The modality effect in a mobile learning environment: Learning from spoken text and real objects

Tzu-Chien Liu, Yi-Chun Lin, Yuan Gao and Fred Paas

Tzu-Chien Liu is a professor at the Department of Educational Psychology and Counseling, National Taiwan Normal University, Taiwan. Yi -Chun Lin is a postdoctoral research fellow at the Department of Educational Psychology and Counseling, National Taiwan Normal University, Taiwan. Yuan Gao is a researcher at the Smart Learning Institute, Beijing Normal University, China. Fred Paas is a professor at the Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, the Netherlands and professorial fellow at the Early Start Research Institute, University of Wollongong, Australia. Address for correspondence: Tzu-Chien Liu, the Department of Educational Psychology and Counseling, National Taiwan Normal University, No.162, Section 1, Heping East Road, Da-an District, Taipei City 10610, Taiwan. Email: tzuchien@ntnu.edu.tw

Abstract

The finding that under split-attention conditions students learn more from a picture and spoken text than from a picture and written text (ie, the modality effect) has consistently been found in many types of computer-assisted multimedia learning environments. Using 58 fifth-grade and sixth-grade elementary school children as participants, we investigated whether the modality effect can also be found in a mobile learning environment (MLE) on plants' leaf morphology, in which students had to learn by integrating information from text and real plants in the physical environment. A single factor experimental design was used to examine the hypothesis that students in a mixed-mode condition with real plants and spoken text (STP condition) would pay more attention to the real plants, and achieve higher performance on retention, comprehension, and transfer tests than the single-mode condition with real plants and written text (WTP condition). Whereas we found that participants in the STP condition paid more attention to observing the plants, and achieved a higher score on the transfer test than participants in the WTP condition, no differences were found between the conditions for retention and comprehension test performance.

Introduction

Mobile technology is increasingly being used in education to assist learning (Chiang *et al.*, 2015; Huang, Chiu, Liu, & Chen, 2011; Liu *et al.*, 2003; Sung, Chang, & Liu, 2016). The portable character of mobile devices makes learning independent of place and time, and enables the incorporation of natural experiences into the learning environment (El-Hussein & Cronje, 2010). An example of such a natural experience is the observation of real objects (eg, art, animals or plants) in both formal and informal learning environments (eg, museum, zoo or botanical garden). With the assistance of a mobile device, learners can profit from receiving multiple information sources (eg, text, picture, audio, objects), and concrete learning experiences, which can potentially facilitate the construction of cognitive schemas and motivation, and consequently learning performance (Hwang, Hung, Chen, & Liu, 2014; Sung, Chang, Lee, & Yu, 2008). However, the cognitive load that is added through the observation of real objects in mobile device supported learning environments can also result in cognitive overload and therefore negatively affect learning performance (Liu, Lin, & Paas, 2014; Liu, Lin, Tsai, & Paas, 2012).

Practitioner Notes

What is already known about this topic

- The modality effect indicates that the combination of spoken text and pictorial information can result in better learning performance than the combination of written text and pictorial information.
- The modality effect has been found in many studies carried out in computer-assisted multimedia learning environments.

What this paper adds

- Examined whether modality principle can be applied in a mobile learning environment (MLE).
- Investigated whether written or spoken text is more suitable for learning from real objects in a MLE.
- Provided recommendations for designing MLEs based on a cognitive load perspective.

Implications for practice and/or policy

- In MLEs, the use of spoken text can reduce the cognitive load on the visual channel, which can facilitate deeper cognitive processing and result in a better transfer performance.
- Students pay more attention to the real plants when learning with spoken text and real plants than when learning with written text and real plants.

In order to solve the potential problem of cognitive overload when learning with a mobile device in the physical environment, Liu, Lin, and Paas (2013) proposed a method to reduce learners' cognitive load by using arrow line cues to focus learners' attention on relevant information in the mobile learning environment (MLE). In addition, Liu *et al.* (2014) showed that the high cognitive load resulting from processing the rich information conveyed by photos of real plants in a MLE could be avoided, and learning performance could be improved by replacing photos with schematic hand drawings of the plants. Moreover, Gao, Liu, and Paas (2016) showed that quick respond (QR) code technology is an effective way to manage high cognitive load in MLEs by eliminating the need to search for target objects.

