7%)
	Mixed	0 (0.0%)	1 (1.7%)	2 (3.4%)	4 (6.9%)	0 (0.0%)
	Not mentioned	3 (5.2%)	1 (1.7%)	7 (12.1%)	16 (27.6%)	5 (8.6%)

Variable		Number of journals (%)				
		Listening Dataset	Speaking Dataset	Reading Dataset	Writing Dataset	Mix Dataset
Software used	Windows Movie Maker	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Microsoft PowerPoint	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)	0 (0.0%)
	Storybird	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (5.2%)	0 (0.0%)
	Microsoft PhotoStory 3	0 (0.0%)	1 (1.7%)	0 (0.0%)	0 (0.0%)	2 (3.4%)
	Adobe (Premiere, Photoshop)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Fiabot!	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)
	GoAnimate	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Meschola	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Kerpoof	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Second Life	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Instagram	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	A Chinese storytelling application	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Mix	1 (1.7%)	1 (1.7%)	3 (5.2%)	3 (5.2%)	1 (1.7%)
	Not mentioned	2 (3.4%)	0 (0.0%)	11 (19.0%)	17 (29.3%)	3 (5.2%)

Theoretical Frameworks and Features

Table 4 demonstrates the component of DST, age group, and type of study in the DST dataset. Overall, it was found that most of the research papers did not report on the seven previously discussed components of DST (n = 44; 75.9%). Among the studies that did report paying heed to the components of DST in the design of the study, six (10.3%) papers reported a mix of components was the greatest number of papers from the writing dataset, followed by the mix dataset (n = 2; 3.4%) and listening dataset (n = 1; 1.7%) respectively. *Point of view* was the least tested component among the

researched components and was only reported in one paper (1.7%) under the writing dataset. In addition, components like *Power of soundtrack* and *Emotional content* were not reported or operationalized in these studies.

While the majority of the studies (n = 29; 50.0%) did not state the age group in which the participants of the study fell into, there were two prevailing age groups investigated on, including 5 to 13 years (n = 12; 20.7%) and 13 to 21 years (n = 10; 17.2%). The types of study included qualitative (n = 17; 29.3%), quantitative (n = 8; 13.8%), commentary (n = 16; 27.6%) and mix (n = 12; 20.7%). There were also a handful of studies (n = 5; 8.6%) which did not explain the type of study done. Notably, qualitative studies were the most frequently used with nine of the papers (15.5%) listed under the writing dataset, four papers (6.9%) from the mix dataset, two papers (3.4%) each from the listening and reading dataset. The listening dataset was noted to not include any qualitative studies.

Table 4
Investigation of Theoretical Variables for DST Dataset

Variable		Number of journals (%)				
		Listening Dataset	Speaking Dataset	Reading Dataset	Writing Dataset	Mix Dataset
Component of DST	Gift of voice	0 (0.0%)	0 (0.0%)	2 (3.4%)	0 (0.0%)	2 (3.4%)
	Point of view	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
	Mix	1 (1.7%)	0 (0.0%)	0 (0.0%)	6 (10.3%)	2 (3.4%)
	Not mentioned	2 (3.4%)	2 (3.4%)	13 (22.4%)	24 (41.4%)	3 (5.2%)
Age group	5 – 13 years	0 (0.0%)	0 (0.0%)	3 (5.2%)	5 (8.6%)	4 (6.9%)
	13 – 21 years	0 (0.0%)	1 (1.7%)	3 (5.2%)	5 (8.6%)	1 (1.7%)
	Mix	0 (0.0%)	0 (0.0%)	2 (3.4%)	4 (6.9%)	1 (1.7%)
	Not mentioned	3 (5.2%)	1 (1.7%)	7 (12.1%)	17 (29.3%)	1 (1.7%)
Type of study	Qualitative	2 (3.4%)	0 (0.0%)	2 (3.4%)	9 (15.5%)	4 (6.9%)
	Quantitative	0 (0.0%)	0 (0.0%)	2 (3.4%)	6 (10.3%)	0 (0.0%)
	Commentary	1 (1.7%)	1 (1.7%)	5 (8.6%)	7 (12.1%)	2 (3.4%)
	Mix	0 (0.0%)	1 (1.7%)	4 (6.9%)	6 (10.3%)	1 (1.7%)

Not mentioned	0 (0.0%)	0 (0.0%)	2 (3.4%)	3 (5.2%)	0 (0.0%)

