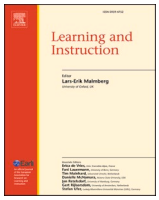
Learning and Instruction 85 (2023) 101749

Contents lists available at ScienceDirect

Learning and Instruction

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



Efficacy of digital picture book enhancements grounded in multimedia learning principles: Dependent on age?

Xuan Li [a,b,*], Adriana G. Bus [c]

[a] University of Houston, 3657 Cullen Blvd., Houston, TX, 77204, USA
[b] Navajo Technical University, P.O. Box 849, Crownpoint, NM, USA
[c] University of Stavanger, 4036 Stavanger, P.O. box 8600, Norway

ARTICLE INFO

Keywords:
Book reading
Digital picture books
Multimedia learning
Story comprehension
Vocabulary development

ABSTRACT

We designed an experiment with stratified randomization to investigate the effects of visual and auditory enhancements in digital picture books on comprehension and incidental word learning. Participants were 183 children aged 3, 4, and 5 years (81 girls and 102 boys) from childcare centers and schools in the Southwest USA. We contrasted the still-image condition (an onscreen picture book with a voice-over reading the narrative aloud) with three enhanced conditions: a digital book that included auditory and visual enhancements, only auditory enhancement, or only visual enhancement. All participants watched and listened thrice to the researcher-assigned digital picture book version within three weeks. The posttests assessed children's story comprehension and book-based vocabulary. The visual and auditory enhancements benefited children's story comprehension and book-based vocabulary. However, a version with auditory and visual enhancements was less beneficial for comprehension than versions with single (auditory or visual) enhancements, particularly in the youngest group.

1. Introduction

Stories are compelling to young children and stimulate their cognitive development. They also open a window into other people's emotions and behaviors, thus incentivizing their ability to understand social situations and emotions (Wilson, 2014). Children as young as three years old start watching and listening to stories on tablets or smartphones in this digital era. As they often do so independently, we need to find out which digital enhancements can support their comprehension and, thus, their interest and engagement (Furenes, Kucirkova, & Bus, 2021). Evidence shows that children 4- or 5-year-old can benefit from dynamic pictures if only they are designed in a way that guides children's visual attention (Sarı, Başal, Takacs, & Bus, 2019; Sun, Loh, & Roberts, 2019; Verhallen, Bus, & de Jong, 2006). Music and sounds may also be helpful because they may make children in this age range attentive to story events and more aware of the character's moods. However, so far, there are hardly any studies testing whether the younger readers of digital stories respond likewise. Therefore, we included children as young as three years in this experiment and studied whether they, just as the 4- to 5-year-olds, benefit from an enhanced digital picture book.

1.1. Visual and auditory digital enhancements in digital books

Due to the sophisticated words and complex grammar in picture books, rare in daily conversations, young children often experience problems understanding narratives (Montag, Jones, & Smith, 2015). Luckily, pictures in picture books mostly tell a similar story, thus sustaining children's comprehension despite the complex language (Brookshire, Scharff, & Moses, 2002; Takacs & Bus, 2018). The picture book format thus aligns with the multimedia learning principle that people learn better from words and pictures than words alone (Mayer, 2014). Furthermore, there is growing evidence that young children understand technology-enhanced picture books with dynamic pictures, music, and sounds even better than books with still images. Studies with Dutch (Verhallen et al., 2006), Turkish (Sarı et al., 2019), and Mandarin-speaking children (Sun et al., 2019) report effects of half a standard deviation or more.

Such enhancements can help children connect the pictures to the narrative and thus improve story comprehension (e.g., Bus & Anstadt, 2021; Christ, Wang, Chiu, & Strekalova-Hughes, 2019; Furenes et al., 2021; Miller & Warschauer, 2014). Details in motion or zooming in on an illustration direct children's attention to parts of the pictures

* Corresponding author. Navajo Technical University, P.O. Box 849, Crownpoint, NM, USA.
E-mail addresses: xli@navajotech.edu (X. Li), Adriana.g.bus@uis.no (A.G. Bus).

https://doi.org/10.1016/j.learninstruc.2023.101749
Received 23 August 2021; Received in revised form 1 December 2022; Accepted 19 February 2023
Available online 2 March 2023

representing what the narrative tells simultaneously, thus enhancing the effects pictures have; in the literature about multimedia learning, this effect is known as *the temporal contiguity principle* (Mayer & Sims, 1994). For instance, one of the spreads in our target book, Little Kangaroo (Van Genechten, 2005), shows Mama Kangaroo and Little Kangaroo looking at a herd of giraffes running across the plain. In the digital version, the herd moves, and a virtual camera zooms in on it, thus making more tangible how overwhelming the running giraffes are for little Kangaroo, as the narrator simultaneously explains.

Second, visual enhancements can encourage self-explanation, one of the multimedia learning principles (Wylie & Chi, 2014). For example, the first scene of the Little Kangaroo story shows Mama Kangaroo jumping cumbersomely with a heavy stomach while the narrator explains, "Mommy Kangaroo has a problem …." This combination of picture and narration makes the reader wonder what Mommy's problem is and why she looks so unhappy, thus stimulating children to generate explanations for Mama Kangaroo's suffering. *Self-explanation* is more encouraged in the digital version of Little Kangaroo than in the still-image version, where Little Kangaroo is visible on the first spread, thus betraying from the start of the narrative why mama Kangaroo is suffering.

