Transforming a Magazine into a Video Involving a Target Audience: A Multiliteracies Case Study in an EFL Context

Yi-Hsuan Gloria Lo, Sy-Ying Lee[*] and Wan-Chu Finny Tien

National Taiwan University of Science and Technology, Taiwan // glorialo@mail.ntust.edu.tw // syying.lee@mail.ntust.edu.tw // B10517007@mail.ntust.edu.tw
[*]Corresponding author

ABSTRACT: In this article, we presented a case study in an EFL context that investigated how a magazine was transformed into a digital video involving a target audience. A group of fifty international students responded to a survey questionnaire developed based on the multimodality framework and some were interviewed to express their preferences for the content and format of the video and to evaluate different versions of the video product. The results show that the transforming process consisted of four stages including (1) collecting target audience's preferences for the video content and format, (2) converting the discourse type from textual to oral, (3) creating multimodal materials for the video, and (4) (re-)composing the video. The target audience's responses revealed that effective multimodal orchestration could provide a better engagement and viewing experience for the target audience. The multiliteracies competency of the video creators and viewers was deepened and expanded through the digital transforming processes and interdisciplinary collaboration, which enabled EFL learners to experience, conceptualize, analyze, and apply the learned and new knowledge. With the ultimate goal to cultivate EFL learners to become multimodal literate citizens in the global society, this study advances our understanding of multimodality and yields significant pedagogical implications for multiliteracies education and educational technology in the EFL context.

Keywords: Video, Target audience, Multimodality, Multiliteracies, Educational technology

1. Introduction

Literacy has long been defined as the ability to communicate and make sense of the world effectively through language. However, the evolution of technology has created a multimodal world, inundated with websites (e.g., YouTube, Google) and social media (e.g., Facebook, Instagram, Twitter), thus leading us into the realm of new media literacy (Hung, 2019; Lin, Li, Deng, & Lee, 2013). Being literate is no longer solely about being able to read and write, the linguistic mode that has been the primary emphasis in most EFL contexts. A literate being in this 21st multimodal world should be equipped with the ability to "understand the effect of all modes of communication that are co-present in any text" (Kress, 2000, p. 337) including visual, aural, spatial and gestural modes, in addition to linguistic mode (New London Group, 1996). This notion suggests the importance of multiliteracies, referred to as the abilities and competencies "to create, critique, analyze and evaluate multimedia texts" (NCTE, 2013). In response to the trend of multimodal literacies, many composition scholars, such as Alexander and Rhodes (2014), believe that "engaging multimodality is a pressingly necessary task for a wide variety of composition" (p. 71). To advance our understanding of the multimodal composing process, the case study echoes this pressing need by examining how meaning-making in a paper format, the most common form of reading and writing production in EFL contexts, can be transformed into a video via digital multimodal composing (DMC) (Jiang, 2017), a semiotic process using digital tools to create multimodal texts combining different modes. With the goal to communicate with a specified target audience via the video to reach out to a larger population (Mills, 2015), this case study aims to cultivate EFL learners to become multimodal literate citizens in the 21st global society.

How can multiple modes be best "orchestrated" or "combined?" This issue may be informed and addressed by the two important questions raised by the 21st century literacy framework (NCTE, 2013). First, "Do students publish in ways that meet the needs of a particular, **authentic audience**?" Second, "Do students solve real problems and share results with **real audience**?" The role of the target audience has been explored in writing composition to some extent (Kakh, Mansor, & Zakaria, 2014; Wong, 2005). However, most of the target audiences in previous research (e.g., Cho & Choi, 2018; Gunel, Hand, & McDermott, 2009; Martinez-Insua, 2019) were not "real" or "authentic;" rather, they were hypothetical because the writers were instructed to imagine a group of audience they were writing to, who may or may not exist. Furthermore, the instructors in most research were often the evaluators to assess and evaluate the final production rather than the target

audience. The role of a specified target audience in the digital multimodal process is far scanter and worth investigation.

In this case study, we had two goals: first, to document the process of transforming a magazine (from static images) into a multimodal video (to moving images) for a specific target audience and second, to collect our target audience's responses to refine the "orchestration" of multiple modes. Specifically, our goals led us to ask two research questions: (1) What is the evolving process of transforming a magazine into a multimodal video when a target audience is involved? (2) What can our target audience's responses to the videos inform us about the orchestration of the multiple modes when transforming a magazine into a video?