Testing Reliability

Table 5 presents data on the reliability of research instruments and data coding practices in the studies in the MALL dataset. The reliability of the studies was significantly under-researched. The papers that did not report reliability consisted of 26 papers (44.8%) from writing dataset, 14 papers (24.1%) from the reading dataset, five papers (8.6%) from the mix dataset, three papers (5.2%) and two papers (3.4%) from the speaking dataset. Of the papers that recorded the reliability statistics, inter-rater reliability (n = 4; 6.9%) was the most frequently used index. Cronbach's α index was also used in 2 papers (3.4%) from the writing and mixed datasets. These papers are Chiang (2020) and Tobin and Blanton (2014) respectively.

Table 5

Reliability indices used in the DST dataset

Reliability Indices	Number of journals (%)				
	Listening Dataset	Speaking Dataset	Reading Dataset	Writing Dataset	Mix Dataset
Cronbach's α	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (3.4%)	1 (1.7%)
Inter-rater reliability	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (9.7%)	1 (1.7%)
Mix	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.7%)	0 (0.0%)
Not mentioned	3 (5.2%)	2 (3.4%)	15 (24.1%)	26 (44.8%)	5 (8.6%)

Effect Size and Outcomes

We also examined the effect size and outcomes reported in papers that conducted quantitative or qualitative studies (i.e., not solely commentary). A total of 38 papers (65.5%) were found to have conducted either quantitative, qualitative or mixed studies.

Of the 38 papers, only three papers (5.2%) reported on the effect size. Therefore, it was not possible to conduct a meta-analysis on the effect size of the studies.

While examining the papers, we discovered that all papers (n = 37; 63.8%) apart from one (Gutierrez et al., 2019; 1.7%) concluded that DST aids in improving language learning. Using t-test with a 95% level of confidence, Gutierrez et al. (2019) discovered that there was no significant difference in the improvement of language learning when using DST compared to traditional storytelling. Furthermore, only 19 (32.8%) of the 38 papers recorded the results of their studies. Thus, there was no written evidence in the remaining 19 papers (32.8%) to support their claims that DST aids in enhancing language learning. Please refer to Appendix D for details of outcomes reported in the publications in the dataset.

Examining Miscellaneous Constructs

While investigating if the application of DST can influence the development of language learners' language skills, we discovered that 24 papers (41.4%) claimed DST also contributed to several other aspects of learning. These constructs included motivation (n = 21; 36.2%), creativity (n = 3; 5.2%), cognitive skill (n = 4; 6.9%), social skill (n = 6; 10.3%), emotional skill (n = 2; 3.4%), transliteracy skill (n = 4; 6.9%), self-efficacy (n = 3; 5.2%), communication (n = 1; 1.7%), and presentation skill (n = 1; 1.7%). Evidently, motivation (n = 21; 36.2%) was the most frequently mentioned aspect that researchers believed DST have made an improvement upon while conducting their studies. However, these claims, lamentably, were made with no supporting evidence in the papers and were only based on researchers' perceptions rather than rigorous data to prove that DST did in fact help with the improvement of these miscellaneous constructs.

Discussion

RQ1: Research Topics and Methods Adopted

Overall, our results indicated that many of the studies did not provide essential information about the topics and methods adopted. These included the target language and research methods used. First, we found that a large number of the 58 studies did not mention the language level (i.e., first language, second language) tested (n=29; 50.0%) or the language in which the study was done (n=27; 46.6%). This highlights a certain negligence in the DST and language research community regarding the importance of languages spoken and language level in DST and language learning

assessment. First and second language play an important role in language learning. Positive transfer indicates the first language knowledge assisting the acquisition of second language, while negative transfer refers to first language negatively impacting second language acquisition (Selinker, 1983). Furthermore, with the addition of a different target language, language performance may be affected as well (Gogolin, 2012). In the event a target language in second language has different linguistic properties compared to first language, learners acquire less vocabulary in both languages than learners who focus on their first language (Gogolin, 2012). Therefore, underestimating the language level and target language may lead to inaccurate observations and data due to intervening variables when studying the effects of DST on language learning.

A total of 38 (65.5%) papers made no mention of the gender of participants when conducting their studies. Similar to language level and target language, there may be a level of unawareness with regards to the effects of gender, on DST and language learning. For example, Wehrwein et al. (2006) found that the majority of males preferred multimodal methods of presentation as compared to females. However, this effect of gender on users' preference for DST, and its consequent effect on learnability, was not examined in any of the studies in the DST dataset. Therefore, this finding is not generalizable beyond the context of the study, and thus has no or minimal replicability.