Little Kangaroo also includes music and sound as multimedia effects. For example, the scene with the giraffes contains the abashing sound of a herd running by. This sound adds to the overwhelming impression of the running herd on the story characters (Brandt, Gebrian, & Slevc, 2012; Furnham & Stephenson, 2007). We could consider such additions a form of *embodiment* (Kucirkova, 2019; Mangen, Hoel, & Moser, 2018; Mangen & Pirhonen, 2022). Children have an auditory experience similar to the story characters, which may help them empathize and better comprehend. At another location in the digital picture book, when Mama Kangaroo says, "look how beautiful the butterflies flutter from flower to flower," peaceful and soft background music simultaneously initiates the kind of feelings the scenery elicits in the story characters. Sounds can also help children understand the negative emotions of the story character. For example, the sharp chirping birds may provoke the same irritation in the reader as little Kangaroo feels and thus explain why she is annoyed by the birds and wants to stay in her Mama's pouch. Moreover, in the future, there may be technical solutions to allow other sorts of additions, such as scents, that added to the scenery might accomplish similar effects as music and sounds (Kucirkova, 2019).

1.2. Possible negative effects of digital enhancements

For app designers, it is tempting to add visual and auditory features, as illustrated above. However, we should wonder whether adding digital features in all possible combinations is effective. Children might be overwhelmed when they need to process several sources of additional information. *The split-attention principle* predicts that it is essential to avoid formats that require inexperienced readers to split their attention between and mentally integrate multiple sources (Mayer, 2014). With an increasing number of features to be processed simultaneously in working memory, inexperienced readers could be most at-risk (Chandler & Sweller, 1996). The dynamic visuals and the music and sound in Little Kangaroo may each help comprehend the story, but simultaneously available, they might quickly overload working memory. Both are relevant for story understanding but, as argued above, might contribute to different story elements. Visual enhancements often aim at reconstructing the character's actions. For example, we see Little Kangaroo disappear in her mom's pouch. At the same time, the peaceful and soft music focuses on her emotions (feeling safe and secure) to explain the action. In addition to processing both enhancements, the reader needs to mentally coordinate those, which requires additional working memory resources (Ayres & Sweller, 2014). In other words, attending to multiple digital additions, not implying the same message, and integrating those may result in cognitive overload, especially when children are younger and inexperienced in comprehending a story's actions and the character's emotions.

Auditory enhancements may also have adverse effects because they use the same modality as the voice-over reading the narrative aloud. Due to the sounds and music, children may hear the words in the oral narrative less well and may fail to store unknown words' pronunciation in memory, complicating rather than facilitating children's learning of new vocabulary (Lehmann & Seufert, 2017; Nguyen & Grahn, 2017). The *modality principle* in multimedia theory predicts that auditory features might hurt children's ability to hear the narrative and learn from the activity (Sarı et al., 2019; Smeets, van Dijken, & Bus, 2014). For instance, in a group aged 4–6 years, Sarı et al. (2019), using Little Kangaroo as one of the stimulus books, found a moderately strong negative effect from combined background sounds and music on book-based vocabulary ($d = -0.56$). We hypothesize that both information sources may compete in the working memory when children need to process music and sounds simultaneously with spoken words through the auditory channel. Processing spoken words may require most auditory channel resources, especially when children's language is less advanced. Consequently, interpreting music and sounds may go at the expense of interpreting the spoken language in younger children. Therefore, in the current study, investigating the effects of music and sounds on story comprehension and learning vocabulary, we tested the impact of music and sounds in a group including younger children than the previous studies primarily targeting 4- to 5-year-olds.

1.3. This Study's hypotheses

This study compared enhanced versions of Little Kangaroo with a version without digital enhancements. Our primary interest was story comprehension as an outcome measure. Some story understanding is fundamental for deriving pleasure from book-reading, thereby developing a habit of book-reading. However, a better grasp may also contribute to incidental word learning because there thus is a richer context for deriving word meanings (Penno, Wilkinson, & Moore, 2002). Therefore, visual and auditory enhancements may also support incidental word learning.

Our final goal is to provide scientifically proven guidelines for app designers and help parents and teachers select and implement picture book apps that can promote even the youngest children's story comprehension and vocabulary learning. To this end, we chose a design that disentangled the impact of visual and auditory additions to test the following hypotheses.

1. First, digital books with enhancements aligning with temporal contiguity, self-explanation, or embodiment benefit story comprehension and word learning more than digital books with only still images.

However, there can be adverse effects of adding auditory and visual enhancements simultaneously.

2. Following the split-attention principle, auditory and visual enhancements simultaneously can decrease story comprehension and word learning. The need to split attention over two additional sources of information and integrate those may particularly harm the youngest children, for whom each process takes relatively more effort.
3. In line with the modality principle, auditory enhancements may interfere with processing the oral narrative, especially in younger children, where processes take much effort; as a result, visual enhancements may be more beneficial than the conditions with auditory enhancements, particularly for word learning.

2. Method

2.1. Design

We carried out an experiment in which children read different versions of a digital story, *Little Kangaroo*, three times during the intervention stage. Before and after the intervention, we tested language skills, and after the intervention, language skills and story comprehension. For the intervention, we created four digital versions of the story. All four versions shared the same illustrations and the same voice-over reading the story text aloud but differed in the digital enhancements.

1. ***still-images***: a voice-over and per spread an overview illustration onscreen just as in the paper book version (see the first column in Table 2) without visual or auditory digital enhancements,
2. ***auditory enhancements***: a voice-over and per spread an overview illustration onscreen just as in the paper book version (see the first column in Table 2) with auditory enhancements as described in the last column of Table 2,
3. ***visual enhancements***: a voice-over and per spread visual enhancements as illustrated under screenshots in Table 2 and described in the fourth column, and
4. ***auditory and visual enhancements***: a voice-over and per-spread visual enhancements as illustrated under screenshots in Table 2 and described in the fourth column and auditory enhancements as described in the last column of Table 2.

2.2. Participants

The sample from an urban area in the Southwest of the USA included 183 children aged three years (n = 79), four years (n = 71), and five years (n = 33). All children spoke English; they were from diverse ethnic backgrounds (31% Asian, 16% Black, 23% Latino, and 29% White); the mothers' education varied from high school (11%) to a Master's Degree or higher (50%); the sample included 81 girls and 102 boys.