Although, audio-based materials constitute an important information source of MLEs, previous studies have seldomly referred to the effects of audio materials in such environments. Therefore, the present study used a cognitive load theory (CLT) perspective to investigate which mode of media (written text or spoken text) is beneficial for learning from real objects, by comparing the learning performance following written and spoken texts embedded in a mobile device supported learning environment with real plants.

CLT is a learning and instruction theory established to coordinate instructional procedures with human cognitive architecture, which consists of a limited working memory and an unlimited long-term memory (Leahy & Sweller, 2016; Paas, Renkl, & Sweller, 2003; Paas & Van Merrienboer, 1994; Sweller, Ayres, & Kalyuga, 2011). The recent CLT model (eg, Choi, Van Merriënboer, & Paas, 2014; Sweller *et al.*, 2011) integrated three types of cognitive load (intrinsic cognitive load, extraneous cognitive load and germane cognitive) to two types. An intrinsic cognitive load, which is caused by the inherent difficulty of learning material, and therefore cannot be reduced without altering the nature of learning materials. In contrast, extraneous cognitive

load, which is the load resulting from inappropriately designed instruction, can be eliminated by optimizing instructional design.

CLT has generated a series of cognitive load effects that provide guidelines for instruction design, focusing on reducing extraneous cognitive load (eg, worked-example effect, split-attention effect, modality effect, redundancy effect), managing intrinsic cognitive load (eg, element interactivity effect, isolated/interacting elements effect) and increasing intrinsic cognitive load (ie, previously called germane cognitive load; eg, variability effect, imagination effect).

The modality effect is one of the instructional design principles of CLT, which is also called the modality principle in Mayer's cognitive theory of multimedia learning (Mayer, 2014). The effect indicates that the combination of spoken text and pictorial information can result in better learning performance than the same information presented as written text and pictorial information (Clark & Mayer, 2011; Ginns, 2005; Mayer, 2014; Sweller *et al.*, 2011).

Two assumptions are often used to explain the modality effect (Rummer, Schweppe, Fürstenberg, Scheiter, & Zindler, 2011). According to the visuospatial load assumption, the modality effect can be explained by a better use of the available processing units for visual and auditory information in working memory, which avoids visuospatial working memory overload when all information is presented in the visual mode (Baddeley, 1992; Mousavi, Low, & Sweller, 1995). According to the temporal contiguity assumption, the modality effect occurs when all information can be held concurrently in working memory, allowing referential connections between elements to be constructed in working memory and encoded in long-term memory (Ginns, 2005). In terms of this assumption, the use of auditory text is considered better than visual text to establish contiguity, because learners are only able to focus on a picture and auditory text simultaneously (Moreno & Mayer, 1999). Rummer *et al.* (2011) have proposed and proven the auditory recency assumption to explain the modality effect. According to this assumption, the modality effect might be caused by early, sensory processes. Although acoustic-sensory information is more durable than visual-sensory information, and acoustic-sensory information will be overwritten by further auditory input. Therefore, a modality effect is restricted to the information presented most recently. An alternative explanation for the modality effect is based on a reduction of extraneous cognitive load by preventing learners from splitting attention between two sources of visual information that need to be mentally integrated for understanding to commence (Ayres & Sweller, 2014; Chandler & Sweller, 1991, 1992; Low & Sweller, 2005; Schnotz & Kürschner, 2007). When the two sources of information are presented to different modalities, eg, by using picture with spoken text, learners can process both sources simultaneously, which makes it easier to mentally integrate the information.

Although, the modality effect has been found in many studies carried out in computer-assisted multimedia learning environments (eg, de Oliveira Neto, Huang, & de Azevedo Melli, 2015; Mayer, Dow, & Mayer, 2003; Mousavi *et al.*, 1995), it should be noted that a reverse modality effect may also occur under some conditions. For example, the transient information effect (Leahy & Sweller, 2011; Singh, Marcus, & Ayres, 2012) indicates that when the learning materials contain long and complicated narrations, learning with written text accompanied with pictures can lead to better learning performance than learning with spoken text accompanied with pictures, because written text can be re-read by the learner. In a recent meta-analysis, Reinwein (2012) showed that the modality effect is moderated by variables, such as type of visualization, pace of presentation and text length, and that some of the results regarding these variables challenge interpretations of the modality effect in terms of Mayer's cognitive theory of multimedia learning and Sweller's CLT. In line with these challenges, Schüler, Scheiter, and Schmidt-Weigand (2011) have argued that the modality effect is not as ubiquitous as one might assume and that recommendations based on the effect should be considered with care.