In terms of the research methods used, about 40.0% of the studies used a mixed number of methods to record data found. In contrast to the previous variables, this signals a greater consideration in the variety of methods used when recording findings. However, it is also noted that only 12.1% of the studies used test or assignment scores to document their results (Radaideh et al., 2020; Liu et al., 2019). This suggests the greater application of subjective research methods like observations, interviews and journal reflections which may influence generalizability and replicability of the conclusions drawn from DST and its effects on language learning. This also ties in with the majority of studies being qualitative studies, which will be discussed in the next section.

Overall, a significant number of studies used the t-test as their instrument of data analysis among quantitative studies (n = 20; 34.4%). While there is certain popularity of using t-test in DST research due to its simplicity in determining the

370   differences in mean between groups (Kim, 2015), there are also limitations to utilizing

371   this statistical test. Generally, when using t-test, there are several data considerations to

372   take into account. Firstly, the data must be free from outliers as it will affect the

373   accuracy and mean significant testing of the results (Jankowski et al., 2017). Secondly,

374   the sample size of the data must be large enough to resemble the sampling distribution

375   of the mean of the normal distribution (Jankowski et al., 2017). However, the studies

376   which conducted t-tests did not carry out or report any analyses that consider these

377   criteria. Furthermore, the interpretation of the results of t-test is directed by the

378   probability, or p-value, of the outcome and does not provide sufficient proof on research

379   (Jankowski et al., 2017).

380   Another element in our coding scheme was the type of software used. Out of 58

381   papers, 33 did not specify the DST software used when carrying out the research. This

382   signals the underestimation of the type of software used, like Storybird (Şimşek, 2017;

383   Chiang, 2020) and Microsoft Photo Story 3 (Flihan, 2013; Pardo, 2014), when teaching

384   using DST. While it is evident that new technologies are becoming more sought-after,

385   and there have been evidence indicating the increased adoption of technologies in

386   language learning (Shadiev & Yang, 2020), it is not possible to adopt a software or

387   application that is suitable for everyone to use (Dam et al., 2005). The type of software

388   used for implementing DST into language learning may be a hindrance to the

389   improvement of learners' skills in the event the software is too complicated. Thus, the

390   reporting of software used for DST and their features in terms of usability and

391   accessibility should be encouraged.

392   Overall, the results of the analysis of the first research question show that

393   previous DST research did made little attempt to analyze and discuss the research topics

394   and methods adopted in studies. This means that there was little consideration on the

395   factors that can affect the results of the studies and cause them to be inaccurate.

RQ2: Theories and Frameworks Adopted

397   In all, the results obtained also showed that there was a lack of important data about the

398   theories and frameworks adopted in DST research. Lamentably, the component of DST

399   investigated or taken into account when designing the materials (i.e. gift of voice, point

400   of view, soundtrack) was not mentioned in more than 75.0% of the DST studies. This

401   information would not only provide information on the pattern in which learners create

their story content, but also support research needs with regards to the type of media and technology used, as well as the interactivity of DST (Kogila et al., 2020). These components also guide learners to think more creatively and include more elements into their storytelling. We understand that due to the limited duration of the studies and sample size, it might have been difficult to assess this variable in detail. However, we strongly encourage that it be reported in DST research.

Only half of the studies stated the age of participants, which leads to an unsettling negligence in the analysis of how age can affect the learning process. Given that children have an advantage over adults in learning languages because of their mental flexibility (Ghazali, 2006) and the plasticity of their brains (Lightbown & Spada, 1999), it behooves researchers to indicate the age demographic of their participants. Even within the category of children, information on ages of participants may prove crucial. For example, study conducted by Genesee (1978) found that a group of older students had fared better in a variety of language tests than younger students after enrolling in the same language program, possibly, as Genesee (1978) argued, due to the older students' greater motivation or level of competency.

We found that there were more qualitative studies compared to other types of studies among DST data. Qualitative research provides detailed descriptions on the participants' feeling and thoughts (Denzin, 1989), as well as researchers' reflections on the study. Particularly on the topic of DST and language learning, it provides greater insight on how much participants language skills have improved when using DST. However, qualitative research does not require a large sample size as compared to quantitative research, which raises the question of the whether we are able to generalize the data to the whole population of the research (Harry & Lipsky, 2014). Furthermore, due to the method of sampling, researchers may be biased when recording participants' claims and observations, resulting in low credibility of results (Rahman, 2016).