2. Literature review

2.1. The orchestration process involved in multimodal literacies

In multimodal literacies, language is acknowledged as "only one of several semiotic tools for communicative purposes" (Yeh, 2018, p. 29). While digital technology has been identified as a powerful mediation to expand and enhance multimodal meaning-making among people nowadays, including language learners, pedagogies integrating digital and conventional literacies have been called for (Hafner, 2014; Jewitt, 2006; Kress, 2003; Mills, 2010). In response to this, many studies have investigated digital multimodal composing (DMC) by examining how students produce digital videos (Campbell & Cox, 2018; Hafner, 2014; Jiang, 2018; Yeh, 2018). The three following studies were selected for further discussion because of the close relevancy they had with our current case study. Each of them reported on how students created videos by following certain steps or processes when producing the video.

Firstly, Hafner (2014) discovered six steps in video making: reading, data collection, scripting and storyboarding, filming, editing, and sharing. Following these steps, his students identified "a wide range of semiotic resources, including moving images and animation, charts and tables for scientific data, subtitles, different camera angles and lighting, background music, sound effects, interesting locations, interesting participants, and facial expression" (p. 669). Results showed that one main source of rhetorical challenge for his students during video production is multimodal orchestration. For instance, the students were concerned if "over-relying on multimedia" or "excess use of technique may annoy the audience." Another challenge was "how to write an interesting script for narration for the documentary" because they viewed "language as an equally important resource" (p. 668). However, the students did not assess the effectiveness of the visuals and script, which is believed to be an important procedure missing in the orchestration process.

In a similar vein, Yeh (2018) mapped out three steps to create a video among EFL college students: composing the scripts, enacting the scripts, and editing the videos. In this study, the students first read information for their selected topics and composed their scripts for their videos. To enact their scripts, the students had to "come up with different innovative ways of presenting their topics through the combination of multiple modes to present their core themes" (p. 30). In the editing stage, they managed to "combine multiple modes such as adding text, pictures, subtitles, effects, narrations, soundtracks, and PowerPoint slides, to tie all their ideas together to construct the videos" (p. 31). While several perceived benefits were found in language acquisition (e.g., vocabulary, speaking, translation and writing), cultural learning, and multimodal capacities (e.g., editing skills for multimodality), how multiple modes were orchestrated or combined to create a multimodal video, however, was not explored in detail.

Different from Hafner's and Yeh's studies which mainly focused on the production processes students underwent, Monte Mŏr (2015) detailed what and how college students in Brazil experienced and learned through video creation using the *"Learning by Design"* framework proposed by Kalantzis and Cope (2005) to theorize how students learn when they engage in digital media production. This framework outlines four knowledge processes to support teaching and learning: experiencing the known and the new, conceptualizing by naming and theorizing, analyzing functionally and critically, and applying appropriately and creatively. Students in Monte Mor's study recorded and compiled their discussions on the educational system and entrance examination they had all experienced. Students performed their ideas using their own voice and facial expressions for real viewers, their peers. They critiqued and evaluated the system based on *the known* and learned *the new* by viewing and discussing others' experiences and arguments. This project was made authentic, unique, and original because of the application of the technical tool, the *Flash* software, which helped design the interactive display of the multiple videos of all student narrators showing on the screen. They eventually created a collective narrative on education, society, and the future job market as a class project. By the same token, our study intended to use this

framework to examine whether these knowledge processes were also experienced by our EFL learners during video production.

2.2. The role of target audience in (video) composing

Research has shown that target audiences play different roles in the composing process and final outcome, resulting in different impacts on the writers themselves as well as the writing outcomes. While the role of target audience is evident in writing composition (the linguistic mode), what would be the role that target audiences can bring in a video composing process and final outcome?