Learning with mobile technology is different from learning in computer-assisted multimedia learning environments, because learners require to establish a relationship between the audio or visual information presented in the mobile device and the visual information associated with real objects in the physical learning environment (Liu *et al.*, 2014). If the information embedded in mobile devices is also presented in a visual mode (eg, written text), the requirement to shift attention between the visual information from the mobile device and real objects may impose a high extraneous cognitive load on the visual channel of working memory. Applying the modality principle to MLEs, it could be argued that the information embedded in the mobile devices should be presented in an auditory mode if the real objects in the physical environment can only be presented in a visual mode.

In this study, the learning effects of using either a visual or auditory mode of presenting learning materials on a mobile device in combination with real plants in the physical environment were investigated. Based on the modality effect, it was hypothesized that learning the concepts of leaf characteristics in a mobile device supported learning environment with real plants would be better with spoken than with written text on the mobile device. It was assumed that participants who were presented with the spoken text on a mobile device and real plants in the physical environment during learning would be able to pay more attention to the plants, resulting in better retention, comprehension and transfer performance than participants who were presented the written text on a mobile device and real plants in the physical environment.

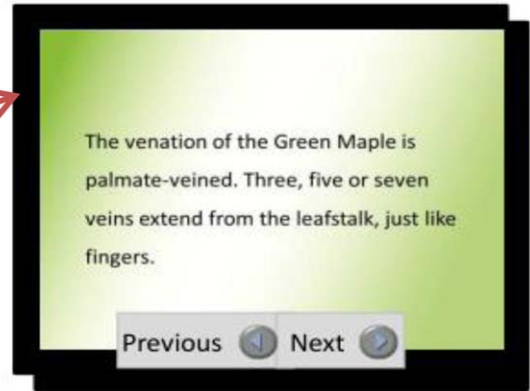## Methods
Participants and design
Fifty-eight fifth-grade and sixth-grade students (25 boys and 33 girls) from a public elementary school in northern Taiwan, mean age 11.05 years (SD = 0.76, range = 10–12), participated in this study. They had some tuition about leaf morphology in their third-grade study and therefore had some prior knowledge about the basic concepts of leaf morphology.

A single factor experimental design was conducted to address the research problem, with presentation format of the learning materials on the mobile device used to learn about real plants as the independent variable. Specifically, the participants were randomly assigned to learn with written text and real plants (WTP condition) or with the spoken text and real plants (STP condition) (See Figure 1 for examples of learning materials in the two conditions). The dependent variables of the current experiment included retention test performance, comprehension test performance, transfer test performance and five variables reflecting how often and for how long participants observed (ie, paid attention to) the plants.

Learning materials and apparatus
The learning topic of this study was plant leaf morphology. Taking six real plants as examples, the contents of the learning materials described three main characteristics of the plants on leaf morphology as well as their subtypes, including four subtypes of venation (parallel-veined, feather-veined, palmate-veined and midrib distinct), four subtypes of margin (entire, dentate, palmately lobed and sinuate) and five subtypes of phyllotaxy (alternate, decussate, distichous, whorled and rosulate).

The learning materials included two versions separately designed for the WTP condition and the STP condition, both of which were presented on a tablet PC. The spoken text was recorded by a female voice with a time length of 3 minutes and 57 seconds. Except for the presentation format (written text vs. spoken text), the learning materials were identical in content for both the WTP condition and the STP condition, including a total of 32 pages on the mobile device to describe the characteristics of plant leaves. The first page presented the introduction of the learning activity. The following 30 pages introduced the leaf morphology knowledge of six plants, with five

(a) WTP condition

The venation of the Green Maple is palmate-veined. Three, five or seven veins extend from the leafstalk, just like fingers.

Previous ◀ Next ▶

(b) STP condition

The Venation of the Green Maple
Please listen to the audio instructions and observe the plants

🔊 Repeat

Previous ◀ Next ▶

Figure 1: Examples of learning materials in the (a) WTP and (b) STP conditions (note that the original materials were presented in the Chinese language) [Colour figure can be viewed at wileyonlinelibrary.com]

pages for each plant. Within the five pages of the introduction for each plant, the first 4 pages were used in order to introduce the plant name and its subtypes of venation, margin and phyllotaxy. In order to avoid the transient information effect, the number of words (ie, Chinese characters) on each page was controlled to be less than 58 characters. The last page for each plant was used to remind the participants that they could not go back to the previous pages once they had pressed the button named "next step." The final page of the learning materials was used to thank the participants.