To conclude, while there have been efforts made to report and evaluate the type of study among DST research, the age group and the components of DST were under-investigated. This highlights a potential lack of awareness among DST researchers on the importance of analyzing such theories and frameworks which can be detrimental to the results of the effects of DST on language learning.

RQ3: Reliability of Instruments

Out of the 58 studies in the DST dataset, close to 90% of studies in the DST dataset did not report the reliability of the instruments used. This raises the question of consistency and precision of instruments used and results obtained. The lack of reliability testing can cause the interpretation of results to be inconsistent and indistinct (Grabowski & Oh, 2018). Inadequate reliability can also cause the true correlation between the instrument used and variable tested may be inaccurate (John & Soto, 2007). Furthermore, due to the low sample size of reliability testing, we were not able to conduct a meta-analysis as there were low numbers of effect size. For the studies which reported reliability of instruments, there was a lack of depth in the testing of reliability, which is consistent with Plonsky and Derrick's (2016) finding that most of the time, researchers failed to report reliability beyond the minimum requirements.

Among the DST dataset that reported reliability, inter-rater reliability was the most commonly used index, followed by Cronbach's α. However, limited further analysis was done by the studies to investigate the reliability of instruments in detail. As such, greater effort needs to be in place to evaluate the reliability of instruments in the field of DST research.

In summary, the reliability of instruments and methods used in the studies was severely under-evaluated. This highlights the negligence and disregard of analyzing the consistency and replicability of DST research, particularly for quantitative research.

RQ4: Outcomes of Review

Overall, our results show that most of the studies in the DST dataset supported the claim that DST enhances language learning although some of the studies did not provide ample evidence to support their claims. 12 studies presented supporting evidence for DST in language learning.

Among the studies that examined listening, Tarinkulu's (2020) study with 26 participants indicated that DST was in fact useful in improving learners' listening skills through researchers' observations and discussions with participants. On the other hand, while Andayani (2019) also agreed that listening skills are enhanced using DST, the study did not provide any supporting evidence. This shows that it is not possible to conclude that DST does aid in improving listening skills due to the lack of evidence in the studies. Furthermore, the DST listening dataset is limited and their results cannot be generalized to the population of language learning and listening.

For studies that investigate speaking skills, only one study in the dataset conducted an experiment and concluded that DST does help with speaking. Yang et al.'s (2020) study had 54 participants and provided quantitative data by conducting proficiency tests and presentations to determine the progress of learners' speaking skills. However, the study did not analyze any components of DST or provide the effect size of the experiment.

In the writing subset, Chiang (2020) reported the improvements of the writing skills of 18 participants which was tested using a written test. Meanwhile, Azis and Husnawadi (2020) interviewed their 28 participants and analyzed their reflections after the experiment before concluding that DST does improve language writing skills. Contrastingly, studies like Diaz (2016) and Lee (2014) did not present sufficient evidence. This highlights the contrast among the studies in the writing dataset, where some studies were more thorough in their research to support their claims with sufficient evidence.

Lastly, for studies that evaluated reading, Oakley et al. (2018) and Batsila and Tsihouridis (2016) with 37 and 51 participants respectively agreed that DST enhance reading skills. Furthermore, both studies provided evidence and obtained their results by conducting reading tests in their research methods. By contrast, although Gutierrez et al.'s (2019) study with 43 participants also concurred that DST was useful, they explained that there was no significant difference in using DST compared to traditional storytelling. This emphasizes that while DST is widely well-received among researchers and educators, whether it necessary to have it replace conventional storytelling is still under-investigated.

In terms of the outcomes of this study across the variables coded, we have seen limited data provided and investigated to determine that DST is useful. These findings resonate with Robin's (2014) evaluation of determining the effectiveness of DST in language learning. Robin (2014) suggested that a large amount of data on DST, and the teachings and learning of DST must be collected and analyzed to fully understand the usefulness of DST. Hence, we urge DST researchers to continue expanding on their studies and findings, to include rigorous methods in their experiments to effectively discover the usefulness of DST in language learning.