Wong's (2005) study showed that the intended target audiences, although imaginary, gave different mental representations, and thus a wide range of writing strategies were employed to serve different rhetorical purposes at different stages of the composing process for their "imagined" or "hypothetical" target audiences. On the other hand, writing-to-learn for different target audiences can have an impact on the learning of the writers themselves. In investigating writers' conceptual understanding of biology (Gunel, Hand, & McDermott, 2009), it is found that students writing (in a hypothetical context) for peers or younger students performed significantly better on conceptual understanding than students writing for the teacher or parents. Furthermore, in Cho and Choi's (2018) study that investigated the effect of audience specifications on the summary writing, it showed that writers with a specified audience outperformed those without one.

Target audience also plays a role to determine what and how language is represented and organized. Martinez-Insua's (2019) research indicated that texts that address learned audiences tend to be contentful (carrying more weight on content), which demands certain background knowledge from the reader in order to create or infer new discourse, whereas texts presented to lay audiences were more contentlight (less weight on content). It is also found that whereas most subject themes are contentful in formal written texts, most of them are contentlight in informal spoken texts. The distinction in different modes of presentation catering to different target audiences can also be found in Hafner's (2014) students' digital video projects.

In Hafner's (2014) study, students were required to create a multimodal scientific documentary to share with "a general audience of nonspecialists" through YouTube, and a written lab report for "a specialist audience." Although Hafner pointed out that his students met the challenge of writing for "an authentic audience" when "combining a range of modes," it is not clear who the target audience was, what language choices were made for the two types of audiences, and whether or not ways of combing a range of different modes to "appeal to their audience" (p. 655) would result from their specified target audience? The same questions could also be asked in response to Yeh's (2018) project aiming at creating a video for "authentic audiences in online communities" (p. 29). In summary, throughout the video composing process, a lack of specification of the target audience may create difficulty in determining appropriate contents, ways of combining modes appealing to the audience's needs, and assessing whether or not it communicates its intended messages effectively.

3. Methodology

3.1. The context of the case study

A case study approach was particularly useful for exploring the particularity and complexity involved in the processes of transforming a magazine to a video because it "allows in-depth, multi-faceted explorations of complex issues in [the] real-life settings" (Crowe, Cresswell, Robertson, Huby, Avery, & Sheikh, 2011, p. 1), which may not be feasibly possible via other approaches. This case study took place in a one-year capstone course proffered for university English majors to conduct a special project on a self-selected topic as a requirement for graduation. Capstone courses have been practiced worldwide to provide students with a culminating, integrative experience of learning to demonstrate what they have acquired in previous years by synthetically applying the learned skills in simulated or real-world situations (Wagennar, 1993). In this course, with the first two authors also the instructors, students in groups chose to design projects in any form that was deemed appropriate and creative to solve a predetermined problem. Among the eleven projects this year, one video-making project, conducted by the third author and her group members, which this research study was based on, fitted our research interest in multiliteracies as an EFL pedagogy.

3.2. Participants

One of the reasons for the increasing number of international students in Taiwan is that they want to experience Taiwan's culture (Pan & Zhang, 2019). Therefore, international students became the target audience of the study. Prior to the study, an invitation message informing of the several stages of cooperation required was posted on the Internet accessible to the international student body. Among the many who replied and expressed their interest and voluntariness to participate in this research, fifty students from different countries studying in graduate level programs were selected based on three criteria: (1) who stayed in Taiwan for at least a semester; (2) who had expressed a strong interest in Hualien; and (3) who were sufficiently competent in English to understand the survey questions and were able to respond to the video with opinions and suggestions.

3.3. Instruments and data collection

Both qualitative and quantitative data were collected using the following instruments.

Magazine. The magazine sought to offer international tourists an in-depth introduction to Hualien. This magazine, a multimodal project prior to the capstone course, featured three modes: linguistic (carefully-crafted words in print), visual (pictures), and spatial (professional layout and aesthetic cover design).

A survey questionnaire. The survey was designed following the five modes in order to understand the target audience's preferences for the what (content) and the how (format) of the video to be created and was distributed to the 50 international students online (see Appendix A at https://drive.google.com/file/d/1HlAPnZwcCupHX4MBG7Il8NlyU61_VARd/view). The first part contains three sections (e.g., natural landscape, the Japanese colonization period, and aboriginal groups). The participants were asked to choose three options in the first section, two options in the second, and one option in the third according to their personal preferences. The second part investigates the format based on the five modes: linguistic (genre, wording of the subtitles), visual (hosting style, subtitles and special effects), visual and spatial (style), aural (background music, sound effects), gestural (gestures), and others (length).