The equipment used for the current experiment was identical for all participants, including an iPad like tablet PC with a 10-inch monitor and a headphone. To avoid differences in the treatment between the two conditions, both the WTP and STP condition were provided

with headphones. The instructional software for the iPad like tablet PC was developed in JAVA.

Measures
The prior knowledge test scores, together with participants' natural science and Chinese scores of last semester were used to verify whether the random assignment of participants to the conditions was successful.

Prior knowledge test
A prior knowledge test was used to evaluate participants' prior knowledge about the basic concepts of leaf morphology. The test consisted of 10 four-item multiple-choice questions. One mark was given for a correctly answered question; the highest possible score was therefore 10 points and the lowest possible score of the test was 0. The internal consistency reliability coefficient (KR-20) of the prior knowledge test was 0.81.

Retention test
A retention test was used to test how well students are able to remember what they have learned. The test consisted of 18 multiple-choice questions. Twelve pictures about the six plants used during the learning phase were presented to the participants. For six pictures of them, the participants were asked to indicate the subtypes associated with venation and margin for the plant displayed in each picture, generating 12 multiple-choice questions with four alternative answers. For the other six pictures, the participants required to indicate the subtypes associated with phyllotaxy, generating six questions with five alternative answers. One point was given for a correct answer and thus the full mark for the retention test was 18, and the lowest possible score was 0. The KR-20 test revealed an acceptable internal consistency reliability of 0.71.

Comprehension test
A comprehension test included two types of tasks: a drawing task in which the students were asked to draw four subtypes of venation and four subtypes of margin that they had acquired during the learning phase, and an assembling task in which the students needed to reconstruct five types of phyllotaxy using artificial leaves and stems. Thus, the drawing and assembling tasks consisted of eight items and five items separately. One mark was given for a correctly completed item for both the drawing and the assembling tasks, and therefore, the full mark for the whole comprehension test with two tasks was 13, and the lowest possible score was 0. The KR-20 test also revealed an acceptable internal consistency reliability coefficient of .70 for the comprehension test performance.

Transfer test
A transfer test, which consisted of 18 multiple-choice questions, was used to test how well the students were able to apply the leaf morphology knowledge that they had acquired during the learning phase to the novel, real plants. Similar to the retention test, the participants were asked to indicate the subtypes associated with venation, margin, and phyllotaxy separately for each of the six plants, with four alternative answers being provided for 12 multiple-choice questions associated with vernation and margin whereas five alternative answers being provided for six multiple-choice questions associated with phyllotaxy. However, all six target plants in this transfer test were novel and they were never presented to the students during the learning phase. One mark was given for a correctly answered question, resulting in a highest possible score of 18 points for the transfer test performance and a lowest possible score of 0. The internal consistency reliability coefficient (KR-20) of the
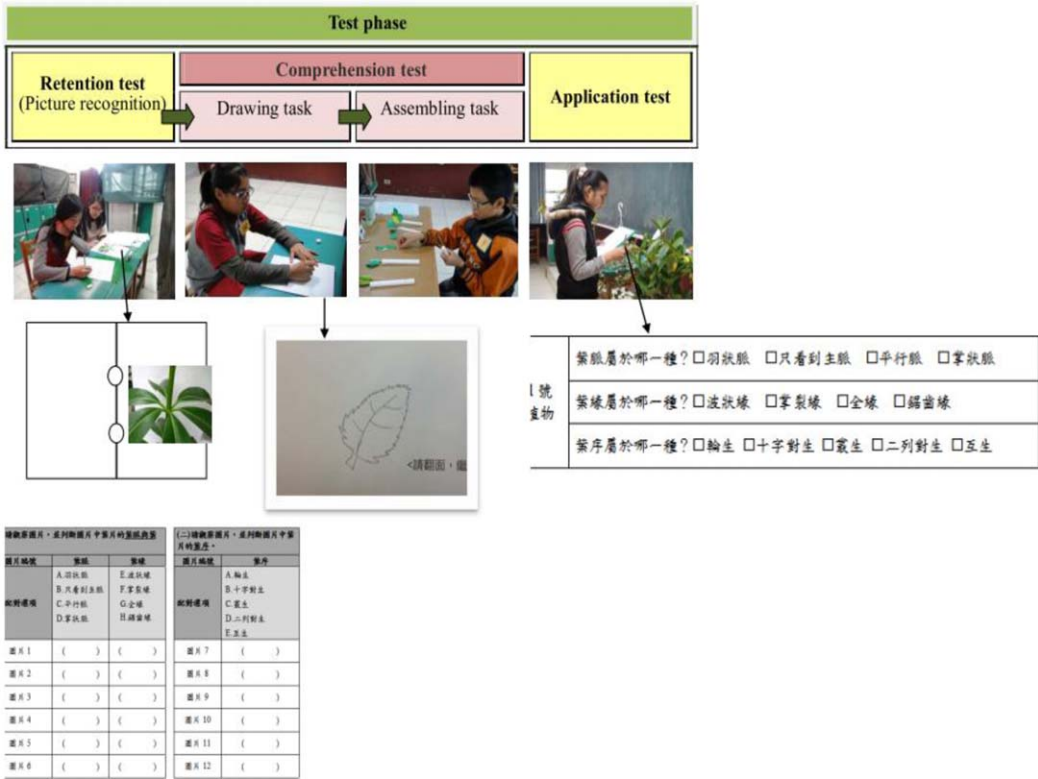
Figure 2: Examples of the retention, comprehension, and transfer tests [Colour figure can be viewed at wileyonlinelibrary.com]