Limitations

This study provided important information on the application of DST in language learning but is not without its limitations. Due to the lack of publications reporting the effect size for their research, we were unable to perform a meta-analysis. This results in a limitation of the research outcomes and thus may affect the analysis of the effects of DST on language learning in adolescents and adults. In addition, we limited the scope of the study to adolescents and adult populations, thus leaving out younger learners. It is recommended that future researchers investigate the utility of DST in children's language learning. Finally, the links between DST as a field of research with other research fields was not examined. Future researchers can apply Scientometric techniques to examine possible relationships. A useful approach is the application of dual overlay maps that visualize the links between different fields of research on a global map of research. Authors (XXXX) present an example of the application of this technique in computer-based educational research.

Conclusion

This study aimed to review the design of studies of DST to investigate its usefulness with regards to language learning and the four main language skills. We started with a extensive literature search in Scopus and identified 58 DST journal articles, excluding the 13 repeated papers, that fitted into our criteria. These studies were then coded and analyzed to address the four research questions we had.

Research question one focused on the research methods and techniques that were adopted in the studies. While there were various topics coded, we found that most papers did not indicate them in their studies. For research question two, the theories and frameworks used in the studies were not provided in the majority of studies. However, qualitative research was notably the most popular form of study, allowing researchers to dive deeper into understanding the thoughts and feelings of participants. Research question three concerned the reliability of instruments in the DST dataset. We discovered that close to 90.0% of studies did not include any reliability testing. Among studies which did investigate reliability, inter-rater reliability was the most popular. Finally, research question four evaluated on the overall outcomes of DST in terms of the language skills and variables coded. While there was sufficient sample size for writing and reading datasets, the listening and speaking datasets were very limited.

529 Furthermore, the absence of rigorous data in the DST dataset made it difficult to
530 conclude that DST is in fact useful in improving language learning.

531 **Recommendations for Future DST Research**

532 Following the findings of our review, we propose several suggestions for future
533 DST research. First, DST and language learning researchers should provide detailed
534 definitions of their research topics and methods. For example, providing the details and
535 explanations of the questions in questionnaires or interviews used to capture data from
536 participants, or specifying the features of DST software used and participants' digital
537 literacy in order to avoid any discrepancies in data collection.

538 Second, in terms of the theories and frameworks adopted by DST researchers,
539 it is important that sufficient explanation and reasoning are provided to support and
540 highlight any possible advantages certain participants might have over others that might
541 affect the accuracy of results. For instance, researchers should analyze the difference in
542 learning abilities between groups divided by factors such as age, gender and language
543 proficiency level before conducting the experiments. The related data (ages and gender
544 of participants; components of DST) should be reported. We encourage future DST
545 research to include a more objective research study framework which spells out the
546 theoretical framework of the study and aligns it with suitable analytical techniques.

547 Finally, to ensure the validity and reliability of DST research, researchers should
548 report the validity and reliability of instruments used in their studies. This is essential
549 as validity and reliability testing ensure the accuracy and consistency of the instruments
550 used and results obtained. It is hoped that with these recommendations, future research
551 can make evident the outcomes and advantages of DST as a learning tool, leading to
552 greater confidence in the method. Demonstrable outcomes will go a long way in the
553 adoption of this promising approach to language learning.

554

555 **References**

556 Alley-Young, G. (2017). Creating digital videos in an ESL learning community to
557 develop communication skills and content area knowledge. *Unplugging the Classroom*,
558 pp.13–35.

559 Azis, Y.A. and Hu (2020). Collaborative Digital Storytelling-based Task for EFL Writing Instruction : Outcomes and Perceptions. *The Journal of AsiaTEFL*, 17(2), pp.562–579.

562 Barrett, H. (2005). *Researching and Evaluating Digital Storytelling as a Deep Learning Tool*. [online] Available at: https://electronicportfolios.com/portfolios/SITEStorytelling2006.pdf.

565 Batra, S. (2013). The Psychosocial Development of Children: Implications for Education and Society — Erik Erikson in Context. *Contemporary Education Dialogue*, 10(2), pp.249–278.

568 Batsila, M. and Tsihouridis, C. (2016). "Once upon a Time there was..." A Digital World for Junior High School Learners. *International Journal of Emerging Technologies in Learning (iJET)*, 11(03), p.42.

571 Boydell, K., Gladstone, B.M., Volpe, T., Allemang, B. and Stasiulis, E. (2012). The Production and Dissemination of Knowledge: A Scoping Review of Arts-Based Health Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, [online] 13(1). Available at: https://www.qualitative-research.net/index.php/fqs/article/view/1711 [Accessed 1 Apr. 2021].