The researchers divided the children whose parents had returned the consent form over the four conditions. Since centers/schools attracted a different population and the number of eligible children per center/school was very different – ranging from 5 to 43 - we divided children equally over the four conditions per center/school so that conditions and centers/schools were not confounded. Furthermore, we assigned the children with the restriction that each condition included a similar share of three-, four- and five-year-olds and about the same percentage of boys. Thus dividing the children over the four conditions, we only knew the child's ID number (allocated to each completed consent form), which center/school the child visited, and the child's age and sex. For the rest, we were blind to child characteristics and assigned the listed ID numbers alternately where there was a choice. We thus composed four groups, considering the restrictions above. For 80% power detecting a Cohen's f of 0.25, with a significance level of 0.05, we needed a minimum of 45 participants in each condition (Faul, Erdfelder, Buchner, & Lang, 2009). The final sample did not fully satisfy this requirement as two groups were slightly smaller, including 43 and 36 children (see Table 1). However, as shown in Table 1, the four groups were similar regarding the number of 3-, 4-, and 5-year-olds and sex. Neither did they differ in ethnicity, the mother's education, and PPVT.

2.3. Materials

2.3.1. The enhanced picture book

From previous studies, we know that 3- to 5-year-old enjoy Little Kangaroo - a story about a Little Kangaroo who does not want to leave her mother's comfortable pouch. The book includes 12 spreads with 43 words on each (SD = 9). In all, the book text contains 511 words. The four digital versions of the book did not have the narrative in print, but there was a female reader's voice-over. In all versions, the pages turned automatically after the voice-over finished reading a page. A professional translator translated the original Dutch story Little Kangaroo into English, and a native English speaker recorded the spoken text.

Table 2 shows the visual and auditory enhancements added to the first two spreads representing learning principles as explained in the introduction to this article. In addition, the table shows screenshots from the film-like visualizations to illustrate the visual enhancements. For example, the first spread shows Mama Kangaroo hopping cumbersomely to emphasize that she is exhausted. Then little Kangaroo suddenly appears from her pouch, and the narrative explains that Little Kangaroo has become too heavy to carry around all day. There are also auditory enhancements. For instance, while the narrator stresses that Mama Kangaroo is exhausted, we hear heavy bongs emphasizing how much effort it takes to hop. While the narrator explains that Little Kangaroo enjoys sitting in her Mama's soft and warm pouch, soothing and pleasant music is audible, emphasizing how comfortable and secure she feels. Furthermore, the environmental sounds at some locations may help concretize the story events. For instance, we cannot see but hear the sound of Little Kangaroo drinking milk in spread two. The music and background sound frequently coincide with the voice-over.

Table 1
Characteristics of participants per condition.

n	Visual + Auditory enhancements	Visual enhancements	Auditory enhancements	Still images	Total	Statistics
	58	46	43	36	183	
Mean age in months	51.40(9.47)	49.59(9.18)	49.91(9.09)	50.75(8.95)	50.46(9.16)	$F(3, 179) = .47, p = .70$
Participants per age: 3, 4, 5 years	24/22/12	22/16/8	18/17/8	14/16/6	78/71/34	$\chi^2 = 1.5$, p = .96
Boys	50%	41%	35%	50%	44%	$\chi^2 = 2.95, p = .40$
Asian	19	10	14	14	57	$\chi^2 = 5.44, p = .79$
Black	10	7	7	6	30	
Latino	14	12	11	5	42	
White	14	17	11	11	53	
High school	9	5	5	1	20	$\chi^2 = 10.04, p = .61$
Associate Degree	2	5	5	6	18	
Bachelor	18	10	10	12	50	
Master's	14	13	11	7	45	
Doctorate	13	12	11	9	45	
PPVT(raw)	66.52(28.11)	66.15(26.88)	65.23(25.56)	63.33 (25.13)	65.50 (26.45)	$F(3, 179) = .08, p = .97$

Notes. PPVT = Peabody Picture Vocabulary Test.We missed information about parent education for five children and ethnic information for one parent.

Table 2
The visual and auditory enhancements added to the first two spreads of the Little Kangaroo story.

Still image	Voice-over	Screenshots	Visual enhancement (film-like)	Auditory enhancement
	Mommy Kangaroo had a problem. And that problem was hiding in her pouch. It was large and heavy,		We see Mama Kangaroo hopping several times. She is jumping clumsily.	While the voice-over reads, the reader hears heavy bongs every time Mama Kangaroo lands on the ground.
	But also really sweet. And it fidgeted with her coat all day.		After a while, she stands still and leans a bit backward. Finally, we see her pouch move.	While the voice-over reads the story, we hear Mama Kangaroo sigh heavily.
	Little Kangaroo had grown too big. It is about time, Mammy thought, that she hops through life on her own two legs.		Little Kangaroo emerges from the pouch, and the camera slowly zooms in on Little Kangaroo.	While the voice-over reads, we hear soft and pleasant music and a swoosh when Little Kangaroo appears.
	But Little Kangaroo did not want to hop. Her Mommy's pouch was nice and soft. And every day, Little Kangaroo got milk.		We can only see Little Kangaroo's ears moving.	We hear the sound of drinking. Simultaneously, we hear pleasant music and the voice-over.
	And a wash.		Mama Kangaroo's tongue moves over Little Kangaroo.	We hear the licking sound while we hear pleasant music and the voice-over.
	And ….the pouch was handy. She did not have to jump everywhere herself.		Mama Kangaroo is holding the Little Kangaroo to see the world while we see her moving forward.	The background music is pleasant and soft. There is the sound of Mama Kangaroo hopping and the voice-over.

Note. Copyright 2007 by *Het Woeste Woud*, Groningen, the Netherlands.