transfer test was 0.74. Figure 2 depicts the retention, comprehension and transfer tests of the research.

Behaviors of observing plants

Regarding the behaviors of observing plants, all learning behavior of participants' during the learning phase, such as watching the Tablet PC, observing plants, watching objects other than plants, was recorded by two video cameras. However, because the present study aimed to investigate plant observation when using different formats of learning materials as well as the learning effects resulting from the plant observation, only the behavior of "plant observation" was coded and transcribed after the experiment for analyses. Two experimenters watched the videos and coded the data. The interrater agreement for the coding of behavior of "plant observation" was high (Cohen's kappa value $\kappa = .89$), and thus the first coder's assessments were used for all analyses.

Two indexes were used to code the behaviors of observing plants, they were (1) "total number of times plants were observed," which showed for how many times in total the participants observed the target plants during learning. One time of observation would be scored when the participant's eyesight contacted with a plant and then left it and (2) "total time spent on observing plants," which reflected the total amount of time the participant's eyesight stayed on the plants. The duration for one time of plant observation was considered to be from the participant's eyesight contacted with the plant to leaving it, and the total time was the sum of each duration. The work for coding the index of "total time spent on observing plants" was completed with the assistance of Microsoft MovieMaker.

Procedure
The procedure of the current study involved three phases: a pre-experimental phase, a learning phase and a test phase. Except for the pre-experimental phase, the experiment was administered on an individual basis.

Pre-experimental phase
One day before conducting the experiment, the participants were given 10 minutes of class time to finish a prior knowledge test on a sheet of paper, which was collected at the end of the test.

Learning phase
Before the learning commenced, the participants were first introduced to the purposes of this experiment. After that, they were provided with five pages of introduction about the three main characteristics (veins, margin, phyllotaxy) of leaves on morphology in a separate notebook computer, which actually had already been learned in their third-grade natural science classes. The reason for reviewing this knowledge before the experiment was that three main characteristics of leaves were the basic foundation for further understanding of subtypes of leaf characteristics. When the participants finished reading those pages, the experimenters asked them to point out the position of the veins, margin and phyllotaxy in a real plant. If the participants could give the right answers, they were allowed to go to the next step (learning about subtypes of three main characteristics of leaves) of the experiment. Otherwise they would be taught about the three main characteristics again till they could correctly point out the right positions.

After making sure that the participants had understood what the veins, margin and phyllotaxy were, they were told that the formal learning would be starting. The six plants were arranged in a row. To learn from the plants, all participants were guided to stand in front of the first plant to start their learning and then move on to the next plant, until the sixth plant. The participants in the WTP condition were guided to learn the subtypes of leaf characteristics by reading the written texts on the Tablet PC and simultaneously observing the real plants. In contrast, the participants in the STP condition were guided to learn the subtypes of leaf characteristics by listening to the spoken texts from the Tablet PC through the headphone and simultaneously observing the real plants. Both conditions of the participants could learn at their own pace by using a touch pen within 11 minutes. All learning processes were recorded in a log file saved in the Tablet PC and their learning behaviors were also recorded by two video cameras. Furthermore, the participants had to learn the six plants sequentially, and they were not allowed to go back to a previous plant once they started studying in the next section about a new plant.