Interviews. Face-to-face interviews were conducted in English with 10 of the 50 respondents to elicit further elaboration on the questionnaire items. After the two videos were produced, one-on-one interviews were held in English with another 10 international students to further understand the target audience's views on the two versions of the video. The first part of the interview investigated how much (on a rating scale of 1 to 10) an interviewee liked the six tourist spots, and the second part seeing how much (on a rating scale of 1 to 10) an interviewee enjoyed the video presentation (overall style, length, hosting style, subtitles, background music, sound effects, and special effects). Lastly, the interviewees were asked to elaborate on their ratings and the reasons for the differences in the ratings (if any). For example, "How did you rate the two versions of "Qingshui Cliff?" Why did you give the second version a higher rating (Score = 9) than that (Score = 5) in the first version? What were your reasons for the different ratings? (See Appendix B at https://drive.google.com/file/d/1K49MbBKv5wkYigLjzGCkO1JEgqm8upmy/view).

Research trip to Hualien. A three-day research trip was conducted in Hualien to explore the place and culture in person and to shoot the video according to the results of the first survey and interview.

Two versions of the video. Two versions of our video were produced, one by the research team and the other with the assistance of a video-editing professional to elicit the target audience feedback (see the Results).

Field notes. The research team kept field notes to document in detail the process of conducting the case study.

3.4. Data analysis

Descriptive statistics (percentages and means) were firstly obtained from the questionnaires. Next, the interview data, transcribed verbatim and coded using the multimodal framework proposed by the New London Group (1996), were then categorized into themes. Lastly, the qualitative results triangulated with the descriptive statistics were used to address the research questions. Specifically, to examine the transformation process from a magazine to a video, the research team first analyzed the field notes by charting the major stages based on the chronological order, which resulted in the emergence of the four stages (see Results: 4.1). Further data analysis for each stage (except stage II) employed the five modes as an analytical framework to analyze the target

audience's preferences (Stage I), the orchestration of the five modes (Stage III), and the differences between the two versions of the video (Stage IV). On the other hand, data analysis for stage II centered on comparing and contrasting the differences between the linguistic texts in the magazine and in the video scripts in relation to formality (Einhorn, 1978), depth (Schallert, Kleiman, & Rubin, 1977), personal references (Strauss, Feiz, & Xiang, 2018), word choice (DeVito, 1965; Miller, 2011), and syntax (Wilkinson, 1971). To address research question 2, when analyzing the target audience's responses to the two versions of the video, the research team first coded qualitative interview data provided by the target audience to analyze the reasons for their preferences, which were then triangulated with the rating results. The results of the analysis were then corroborated with relevant data analysis gathered from the previous stages, based on which three major themes were generated (see Results: 4.2).

4. Findings

4.1. RQ1: What is the evolving process of transforming a magazine into a multimodal video when a target audience is involved?

Grounded in the five modes of multimodality, a four-stage framework for transforming a magazine into a multimodal video emerged, as presented below.

4.1.1. Stage I: Collecting target audience's preferences for the video content and format

In the first part, our target audience expressed *what* they wanted to see in the video. For "natural landscape," the ranking results showed that Qingshui Cliff (40%), Taroko Gorge (38%), and Yanzikou Trail (42%) were ranked first to third, respectively. For "the Japanese colonization period," Pine Garden was ranked first (54%) followed by Ji-An Shrine (40%). For "aboriginal groups," the Amis was ranked first (64%).

In the second part, the target audience were asked about nine aspects related to *how* they wanted the video to be presented. For what type of genre to introduce Hualien, more than half our target audience (52%) hoped a micro film in 7 to 9 minutes long (64%) and supplemented with interesting and humorous elements to be made, so it could be "less boring" (interviewee 5). Most of the target audience (80%) wanted to have a host and expected the host to be a Taiwanese who speaks English and could interact "naturally with the local people" (Interviewee 8). As for whether the subtitles were needed, 62% of the target audience chose "subtitles throughout the video." The subtitle featuring both Chinese and English was more preferred because it looked "more professional" (Interviewees 4, 6, 7, 9). Moreover, the subtitle wording should be simple, fact-based (74%), instead of rich, detail-oriented descriptions. Among the five options of background music, country music was the most popular (60%), because of its "more relaxing tone" (Interviewee 2) and "a brisk rhythmic pace" (Interviewee 8). Interestingly, most of our target audience did not appreciate any special effects (66%) nor sound effects (80%). Collecting the target audience's feedback before producing the video was pioneering as it put the target audience's exact needs into consideration, leading the following orchestration process to be more purposeful, yet more complicated.