2.4. Measures

2.4.1. Picture vocabulary test

The participating children completed a pre-test, the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007). For each word, children chose one of four pictures. For example, the researcher asked the child, "please point to the picture that shows 'baby.'" The test consists of 228 items equally distributed across 19 item sets. Each item set contains 12 items of increasing difficulty. The examiner (first author) started with the set prescribed for the age, returning to a preceding set if the child made one or more errors. The examiner continued with the following sets until the child made eight or more errors within a set (Dunn & Dunn, 2007). The test-retest reliabilities yielded correlations between 0.92 and 0.96 (very high), and the average internal consistency reliability was 0.95. In the analyses, we used the raw PPVT scores. The administration took approximately 10–15 min per child.

Story Comprehension Test We designed a story comprehension test based on Paris and Paris's (2003) Narrative Comprehension Instrument. It included five explicit and five implicit questions. The explicit questions related to.

1. The story characters (Who are the two kangaroos in this picture?)
2. Settings (Where does the story happen?)
3. Initiating event (Tell me what happens here, pointing at the exhausted Mama Kangaroo)
4. Problem (Look at this picture – Little Kangaroo disappearing in the pouch; if you were telling someone this story, what would you say? Why did this happen?)
5. Outcome resolution (At the picture of the Little Kangaroo running after another little kangaroo: What happened here? Why did this happen?).

The implicit questions related to.

1. Feelings (Tell me what the Little Kangaroo feels in this picture. Why do you think so?)
2. Causal inference (Why did the Little Kangaroo hop back into the Mommy Kangaroo's pouch?)
3. Dialogue (What do you think the Mommy and Little Kangaroo are saying here? Why would they say that?)
4. Prediction (This is the last picture in the story. What do you think happens next? Why do you think so?)

5. Theme (If you were the Mama Kangaroo, what would you tell the Little Kangaroo when she does not want to hop out of the pouch? Why would you tell her that?).

In coding the answers, we followed the Paris and Paris coding scheme. When a child's response to a question addressed the primary story event, the child earned two points. If a child only provided part of the story event to a question, the child earned one point. If a child's response was inappropriate or failed to identify the story event, the child earned no points. The average intraclass correlation between the scores of two independent coders was 0.96 [95%CI: 0.90; 0.98]. Cronbach's Alpha reliability equaled 0.76.

2.4.2. Book-based vocabulary test

For the vocabulary posttests, 20 words, including some phrases, were selected from the story, encompassing two nouns, eleven verbs, and seven adjectives. The tests included the following words: fluttering, pouch, hopping, swing, whistling, exhausted, empty, safe, hop in, noisy, take out, run across, carry, heavy, wet, chirping, dangerous, fidget, pride, and stay close. We checked the list of Preschool Non-Disabled Children's Vocabulary provided by the College of Education and Human Sciences, University of Nebraska-Lincoln (2019, November 20). Our selected words have a low frequency or do not appear on the list. According to the English Lexicon Project Web Site focusing on adults (Cognitive Psychology Laboratory, Washington University in St. Louis, 2019, October 19), 13 out of the 20 words have a low frequency or do not appear on the list, while seven words indicate high frequency.

Using pictures from the target story, we designed a test of *receptive* word knowledge. The researcher asked the child to point to the image that shows the word spoken aloud by the researcher, each time having a choice of four pictures. For every correct target word, the child received one point. We also tested *expressive* knowledge of the same words. To that end, the examiner showed the child a picture from the computer screen and read an incomplete sentence aloud. For example, "Mama Kangaroo has been carrying the baby kangaroo for a whole day; she is … ?" Then, the examiner asked the child to complete the sentence with the target vocabulary word "exhausted." If the children provided a synonym, the examiner asked whether they knew another word. We did not award synonyms. For every correct target word, the child received one point. The average intraclass correlation between the scores of two independent coders was 0.93 [95%CI: 0.89; 0.98].

As scores on expressive vocabulary in the youngest age group included many zero scores, we created one scale including receptive and expressive items. With the ltm Package, we calculated that Cronbach's Alpha equaled 0.83. We used the sum score in all further analyses to indicate children's book-based vocabulary.

2.5. Procedure

2.5.1. Data collection methods and research design

After obtaining approval for the study from six childcare center directors and three school principals in the university's neighborhood and ethical approval for this study from the Internal Review Board at the university of Houston, the researcher (the first author) approached parents of all children in the age range of 3–5 for participation. They received a written explanation of the study and, if they wished, they could contact the researchers for more information. About fifty percent of the parents returned the form providing consent. There was no need to exclude children due to severe cognitive or physical problems. In addition, the researcher obtained child assent before the research activities started. The researcher explained that the children would repeatedly listen and watch a story and answer questions about the story afterward. They were free to withdraw or stop participating at any time.

The researcher pre-tested each child with the PPVT-4 in a quiet room at the care center or school during regular school hours where only the child and the researcher were present. In the same session, the researcher checked whether children were unfamiliar with the Little Kangaroo story by asking whether they knew the characters while looking at the book's cover page. None of the children was familiar with the story.

One week later, the intervention began. Each child listened and watched a version of Little Kangaroo on three different days. So, each participant read the story three times. Previous studies (e.g., Verhallen et al., 2006) show that three to four repetitions are indispensable to comprehending a story and learning new words for young children. There were, on average, 5.5 days between the first and the second session ($SD = 0.48$) and 5.5 days between the second and third sessions ($SD = 0.48$). These sessions took place in a quiet room during regular school hours. Only the participants and the researcher were present. The researcher brought the children to the room and returned them to their classrooms when the session was over. Teachers or other adults were not allowed to enter the intervention room during the sessions.

In the first two sessions, four children from different conditions were present. Each child read the story independently. The children received headphones to avoid interaction and distraction from each other. The researcher started the program for each child and explained that they would see and hear a story. The researcher provided help if the program caused problems. There was no opportunity to respond to or discuss the story. If a child tried to look at the neighbor's screen, the researcher directed the child's attention to his or her computer screen. The researcher gently calmed the children when they started talking or made noise while watching the story. As the story lasted slightly longer than 5 min, the complete session did not take more than 10 min.