Test phase
The test phase immediately followed the learning phase in order to evaluate the participants' understanding of the subtypes of the leaf characteristics, including a retention test, a comprehension test and a transfer test. The three tests were all paper-and-pencil tests and they were administered successively within a total time limit of 19 minutes (4 minutes for the retention test, 10 minutes for the comprehension test and 5 minutes for the transfer test). Similar to the learning phase, the test phase was also administered on an individual basis.

Data analyses
Analyses of comparison of similar conditions
Three independent *t*-tests were separately conducted on participants' prior knowledge test scores, natural science scores and Chinese language scores of last semester after the pre-experimental phase in order to examine the equality of the two conditions in all three aspects related to learning about leaf morphology.

Analyses of learning performance

A series of one-way analyses of covariance (ANCOVAs) was conducted to compare the differences between the two groups' retention, comprehension and transfer test scores, while reliably removing the influence of learning time on learning performance. The significance level was set at 0.05.

Analyses of plant observation behaviors

A series of independent *t*-tests were conducted on the two indexes, examining the differences on the behaviors of "plant observation" between the two conditions. The significance level was also set at 0.05, unless mentioned otherwise.

Results

Analyses of comparison of similar conditions

Table 1 shows the means and standard deviations. The independent *t*-test for the three test scores used to verify whether the experimental conditions were similar. The results indicated that there were no significant differences between conditions on the prior knowledge test scores, $t(56) = -0.63$, $p > .05$, Nature Science scores, $t(56) = 0.18$, $p > .05$, and Chinese Language scores, $t(56) = -0.50$, $p > .05$, thereby confirming that the random assignment of participants to conditions was successful.

Learning time

Regarding the time spent on the learning phase. The results of an independent *t*-test indicated a significant effect of the presentation format of the learning materials (WTP condition vs. STP condition) on learning time, $t(56) = -5.43$, $p < .001$, Cohen's $d = 1.43$, with participants in the STP condition ($M = 6.49$, $SD = 1.26$) spending more time on learning than the participants in the WTP condition ($M = 4.68$, $SD = 1.28$).

Learning performance during test phase

Table 2 shows the means and standard deviations for learning performance during the test phase. The learning performance data were analyzed using a series of ANCOVAs with the retention, comprehension and transfer test scores as dependent variables. The preliminary check was conducted to ensure that there was no homogeneity of the regression slopes for each of the dependent variables, $F(1, 54) = 0.17$, $p > .05$; $F(1, 54) = 0.24$, $p > .05$; $F(1, 54) = 2.18$, $p > .05$. Regarding learning performance, after adjusting for the learning time, the results failed to indicate any significant difference of presentation format of learning materials on either the retention test, $F(1, 55) = 0.32$, $MSE = 6.56$, $p > .05$, $\eta_p^2 = .01$, or the comprehension test, $F(1, 55) = 0.26$, $MSE = 4.15$, $p > .05$, $\eta_p^2 = .01$. With regard to the transfer test, after adjusting for the learning time, there was a significant difference between the two groups regarding their scores, $F(1, 55) = 7.08$, $MSE = 4.45$, $p < .05$, $\eta_p^2 = .11$. The adjusted mean score of the transfer test was higher in the STP condition ($M = 11.33$) than in the WTP condition ($M = 9.50$).

Table 1: The means and standard deviations for prior knowledge test scores, nature science scores and Chinese language scores

			WTP condition		STP condition	
Variable	Min	Max	M	SD	M	SD
Prior knowledge	0	10	5.59	2.56	5.97	1.99
Nature science	74	96	88.82	5.74	88.54	5.76
Chinese language	73	98	89.53	5.89	90.28	5.68

Table 2: The means and standard deviations for test performance during test phase

Variable	Min	Max	WTP condition		STP condition	
			M	SD	M	SD
Retention test	5	18	9.21	2.32	10.52	2.86
Comprehension test	3	12	7.48	2.11	8.34	1.99
Drawing task	3	8	5.66	1.45	5.76	1.43
Assembling task	0	4	1.83	1.07	2.59	1.05
Transfer test	6	16	9.24	1.96	11.59	2.28

The degree of attention on the plants during learning phase

A series of independent t-tests were used to analyze participants' attention on the plants during learning. The results indicated that the participants in the STP condition significantly outperformed the participants in the WTP condition on both two indexes, $t(56) = 3.14$, $p < .01$, Cohen's $d = 0.82$, for the total number of times plants were observed, $t(56) = 11.86$, $p < .001$, Cohen's $d = 3.12$, for the total time spent on plant observation. Table 3 presents the means, standard deviations and the analysis results of independent t-tests for the both indexes of the behaviors of plant observation.