4.1.2. Stage II: Converting the discourse type from textual to oral

Based on our participants' preferences, the scripts were composed and featured simple, fact-based oral language. To communicate with a specified target audience in an authentic situation, we looked into the differences between written and oral language as laid out by applied linguists with regards to formality, depth, personal references, word choice, and syntax. Table 1 below presents the major differences in these five aspects when converting the written language in the magazine to the oral language for the video script.

For *formality*, the sentence describing the Pine Garden was originally phrased formally with a logical explanation that included the geographical information about the garden; but in oral language, it was marked by an informal conjunction "so," which initiated the introduction to the garden. As an introductory video, the *depth* of language may be compromised for an easier listening experience, allowing the visual presentation in the video to present how the landscape looked like. *Personal reference* was also one feature often used in films or videos. The written text that introduced the Yanzikou trail was a plain, objective description and the oral form using the personal pronouns as "we" and "you" made it more conversational and interactive. *Word choice* was an important indicator of the genre for a specific purpose. The statement describing the Amis people's different

festivals contained less common words "glamorous" and "celebration" and a long phrase "a variety of." In the oral form, the description was simpler with fact-based language, corresponding to the target audience's feedback. Lastly, sentences with complex *syntax* were considered a hindrance to comprehension and the revision was aimed to make sentences syntactically simpler and more straightforward, as shown in Table 1.

Table 1. Differences between written and oral language

Difference	Written language (magazine content)	Oral language (video script)
1. Formality	More formal Pine Garden: "The Pine Garden lies nearby in the plain and the south of Melian District at the north-eastern corner of Hualien city…"	Less formal Pine Garden: "**So**, **we** are going to learn about the Japanese colonial history in Hualien. The Pine Garden lies in the north-eastern corner of Hualien city…"
2. Depth	Greater precision and detail Qingshui Cliff: "Qingshui Cliff is a coastal cliff above sea level in Xiulin township, Hualien county. It is 21 kilometers in length and rises 24 kilometers from the Pacific Ocean…. The best view of this can be observed from the Suhua Highway, which crawls below the magnificent vertical cliff."	Less detailed descriptions Qingshui Cliff: "Hey! Look at the steep cliff and the clean water. Here is the Qingshui Cliff. Now we are on the platform of the Suhua Highway, and the best view of the cliff can be observed from here."
3. Personal reference	Fewer personal references (i.e., pronouns of the first- and second-person singular and plural) Yanzikou Trail: "The 1.4km long Yanzikou trail offers another equally heart-stopping experience."	More personal references (i.e., pronouns of the first- and second-person singular and plural) Yanzikou Trail: "Now we are on the Yanzikou trail. In case of falling rocks, you will need this. Now you are safe. Let's go!"
4. Word choice	Longer and less common words The Amis: "The Amis hold a variety of glamorous festivals for celebration. These festivals include Sea Festival, also known as Catching-Fish Festival, Harvest Festival, and other minor festivals."	Shorter and more common words The Amis: "They hold various festivals, such as Catching-Fish Festival, Harvest Festival, Water-Fetching Festival, etc."
5. Syntax	Syntactically more complex Taroko Gorge: "Each passage informed me that the gorge was a natural wonder, which has captivated the human race since its discovery."	Syntactically simpler Taroko Gorge: "The gorge has been a natural wonder since its discovery."

4.1.3. Stage III: Creating multimodal materials for composing the video

In this section, the analysis focused on what role the target audience played, how different modes were orchestrated, and most essentially, how the two interplayed during this stage. Our analysis showed that the multimodality framework enabled the research team to pay attention to a chosen mode while assembling other modes to strengthen the given mode.