Children listened a third time to the story during the third intervention session. Only one child was allowed at this session because the researcher assessed book-based vocabulary and story comprehension immediately after the reading. This session lasted approximately 15 min per child. Unfortunately, four children did not want to participate, and we excluded them.

2.6. Data diagnostics and analytical strategies

We applied a Shapiro-Wilk test and visually inspected histograms, Q-Q plots, and boxplots to check whether the scores on PPVT, story comprehension and vocabulary were normally distributed for the four experimental conditions. Both dependent measures satisfied all conditions for carrying out a multiple regression. Table 1 describes the means and standard deviations of the background variables (sex, age, PPVT, and parent education) per condition. None of them revealed a significant difference between the four conditions. In Table 3, the high intercorrelations between PPVT, book-based vocabulary, and story comprehension are noteworthy ($r \geq 0.62$).

Next, we inspected whether Cook's distance, tolerance, and VIF satisfied conditions for multiple regression for story comprehension and vocabulary. We also examined the scatterplot of standardized predicted values versus standardized residuals to show that the residuals were approximately normally distributed. Finally, PPVT raw score was centered.

We regressed the two posttest outcome variables (vocabulary and story comprehension) on age in years, sex, PPVT (mean-centered), and conditions. We used a linear mixed-effect model because the children

Table 3
Correlations between the dependent variables and background variables.

	1	2	3	4	5
1. Sex	–				
2. Age	.17*	–			
3. PPVT	.04	.66**	–		
4. Parent education	−.02	.05	.27**	–	
5. Book-based vocabular	.11	.49**	. 74**	.28**	–
6. Story comprehension	−.04	.58**	.62**	.16*	.71**

*$p < .05$. **$p < .001$.

were nested within centers or schools. We used the *lme4* Package in r (Bates, Maechler, Bolker, & Walker, 2015). Using Helmert coding, we created three orthogonal contrasts between conditions aligned with the three hypotheses.

1. The three enhanced conditions versus the still-image condition,
2. The condition with combined visual and auditory enhancements versus the conditions with visual or auditory enhancements alone, and
3. Visual versus auditory enhancements.

As children were grouped within centers, we first inspected the random effects of centers/schools (the intraclass correlation coefficient). Then, where fixed effects had a significant impact, we calculated Hedges'*g* as an indicator for the effect size. For computing the effect sizes, we used the *esc* package in R (Lüdecke, 2019).

3. Results

3.1. Random effects

We checked clustering in our data by analyzing the intercept-only model – a baseline model with no predictors. We found variations in comprehension and vocabulary associated with differences between centers/schools. The ICCs, or Intraclass Correlation Coefficients, were 15% and 29%. So, students in the same centers/schools were more alike on the two dependent variables evidencing a random intercept. The correlations between the fixed intercept and fixed slopes for the predictors were low. So, there was no evidence for random slopes.

3.2. Background variables

As a first step, we regressed the two posttest outcome variables (vocabulary and story comprehension) on background variables: parent education, sex, age in years, and PPVT. We did not include parent education in the final regression models because it did not affect the dependent variables. However, sex approached significance for comprehension and was therefore included. PPVT had a significant effect on story comprehension ($t(137) = 6.19$, $p < .001$) and vocabulary ($t(141) = 11.05$, $p < .001$). Furthermore, age significantly affected story comprehension ($t(172) = 2.78$, $p = .006$).

3.3. Enhancements vs. still-image condition (hypothesis 1)

After the background variables, we entered in a second step the three contrasts and, next, the interaction between age in years and those contrasts. The results in Table 4 show that the contrast enhancements vs. still images caused an effect on comprehension and vocabulary. The enhanced versions scored 0.30 and 0.22 standard deviations higher than the still image version. However, concerning vocabulary, not all age groups benefited from the enhanced conditions to the same extent, $t(169) = -2.75$, $p = .007$. The estimated marginal means per age group (Table 5) show that the enhanced conditions were most profitable for the three-year-old but much less for the four-year-old and no longer in favor of the enhanced conditions in the five-year-old.

3.4. Combined vs. single enhancements (hypothesis 2)

On comprehension, a single enhancement – visual or auditory – revealed higher scores ($M = 9.45$, $SE = 0.50$) than both enhancements simultaneously present ($M = 8.72$, $SE = 0.56$), $t(169) = -2.38$, $p = .018$. The effect size equaled 0.16. There was not such an effect on vocabulary, $t(168) = -0.84$, $p = .40$. For story comprehension, the interaction between age in years and the contrast combined enhancements vs. a single enhancement was also significant, $t(169) = 2.15$, $p = .033$; see Table 4. Table 5 shows that the three-year-old benefited more from a single enhancement than both simultaneously. On the other hand, four- or five-year-old understood slightly more when both enhancements were available, even though not significantly.

Table 4
Results of Regressing Vocabulary and Story Comprehension on Sex, Age in years, PPVT Score (Centered), and Experimental Conditions Using a Mixed Model.