Discussion and conclusion

This study aimed to examine the modality effect in a MLE. The results partially supported the modality effect hypothesis that learners would learn better about the morphology of plants when the learning materials on the mobile device that were associated with the real plants in the physical environment were presented as spoken text rather than written text. However, the superiority of dual modes relative to single mode of presentation of learning materials was only significant for the transfer test, and not for the retention and the comprehension test.

These results can be interpreted in terms of the purpose that the tests served. Whereas, the retention and the comprehension tests were used to measure how well the participants remembered the target plants associated learning materials, the transfer test was used to measure how well they could apply the knowledge that they had acquired in the MLE. Mayer (2001, 2005) proposed that such a transfer test can be used as an indication of the depth of processing during learning. Accordingly, compared to the retention and the comprehension tests, the transfer test can be considered to be more sensitive to indicate the degree to which the participants processed the information in working memory and stored it in cognitive schemas in long-term memory for retrieval when required.

The findings of the modality effect in the transfer test can be explained with CLT. When learning about the leaf characteristics of plants in a physical learning environment, both the learning materials carried by mobile devices and the real plants necessarily contributed to the learning and required to be dealt with simultaneously by working memory resources. When the written

Table 3: The means and standard deviations for the behaviors of observing plants

Variables	Min	Max	WTP condition		STP condition	
			M	SD	M	SD
Total number of times plants were observed	11	116	33.31	16.03	49.28	22.22
Total time spent on observing plants (s)	18.46	277.89	59.48	30.27	187.33	49.52

texts were presented to the participants as narration, the information associated with the spoken materials can be processed in the verbal channel, thereby freeing working memory capacity in the visual channel that can be devoted to processing the information associated with the real plants more deeply. In this way, using spoken text to learn plants reduced the cognitive load on the visual channel, facilitating deeper cognitive processing and resulting in a better learning performance. In contrast, using written text to learn about plants in the WTP condition might have caused a higher cognitive load in the visual channel of working memory because both sources of information needed to be processed in this system, thereby impairing learning. The better transfer performance of STP condition could also be explained by the avoidance of split attention. Comparing to the participants in the WTP condition, participants in the STP condition could listen to the spoken text and observe the corresponding plants simultaneously. This could have reduced the cognitive resources needed for searching and integrating verbal and pictorial information, thereby leaving more cognitive resources for deep learning. Moreover, the better assembling task performance for participants in the STP condition could be explained by the auditory recency effect (Rummer *et al.*, 2011), because the subtype of phyllotaxy for each plant was introduced on the last page of the four pages used to introduce the plant name and its subtypes of venation, margin and phyllotaxy.

The analyses of attention behaviors on the plants further supported this explanation. The participants in the STP condition with spoken text were found to pay significantly more attention to the real plants than the participants in the WTP condition with written text. The participants receiving a mixed mode of information in the STP condition could observe the real plants while at the same time listening to the spoken materials. Therefore, it was easier for participants in the STP condition to mentally integrate the information from both resources than for participants in the WTP condition, who were unable to attend to both resources at the same time.

Another point that is worth to discuss is that field-based mobile learning might benefit transfer tasks more than retention and comprehension tasks. In this study, participants in the STP condition spent more learning time on the learning task than participants in the WTP condition, but they did not get significantly better retention and comprehension performance. One possible reason is that it might be less cognitively efficient to remember the subtypes of venation, margin and phyllotaxy of the same six plants shown on the pictures and correctly drawing the venation and margin of the plants with the use of field-based mobile learning. Educators should consider the instructional objectives when deciding upon whether to use field-based mobile learning.

A limitation of the present study was lack of a direct measure of cognitive load. Although the indexes of plants observation recorded during the learning phase provided objective evidence for the distribution of participants' attention over the learning processes and contributed to the explanations of the research findings, future research is suggested to use cognitive load evaluation techniques such as cognitive load self-rating scales (see Paas *et al.*, 2003) to collect more direct data. Nevertheless, the findings of the present study about the modality effect in a physical learning environment may have general implications for using mobile devices in an educational context. In addition, whether spoken text is in general better than written text when presented on tablets or whether this depends on additional "pictorial input" (ie, multimedia effect) could not be answered in the current study because the real plants were shown in both conditions. Future studies focused on mobile learning are recommended to explore the effects of spoken text and written text on learning in conditions with or without pictorial input.