Linguistic mode. The written discourse in the magazine has been extensively shifted to an oral genre for video scripts in relation to formality, depth, personal references, word choice, and syntax. However, the oral scripts would not be effectively presented without taking *the aural mode* into play. To familiarize themselves with the oral language during filming, the two hosts had to rehearse the oral script several times to make sure they could enunciate the words clearly and that they were familiar with their lines. Also, *to engage their target audience*, the two hosts added some impromptu phrases to make their speech more natural, for example, adding the impromptu sentence "Here is the Qingshui Cliff" to signal to the audience that the introduction to Qingshui Cliff was about to start.

Visual mode. To create the most appropriate multimodal materials for the most effective visual experience for the target audience, the research team assembled different modes to achieve the communicative purpose. For example, linguistic mode was added to attract our target audience's attention when talking about Taroko Gorge; one of the hosts opened the introduction by announcing, "Hi! We are on the Shakadang Trail, part of the Taroko Gorge. Let's go inside for an adventure." (Figure 1a). Another example is taking *spatial mode* into consideration

to enhance the best visual experience. For instance, when talking about the Amis's weaving techniques, one of the hosts stood in front of a weaving handicraft, and the camera captured the handicraft so that *the audience* could view the handicraft during the introduction (Figure 1b).

Aural mode. The aural mode was closely interrelated to the *visual mode* in the composing process because the more sources that the visual mode could capture, the more aural input there was to enrich the aural experience during the output stage (see Figure 1c, 1d, & 1e). For example, using both a camera and a smartphone not only expanded the visual experience but also the aural sources. In addition, to ensure that the *target audience* could experience the oral script (the aural mode) provided by the two hosts, they used voice amplifiers. Amplifiers were especially important when the surrounding was noisy, such as cicadas chirping in the trees (Figure 1e), as experienced in the Pine Garden. This also explains why the hosts did not begin videotaping until all people and cars had left the area.
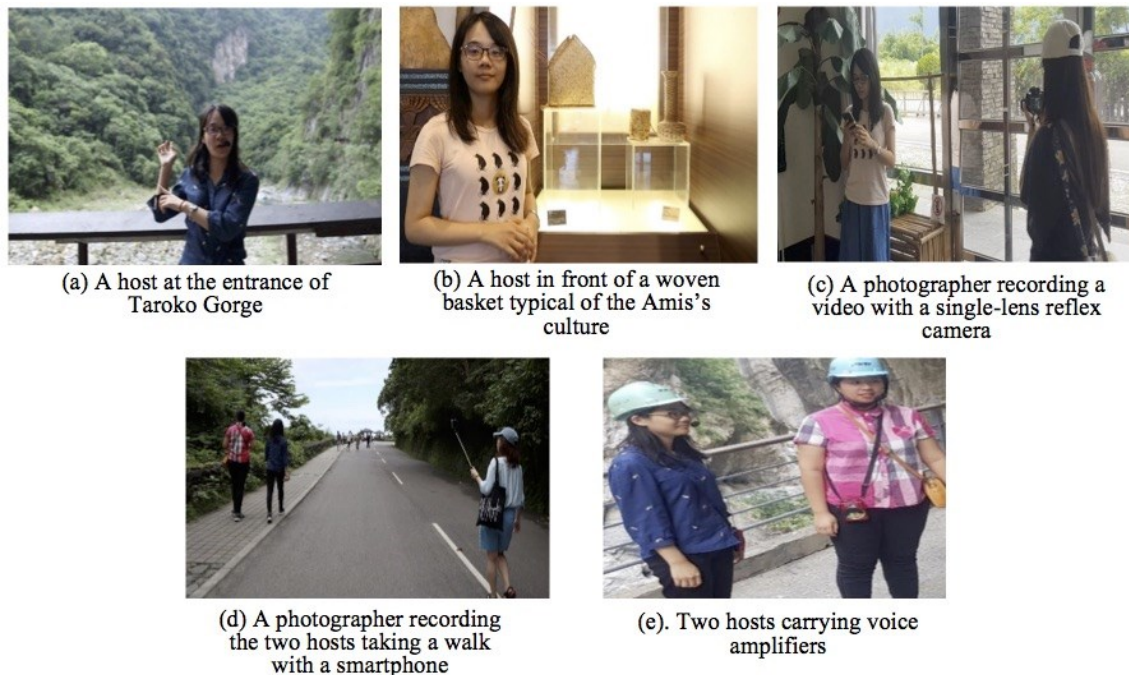