	Vocabulary ($N = 178$)		Comprehension ($N = 178$)	
	Intercept-only model	Model with predictors	Intercept-only model	Model with predictors
Fixed Effects[1]				
Intercept	19.99(1.26) ***	18.10(2.28) ***	8.71(0.70) ***	−1.22(3.43)
Sex		−0.34(0.59)		−0.82(0.51)
Age in years		−0.65(0.58)		2.41(0.86) **
PPVT		0.17(0.02) ***		0.08(0.01) ***
-Contrasts				
Enhancements vs. still image		3.19(0.98) **		7.75(4.10)*
Combined vs. single		−0.92(1.10)		−7.89(3.31) *
Visual vs. auditory		−0.31(2.03)		0.95(3.50)
*-Interactions Age * contrasts*				
Age * Enhancements vs. still image		−0.71(0.26) **		−1.44(1.03)
Age * Combined vs. single		0.05(0.28)		1.85(0.86)*
Age * Visual vs. auditory		0.03(0.53)		−0.22(0.92)
Random effects[2]				
Intercept	12.00(3.46)	1.00(1.00)	3.20(1.79)	0.47(0.69)
Variation	29.08(5.39)	13.97(3.74)	17.81(4.22)	10.35(3.22)

[1] estimate and, in brackets, standard errors. Significance codes: '***' 0.001, '**' 0.01, '*' 0.05.
[2] variance and, in brackets, standard deviations.

3.5. Visual vs. auditory enhancements (hypothesis 3)

The third contrast between auditory and visual enhancements did not cause significant effects; see Table 4. In the conditions with visual enhancements ($M_{visual} = 9.54$, $SE = 0.63$) and auditory enhancements ($M_{auditory} = 9.34$, $SD = 0.63$), children scored similarly on story comprehension, $t(169) = 0.27$, $p = .79$). Likewise, scores on vocabulary were similar for visual ($M_{visual} = 21.1$, $SE = 0.78$) and auditory enhancements ($M_{auditory} = 21.4$, $SE = 0.79$), $t(167) = -0.15$, $p = .88$).

4. Discussion

4.1. Main findings

The digital book Little Kangaroo has enhancements designed to support *story comprehension* in unique ways: visual enhancements help synchronize the narration with illustrations and stimulate self-explanations, and environmental sounds and background music can highlight the story characters' emotions or feelings such as pleasure, stress, or anxiety thus increasing embodiment and empathy with the story characters. The current findings are consistent with previous studies showing that these ways of enhancing digital books help children understand the story plot (cf. Sarı et al., 2019; Sun et al., 2019; Verhallen et al., 2006). In addition, the enhancements also promote incidental *word learning*.

However, there are also new findings. One remarkable result is that combining visual and auditory enhancements is less effective than one enhancement at a time, either visual or auditory. This result particularly applies to story comprehension in the youngest group. The three-year-olds' comprehension benefits more from one enhancement at a time

Table 5
Estimated marginal means (and standard errors) for vocabulary and comprehension per age group and hypothesis.

		n(3/4/5)	Three years	Four years	Five years	Pooled
Vocabulary *(Maximum score = 40)*						
Hypothesis						
1	*Enhanced* picture book	63/53/28	20.20 (0.75)	21.3 (0.72)	19.20 (1.03)	20.20 (0.52)
	Still image	13/16/5	16.20 (1.23)	19.3 (1.08)	21.20 (1.88)	18.9 (0.85)
Hypothesis						
2	*Combined* (visual *and* auditory enhancements)	22/22/12	18.00 (0.99)	21.30 (0.93)	16.8 (1.35)	18.70 (0.66)
	Single (visual *or* auditory enhancements)	41/31/16	21.40 (0.82)	21.30 (0.85)	20.90 (1.35)	21.20 (0.60)
Hypothesis						
3	*Visual* enhancements	23/15/8	21.30 (1.01)	20.90 (1.14)	21.10 (1.57)	21.10 (0.87)
	Auditory enhancements	18/16/8	21.50 (1.11)	21.70 (1.11)	21.00 (1.51)	21.04 (0.79)
Comprehension *(Maximum score = 20)*						
Hypothesis						
1	*Enhanced* picture book	61/55/28	7.55 (0.66)	9.73 (0.62)	10.37 (0.90)	9.21 (0.46)
	Still image	13/16/5	5.20 (1.07)	8.20 (0.94)	9.52 (1.62)	7.64 (0.74)
Hypothesis						
2	*Combined* (visual *and* auditory enhancements)	21/22/12	5.49 (0.86)	10.17 (0.80)	10.51 (1.15)	8.72 (0.56)
	Single (visual *or* auditory enhancements)	40/33/16	8.63 (0.71)	9.44 (0.71)	10.27 (0.99)	9.45 (0.50)
Hypothesis						
3	*Visual* enhancements	19/16/8	9.99 (0.91)	9.08 (0.92)	10.56 (1.28)	9.54 (0.63)
	Auditory enhancements	21/17/8	8.36 (0.87)	8.36 (0.87)	9.90 (1.32)	9.34 (0.63)

than from both simultaneously, while four- and five-year-olds tend to benefit more from both. Concerning word learning, there were no adverse effects of combined enhancements.

We expected other adverse effects from sounds and music, interfering with word learning. For example, as sounds and music use the same modality as the voice-over, children may fail to store the words' pronunciation in memory - an essential step in learning new words. However, surprisingly, in this study, visual enhancements are as helpful as auditory enhancements. In this respect, the current findings differ from Sarı and colleagues' (2019) study, showing that music and sounds negatively influence children's vocabulary learning while visual enhancements have a positive effect.

4.2. Theoretical implications

Our findings support the hypothesis that technology-enabled enhancements, as in Little Kangaroo, help young learners build mental representations of a story. Furthermore, we find positive effects on incidental word learning from repeatedly listening to a digitally enhanced story. The findings corroborate the relevance of multimedia learning principles in digital picture books. The intuitively created visual enhancements align with temporal contiguity and the importance of stimulating self-explanation (Mayer, 2014). The music and sounds align with the principles of embodiment and may thus stimulate children to empathize with the story characters (Kucirkova, 2019; Mangen et al., 2018; Mangen & Pirhonen, 2022).