Statements on open data, ethics and conflict of interest

All data of this research can be accessed by sending request e-mails to the corresponding author. The applicants will be asked to clearly describe their research purposes in the request e-mails.

All research data are anonymous. Participants knew that their participation in this research was voluntary and they could stop participating at any time.

There is no conflict of interest in this study.

References

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). New York, NY: Cambridge University Press.

Baddeley, A. (1992). Working memory. *Science, 255,* 556–559.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8,* 293–332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62,* 233–246.

Chiang, F.-K., Zhu, G., Wang, Q., Cui, Z., Cai, S., & Yu, S. (2015). Research and trends in mobile learning from 1976 to 2013: a content analysis of patents in selected databases. *British Journal of Educational Technology, 47,* 1006–1019.

Choi, H. H., Van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review, 26,* 225–244.

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning* (3rd ed.). San Francisco, CA: Wiley.

de Oliveira Neto, J. D., Huang, W. D., & de Azevedo Melli, N. C. (2015). Online learning: audio or text? *Educational Technology Research and Development, 63,* 555–573.

El-Hussein, M. O. M., & Cronje, J. C. (2010). Defining mobile learning in the higher education landscape. *Educational Technology and Society, 13,* 12–21.

Gao, Y., Liu, T. C., & Paas, F. (2016). Effects of mode of target task selection on learning about plants in a mobile learning environment: effortful manual selection versus effortless QR-Code selection. *Journal of Educational Psychology, 108,* 694–704.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15,* 313–331.

Huang, Y. M., Chiu, P. S., Liu, T. C., & Chen, T. S. (2011). The design and implementation of a meaningful learning-based evaluation method for ubiquitous learning. *Computers and Education, 57,* 2291–2302.

Hwang, G.-J., Hung, P.-H., Chen, N.-S., & Liu, G.-Z. (2014). Mindtool-assisted in-field learning (MAIL): an advanced ubiquitous learning project in Taiwan. *Educational Technology and Society, 17,* 4–16.

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25,* 943–951.

Leahy, W., & Sweller, J. (2016). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science, 44,* 107–123.

Liu, T. C., Lin, Y. C., & Paas, F. (2013). Effects of cues and real objects on learning in a mobile device supported environment. *British Journal of Educational Technology, 44,* 386–399.

Liu, T. C., Lin, Y. C., & Paas, F. (2014). Effects of prior knowledge on learning from different compositions of representations in a mobile learning environment. *Computers and Education, 72,* 328–338.

Liu, T. C., Lin, Y. C., Tsai, M. J., & Paas, F. (2012). Split-attention and redundancy effects in mobile learning in physical environments. *Computers and Education, 58,* 172–180.

Liu, T. C., Wang, H. Y., Liang, T., Chan, T. K., Ko, H. W., & Yang, J. C. (2003). Wireless and mobile technologies to enhance teaching and learning. *Journal of Computer Assisted Learning*, *19*, 371–382.

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147–158). New York, NY: Cambridge University Press.

Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York, NY: Cambridge University Press.

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Mayer, R. E., Dow, G., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: what works in the design of agent-based microworlds? *Journal of Educational Psychology*, *95*, 806–813.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology*, *91*, 358–368.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, *87*, 319–334.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist*, *38*, 1–4.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63–71.

Paas, F., & Van Merrienboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351–371.

Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research*, *41*, 1–32.

Rummer, R., Schweppe, J., Fürstenberg, A., Scheiter, K., & Zindler, A. (2011). The perceptual basis of the modality effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *17*, 159–173.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*, 469–508.

Schüler, A., Scheiter, K., & Schmidt-Weigand, F. (2011). Boundary conditions and constraints of the modality effect. *Zeitschrift für Pädagogische Psychologie*, *25*, 211–220.

Singh, A. M., Marcus, N., & Ayres, P. (2012). The transient information effect: Investigating the impact of segmentation on spoken and written text. *Applied Cognitive Psychology*, *26*, 848–853.

Sung, Y. T., Chang, K. E., Lee, Y. H., & Yu, W. C. (2008). Effects of a mobile electronic guidebook on visitors' attention and visiting behaviors. *Educational Technology and Society*, *11*, 67–80.

Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: a meta-analysis and research synthesis. *Computers and Education*, *94*, 252–275.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Berlin, Germany: Springer.