(a) A host at the entrance of Taroko Gorge

(b) A host in front of a woven basket typical of the Amis's culture

(c) A photographer recording a video with a single-lens reflex camera

(d) A photographer recording the two hosts taking a walk with a smartphone

(e). Two hosts carrying voice amplifiers

Figure 1. Multimodal materials including linguistic, visual and aural modes

Gestural mode. The target audience desired to see hosts in the film. Therefore, facial expressions, gestures, and movements became an essential part of the meaning-making process. For example, having the hosts to walk along the trail and to have a casual conversation was to foster a relaxing feeling, which once again was strengthened via an assembly of four other multimodal modes to achieve this communicative purpose. As always, to engage the target audience, the hosts made good use of gestures and facial expressions corresponding to the ambience set for the filming. For example, at the Ji-An Shrine, one of the hosts signaled the numbers 1, 2, 3 with her fingers (Figure 2a) to emphasize the three functions of the shrine to be introduced and smiled to demonstrate the amiability portrayed throughout the filming.

(a) A host speaking with gestures in a hall inside Ji-An Shrine

(b) A host reciting a poem about Taroko Gorge in the background

(c) A host standing on the right, instead of in the middle

Figure 2. Multimodal materials including linguistic, gestural, and spatial modes

Spatial mode. How the hosts position themselves on the screen in the filming process was the central issue informed by the spatial mode. At times, they decided to leave the whole screen/space to the audience. For example, when reciting a poem about Taroko Gorge, the photographer captured the natural scenery only and did not include the host so that the poem could be put on the screen (linguistic mode) with the natural scenery as the backdrop (visual mode) and the recitation (aural mode) as the background voice (Figure 2b). Even when the host was part of the screen, the major focus was on *the audience*. For example, when introducing the distinctive foods of the Amis, the host stood on the right instead of in the middle to make sure that the foods could be best viewed by the audience (Figure 2c).

4.1.4. Stage IV: (Re-)composing the two versions of the video

When the first version of our video was produced, an internal review by the research team found that three aspects of the video needed to be improved. First, the pace was too slow. Second, only one song was used for background music and it became monotonous; third, the color and size of the subtitles made reading difficult. Although the research team had the ability to self-critique the self-made video, they did not have the enough digital composing skills to resolve the problems identified. Therefore, they decided to turn to a video-editing professional who was well informed by the research team about the preferences of our target audience and the communicative purpose of the video to help create a second version of the video in order to eliminate the flaws identified.

Compared with the first version, the differences in the second version include using a new editing software to ensure the quality, maintaining the overall style throughout the film, changing the font of the subtitles, inserting transitional clips to different scenes, creating ear-catching sound effects, adding dazzling special effects, and adjusting the saturation and exposure of the output (see Table 2).

Table 2. Differences between the two versions of our video

Aspect	Version 1	Version 2
Editing software	PowerDirector 365	Final Cut Pro X
Overall style (visual & spatial mode)	A documentary coupled with funny elements at the END of the video clip	A documentary integrated with funny elements THROUGHOUT the video clip
Subtitle (visual mode)	English (font: Times New Roman) and Chinese (font: DFKai-SB) subtitles	English and Chinese subtitles (font: GenJyuuGothic)
Background music (soundtrack) (aural mode)	One song through the video	Seven different songs in the video (to mark different sections)
Transition (visual mode)	Not well-defined	Clean-cut
Sound effects (aural mode)	No	Yes
Functional special effects (visual mode)	Simple	Dazzling
Ornamental special effects (visual mode)	No	Yes
Color saturation (visual mode)	No adjustment	Adjustment of color saturation and exposure

4.2. RQ2: What can our target audience's responses to the videos inform us about the orchestration of multiple modes when transforming a magazine to a video?