A unique result is that we find evidence for cognitive overload caused by several additional sources simultaneously: auditory and visual additional information is less helpful than one source at a time. This result aligns with the theory that multiple sources of information appearing simultaneously but not implying the same message can be challenging (e.g., Chandler & Sweller, 1996). Apart from processing each enhancement separately, children need to coordinate information elicited by each enhancement, which adds cognitive load to their memory and may interfere with apprehending the information. Attuning all sources of additional info, children experience an extra burden on their working memory. Therefore, eliminating one source (visual or auditory enhancements) reinforces the youngest children's comprehension.

Especially the three-year-olds' comprehension suffers from both enhancements simultaneously because processing and integrating all information can be too much for them. Older children may experience the same adverse effects on comprehension of more complex stories than Little Kangaroo. We did not find that children's incidental word learning suffers from combined visual and auditory enhancements. However, if Penno et al.'s theory that understanding the story helps incidental word learning applies, we may expect that simultaneously processing different enhancements can also strain word learning.

Unlike previous studies (Sarı et al., 2019; Smeets et al., 2014), our findings suggest that auditory enhancements can be as effective as visual enhancements in promoting story comprehension and incidental word learning from repeated encounters with a story. The negative results Sarı et al. (2019) reported do not necessarily contradict the current conclusion that auditory enhancements, just as visual enhancements, can provide an enriched context for deriving the meaning of unknown words. The mixed results across the studies may relate to the music and sound quality in the target books, which can easily vary. The loudness, for instance, may differ, and, dependent on loudness, the effects on word learning may be more positive or adverse (Lehmann & Seufert, 2017; Nguyen & Grahn, 2017). It is possible that in the here used digital book, the loudness of music and sounds may have been better attuned to the oral narrative than in the digital books used by Sarı et al. (2019).

4.3. Limitations and future directions

Our study used only one picture book to test the effects of digital enhancements. Therefore, we cannot exclude that the format of the enhancements is unique for the here used book. Future studies may consider including similar digital picture books to test the efficacy of the auditory and visual enhancements and the combination of both.

As the same researcher guided the intervention and conducted the assessments, she was not blind to the experimental condition during testing, which may have influenced the test results. It would have been more elegant if the researcher doing the tests had been blind to the experimental condition.

The finding that simultaneous audio and visual enhancements degrade learning may not hold in contexts where children can control the pace of new information. For example, a previous study (Bus & Anstadt, 2021) showed no story comprehension gap between linguistically more and less advanced children when the digital picture book enables control and thus more time to process information.

The study corroborates the efficacy of combined music and environmental sounds, but whether the music, the sounds, or both have effects is not tested. Previous studies report that music has a positive impact on children's cognitive development (Trehub, 2003), reading comprehension (Su et al., 2017), and attention attraction (Shih, Huang, & Chiang, 2009). However, research on environmental sounds' effect on young children's comprehension is unavailable. Therefore, we need further studies investigating the separate impact of music and background sounds in digital picture books to support children's story comprehension and incidental word learning.

Finally, it may be essential to explore the effects of the loudness of music and sounds compared to the oral narrative in future studies.

4.4. Practical implications

Well-designed technology-enhanced picture books, including visual and auditory enhancements, can support young children's story comprehension and word learning. These enhanced digital books aligning with multimedia learning principles and embodied cognition are a promising source for families where parents are less inclined to share stories with their young children (e.g., Justice, Logan, & Damschroder, 2015). To prevent the adverse effects of a lack of book reading on children's language and literacy development and, in the long run, their academic success, it seems a good idea to implement digital books in the homes and the early childhood curricula (Hoel & Tønnessen, 2019). Digital books can meet all young children's need to hear stories even when independent digital-book reading is less optimal than the shared reading of picture books (Avelar et al., 2022; Dore et al., 2018).

The current findings enhance our understanding of an appropriate digital picture book design. Visual and auditory enhancements, like in Little Kangaroo, concretizing story events, arousing children's curiosity and speculation about story events, and highlighting characters' feelings and emotions, can help them understand the story and learn new vocabulary. However, this study's main lesson is that designers must be careful with adding additional sources. Children's comprehension may diminish when digital books combine visual enhancements with music and sounds throughout the story. This effect may occur because the additional information sources in digital books easily exceed the optimal working memory load, creating a cognitive overload and decreasing learning performance (Mayer, Heiser, & Lonn, 2001). Another lesson from this study is that the app designer should consider the child's age. Three-year-olds most need enhancements to comprehend the story and learn new vocabulary from independent reading sessions but are also the ones whose memory is easily overloaded.

CRediT author statement

Xuan Li: Conceptualization, Investigation, Project administration, Data Curation, Formal analysis, Writing- Original Draft preparation. **Adriana G. Bus**: Methodology, Supervision, Validation, Data analysis, Writing- Review & Editing.

References

Avelar, D., Dore, R. A., Schwichtenberg, A. J., Caroline, K., Roben, P., Hirsh-Pasek, K., & Roberta Michnick Golinkoff. (2022). Children and parents' physiological arousal and emotions during shared and independent e-book reading: A preliminary study. *Int. J. Child-Comput. Interact., 33*, Article 100507. https://doi.org/10.1016/j.ijcci.2022.100507

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135–146). Cambridge, UK: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in Psychology, 3*, 327. https://doi.org/10.3389/fpsyg.2012.00327, 327.

Brookshire, J., Scharff, L. F. V., & Moses, L. E. (2002). The influence of illustrations on children's book preferences and comprehension. *Reading Psychology, 23*(4), 323–339. https://doi.org/10.1080/713775287

Bus, A. G., & Anstadt, R. (2021). Toward digital picture books for a new generation of emergent readers. *AERA Open, 7*. https://doi.org/10.1177/23328584211063874

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*(2), 151–170. https://doi.org/10.1002/(SICI)1099-0720(199604)10:2<151::AID-ACP380>3.0.CO;2-U

Christ, T., Wang, C., Chiu, M. M., & Strekalova-Hughes, E. (2019). How App books' affordances are related to young children's reading behaviors and outcomes. *AERA Open, 5*, 10.1177%2F2332858419859843.