576 Chamot, A. (2004). Issues in Language Learning Strategy Research and Teaching. *Electronic Journal of Foreign Language Teaching*, [online] 1(1), pp.14–26. Available at: https://e-flt.nus.edu.sg/wp-content/uploads/2020/09/v1n12004/chamot.pdf.

579 Chung, S.-K. (2006). *The International Journal of Arts Education Digital Story telling in Integrated Arts Education*. [online] Available at: https://ed.arte.gov.tw/uploadfile/Periodical/1320_arts_education41_033050.pdf.

582 Cooke, A., Smith, D. and Booth, A. (2012). Beyond PICO: The SPIDER Tool for Qualitative Evidence Synthesis. *Qualitative Health Research*, 22(10), pp.1435–1443.

584 Erikson, E.H. (1963). *Childhood and society. 2nd ed.* U.S.: Norton.

Gass, S. and Selinker, L. (1986). Language Transfer in Language Learning, eds. Rowley, MA: Newbury House, 1983. Pp. xii + 374. *Studies in Second Language Acquisition*, 8(1), pp.110–112.

Genesee, F. (2021). IS THERE AN OPTIMAL AGE FOR STARTING SECOND LANGUAGE INSTRUCTION? *McGill Journal of Education / Revue des sciences de l'éducation de McGill*, [online] 13(002). Available at: https://mje.mcgill.ca/article/view/7201.

Ghazali, F.A. (2006). *First Language Acquisition Vs Second Language Learning: What Is the Difference?* [online] Available at: https://usir.salford.ac.uk/id/eprint/22469/1/First_Language_Acquisition_Vs_Second_Language_Learning.pdf.

Girmen, P. and Kaya, M.F. (2019). Using the Flipped Classroom Model in the Development of Basic Language Skills and Enriching Activities: Digital Stories and Games. *International Journal of Instruction*, [online] 12(1), pp.555–572. Available at: https://files.eric.ed.gov/fulltext/EJ1201167.pdf [Accessed 24 Mar. 2019].

Gogolin, I. (2012). First-Language and Second-Language Learning. *The Encyclopedia of Diversity in Education*, [online] pp.915–918. Available at: https://www.researchgate.net/publication/275029111_First-Language_and_Second-Language_Learning.

Hett, K. (2012). *Technology-supported literacy in the classroom: Using audio books and Digital Storytelling to enhance literacy instruction.* 3rd ed. Illinois Reading Council Journal, pp.3–13.

Jankowski, K.R.B., Flannelly, K.J. and Flannelly, L.T. (2017). Thet-test: An Influential Inferential Tool in Chaplaincy and Other Healthcare Research. *Journal of Health Care Chaplaincy*, 24(1), pp.30–39.

John, O.P. and Soto, C.J. (2007). The importance of being valid: Reliability and the process of construct validation. *Handbook of Research Methods in Personality Psychology*, [online] pp.461–494. Available at: https://www.colby.edu/psych/wp-content/uploads/sites/50/2013/08/John_Soto_2007.pdf.

Kim, T.K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology*, [online] 68(6), p.540. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4667138/.

Kogila, M., Ibrahim, A.B. and Zulkifli, C.Z. (2020). A Powerful of Digital Storytelling to Support Education and Key Elements from Various Experts. *International Journal of Academic Research in Progressive Education and Development*, 9(2).

Lightbown, P.M. and Spada, N. (2013). *How languages are learned*. 4th ed. Oxford Oxford University Press.

Lim, P.R. and Md Noor, N. (2019). Digital Storytelling as a Creative Teaching Method in Promoting Secondary School Students' Writing Skills. *International Journal of Interactive Mobile Technologies (iJIM)*, 13(07), p.117.

Liu, C.-C., Yang, C.-Y. and Chao, P.-Y. (2019). A longitudinal analysis of student participation in a digital collaborative storytelling activity. *Educational Technology Research and Development*, 67(4), pp.907–929.

Mackey, A. and Gass, S.M. (2016). *Second language research : methodology and design*. New York: Routledge.

PLONSKY, L. and DERRICK, D.J. (2016). A Meta-Analysis of Reliability Coefficients in Second Language Research. *The Modern Language Journal*, 100(2), pp.538–553.

Plonsky, L. and Oswald, F.L. (2011). How to do a Meta-Analysis. *Research Methods in Second Language Acquisition*, pp.275–295.