Cognitive Psychology laboratory. (2019). St. Louis: Washington University. English Lexicon Project Website https://elexicon.wustl.edu/.

College of Education and Human Sciences. (2019). *University of Nebraska-lincoln. Preschool non-disabled children's vocabulary*. Retrieved from https://cehs.unl.edu/documents/secd/aac/vocablists/VLNONAACCU2.pdf.

Dore, R. A., Hassinger-Das, B., Brezack, N., Valladares, T., Paller, A., Vu, L., et al. (2018). The parent advantage in children's e-book comprehension. *Early Childhood Research Quarterly, 44*, 24–33. https://doi.org/10.1016/j.ecresq.2018.02.002

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. San Antonio, TX: Pearson Assessments.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research, 91*(4), 483–517. https://doi.org/10.3102/0034654321998074

Furnham, A., & Stephenson, R. (2007). Musical distracters, personality type and cognitive performance in school children. *Psychology of Music, 35*, 403–420. https://doi.org/10.1177/0305735607072653

Hoel, T., & Tønnessen, E. S. (2019). Organizing shared digital reading in groups: Optimizing the affordances of text and medium. *AERA Open, 5*. https://doi.org/10.1177/2332858419883822

Justice, L. M., Logan, J. R., & Damschroder, L. (2015). Designing caregiver-implemented shared-reading interventions to overcome implementation barriers. *Journal of Speech, Language, and Hearing Research, 58*, 1851–1863. https://doi.org/10.1044/2015_JSLHR-L-14-0344

Kucirkova, N. (2019). How could children's storybooks promote empathy? A conceptual framework based on developmental psychology and literary theory. *Frontiers in Psychology, 10*, 121. https://doi.org/10.3389/fpsyg.2019.00121

Lehmann, J. A. M., & Seufert, T. (2017). The influence of background music on learning in the light of different theoretical perspectives and the role of working memory capacity. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01902

Lüdecke, D. (2019). *Esc: Effect size computation for meta analysis*. https://doi.org/10.5281/zenodo.1249218 Version 0.5.1. https://CRAN.R-project.org/package=esc

Mangen, A., Hoel, T., & Moser, T. (2018). Technologies, affordances, children, and embodied reading: A case for interdisciplinarity. In N. Kucirkova, J. Rowsell, & G. Falloon (Eds.), *The routledge handbook of learning with technology in early childhood* (pp. 235–248). New York, NY: Routledge.

Mangen, A., & Pirhonen, A. (2022). Reading, writing, technology, and embodiment. In S. L. Macrine, & J. M. B. Fugate (Eds.), *Movement matters. How embodied cognition informs teaching and learning* (pp. 103–117). Cambridge, MA: The MIT Press.

Mayer, R. E. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 169–182). Cambridge, UK: Cambridge University Press.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198. https://doi.org/10.1037//0022-0663.93.1.187

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401.

Miller, E. B., & Warschauer, M. (2014). Young children and e-reading: Research to date and questions for the future. *Learning, Media and Technology, 39*(3), 283–305. https://doi.org/10.1080/17439884.2013.867868

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science, 26*, 1489–1496. https://doi.org/10.1177/0956797615594361

Nguyen, T., & Grahn, J. A. (2017). Mind your music: The effects of music-induced mood and arousal across different memory tasks. *Psychomusicology: Music, Mind, and Brain, 27*(2), 81. https://doi.org/10.1037/pmu0000178

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly, 38*(1), 36–76. https://doi.org/10.1598/rrq.38.1.3

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*, 23–33.

Sarı, B., Başal, H. A., Takacs, Z. K., & Bus, A. G. (2019). A randomized controlled trial to test efficacy of digital enhancements of storybooks in support of narrative comprehension and word learning. *Journal of Experimental Child Psychology, 179*, 212–226. https://doi.org/10.1016/j.jecp.2018.11.006

Shih, Y. N., Huang, R. H., & Chiang, H. S. (2009). Correlation between work concentration level and background music: A pilot study. *Work, 33*(3), 329–333. https://doi.org/10.3233/WOR-2009-0880

Smeets, D. J., van Dijken, M. J., & Bus, A. G. (2014). Using electronic storybooks to support word learning in children with severe language impairments. *Journal of Learning Disabilities, 47*(5), 435–449. https://doi.org/10.1177/0022219412467069

Su, Y. N., Kao, C. C., Hsu, C. C., Pan, L. C., Cheng, S. C., & Huang, Y. M. (2017). How does mozart's music affect children's reading? The evidence from learning anxiety and reading rates with e-books. *J. Educ. Technol. Soc., 20*(2), 101–112.

Sun, H., Loh, J., & Roberts, A. C. (2019). Motion and sound in animated storybooks for preschoolers' visual attention and Mandarin language learning: An eye-tracking study with bilingual children. *AERA Open, 5*, 1–19. https://doi.org/10.1177/2332858419848431

Takacs, Z. K., & Bus, A. G. (2018). How pictures in picture storybooks support young children's story comprehension: An eye-tracking experiment. *Journal of Experimental Child Psychology, 174*, 1–12. https://doi.org/10.1016/j.jecp.2018.04.013

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience, 6*(7), 669–673. https://doi.org/10.1038/nn1084

Van Genechten, G. (2005). *Kleine kangoeroe [Little Kangaroo]. Hasselt, Belgium: Clavis (animated by Het Woeste Woud* (Groningen, The Netherlands)).
Verhallen, M. J. A. J., Bus, A. G., & de Jong, M. T. (2006). The promise of multimedia stories for kindergarten children at risk. *Journal of Educational Psychology, 98*, 410–419. https://doi.org/10.1037/0022-0663.98.2.410
Wilson, E. O. (2014). *The meaning of human existence*. WW Norton & Company.
Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 413–432). Cambridge, UK: Cambridge University Press.