Rahman, S. (2016). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language "Testing and Assessment" Research: A literature review. *Journal of Education and Learning*, [online] 6(1), p.102. Available at: https://files.eric.ed.gov/fulltext/EJ1120221.pdf.

Robin, B.R. (2008). Digital Storytelling: A Powerful Technology Tool for the 21st Century Classroom. *Theory Into Practice*, 47(3), pp.220–228.

641 Robin, B.R. (2014). *The educational uses of digital storytelling.* [online] University of

642 Houston. Available at: https://www.researchgate.net/publication/228342171.

643 Rubino, I., Barberis, C. and Malnati, G. (2018). Exploring the values of writing

644 collaboratively through a digital storytelling platform: a mixed-methods analysis of

645 users' participation, perspectives and practices. *Interactive Learning Environments*,

646 26(7), pp.882–894.

647 Sarkis-Onofre, R., Catalá-López, F., Aromataris, E. and Lockwood, C. (2021). How to

648 properly use the PRISMA Statement. *Systematic Reviews*, 10(1).

649 Shadiev, R. and Yang, M. (2008). *Sustainability*. [online] Mdpi.com. Available at:

650 https://www.mdpi.com/journal/sustainability.

651 Shin, B.-J. and Park, H.-S. (2008). The Effect of Digital Storytelling Type on the

652 Learner's Fun and Comprehension in Virtual Reality. *Journal of The Korean*

653 *Association of Information Education*, [online] 12(4), pp.417–425. Available at:

654 http://www.koreascience.or.kr/article/JAKO200832056741617.page [Accessed 4 Dec.

655 2021].

656 Signes, C. (2008). *PRACTICAL USES OF DIGITAL STORYTELLING*. [online]

657 Available   at:

658 https://www.uv.es/gregoric/DIGITALSTORYTELLING/DS_files/DST_15_ene_08_f

659 inal.pdf.

660 Simpson, R., Dam, A.V. and Becker, S. (2005). *Next-Generation Educational*

661 *Software: Why We Need It and a Research Agenda for Getting It*. [online]

662 er.educause.edu. Available at: https://er.educause.edu/articles/2005/1/nextgeneration-

663 educational-software-why-we-need-it-and-a-research-agenda-for-getting-it [Accessed

664 4 Dec. 2021].

665 Sukovic, S. (2014). iTell: Transliteracy and Digital Storytelling. *Australian Academic*

666 *& Research Libraries*, 45(3), pp.205–229.

667 Tanrıkulu, F. (2020a). The Effect of L2 Listening Texts Adapted to the Digital Story

668 on the Listening Lesson. *Turkish Online Journal of Distance Education*, pp.1–18.

669 Tanrıkulu, F. (2020b). Students' perceptions about the effects of collaborative digital
670 storytelling on writing skills. *Computer Assisted Language Learning*, pp.1–16.

671 Tarone, E., Swain, M. and Fathman, A. (1976). Some Limitations to the Classroom
672 Applications of Current Second Language Acquisition Research. *TESOL Quarterly*,
673 10(1), p.19.

674 Wehrwein, E.A., Lujan, H.L. and DiCarlo, S.E. (2007). Gender differences in learning
675 style preferences among undergraduate physiology students. *Advances in Physiology*
676 *Education*, 31(2), pp.153–157.

677 Xu, Y., Park, H. and Baek, Y. (2011). A New Approach Toward Digital Storytelling:
678 An Activity Focused on Writing Self-efficacy in a Virtual Learning Environment.
679 *Educational Technology & Society*, [online] 14(4), pp.1176–3647. Available at:
680 https://2learner.edu.vn/~longld/References%20for%20TeachingMethod&EduTechnol
681 ogy%20-%20Tai%20lieu%20PPDH%20&%20Cong%20Nghe%20Day%20Hoc/(Pap
682 er)%20-%20Bai%20viet%20tham%20khao%20-%20Digital%20Storytelling%20(DS
683 T)/DO-A%20New%20Approach%20Toward%20Digital%20Storytelling.pdf
684 [Accessed 4 Dec. 2021].

685 Yamaç, A. and Ulusoy, M. (2016). The effect of digital storytelling in improving the
686 third graders' writing skills. *International Electronic Journal of Elementary Education*,
687 [online]   9(1),   pp.59–86.   Available   at:
688 https://www.iejee.com/index.php/IEJEE/article/view/145 [Accessed 4 Dec. 2021].