The analysis of the target audience's responses to the two versions of the video informed us of three major insights into the orchestration of multimodal modes involving a target audience. First, the results show that content and format were interdependent during the orchestration process, greatly enhancing the effectiveness of the product. Second, the target audience's feedback served as critical information for the video creators to reconsider their multimodal meaning-making process. Finally, interdisciplinary collaboration with a professional manifested a new possibility for experiencing multiliteracies.

4.2.1. The content and the format as mutually conducive

Most of the interviewees rated the content of the second version ($M = 7.8$) much higher than the content in the first version ($M = 6.4$) even though the content about the six scenic spots was all generated from the first version (see Table 3) Why so? A short introductory video clip with funny and interactive elements incorporated with mini no-good (NG) clips was inserted before each scenic spot was presented while the NG clips could not be seen until the very last part of the video in the first version. When Interviewee 5 was asked why she preferred the Ji-An Shrine in the second version ($M = 8$) to that in the first version ($M = 6$), she responded that "the special effects and sound effects" of the second version and "the NG clips inserted throughout the video" made the content "more attractive" and "more appealing." Likewise, the interviewees liked the hosting style of the second version ($M = 7.7$) much more than that of the first version ($M = 5.6$), although the two hosts had exactly the same scripts, facial expressions, and movements. Interviewee 2 indicated that the hosting style in the second version was "a lot better" and "more relaxing thanks to the great editing." As indicated, the re-composing of the content through a different orchestration of multiple modes resulted in a better perception of the hosting style to be "more natural and appealing," (Interviewee 3) and "more interesting and attractive" (Interviewee 8).

Table 3. Target audience's responses to the videos

Item		First version (*M*)	Second version (*M*)
WHAT (content)	Qingshui Cliff	6.3	7.5
	Shakadang Trail	6.4	7.7
	Yanzikou Trail	6.3	7.6
	Pine Garden	6.0	7.6
	Ji-An Shrine	6.6	8.0
	Hualien Indigenous Tribe Museum	6.6	8.2
HOW (format)	Overall style	6.4	7.5
	Length	6.1	7.8
	Hosting style	5.6	7.7
	Subtitles	5.6	8.0
	Background music	5.5	8.7
	Sound effects	4.7	8.2
	Special effects	5.2	7.9

4.2.2. Gaining feedback from the target audience as renegotiation of meaning-making

When the target audience was first asked whether or not they would like to include special effects and sound effects as part of the video composing, the "no special effects" option won out (66%) in comparison with other choices provided, and 80% of them did not want sound effects. The major reason was that they were afraid that the effects might distract the viewers' attention to the content of the video. Interestingly, when the target audience was asked to express their preferences for the two versions they rated the second version ($M = 8.2$) (with added special and sound effects) noticeably higher than the first version ($M = 4.7$) featuring no effects based on the results of the survey. The special effects made the second version "fun and not too serious," (Interviewee 2), "funny and more entertaining," (Interviewee 5), and lent more "visual appeal" (Interviewee 6). The possible explanation for this discrepancy may be that the majority of the target audience might have expressed their preferences based on their previous experience of special and sound effects when viewing videos. Without the chance to compare and evaluate the two versions of the video, the target audience would not have learned that appropriate special and sound effects would not distract viewers' attention. Rather, it would enhance the interactivity between the viewers and the video, leading to a better engagement and viewing experience. Likewise, the research team would not have been able to learn that meaning-making was a recursive process of re-negotiation and re-adaptation had it not been for the involvement of the target audience.

4.2.3. Interdisciplinary collaboration as new possibilities for learning

The responses of the target audience highlighted the importance of interdisciplinary collaboration. Due to the insufficient digital composing skills, the composing process would have ended after the research team made their first trial had it not been for the collaboration with a professional with an expertise in video composing. Without the collaboration, the research team probably would not have been able to "fix" the problems even when they had the ability to identify the flaws of their own video composing. Second, the collaboration with a professional helped the research team realize how content could be received by the target audience differently with different

multimodal composing skills. The research team would not have known how many levels/possibilities of meaning-making could be achieved through assembling different multimodal modes and how their target audience would respond differently to the two versions of the video with different orchestration. The learning would not have been possible without the collaboration with a professional with the relevant expertise required for digital compositing.

5. Discussion

This case study, similar to Hafnar (2014) and Yeh (2018), yielded results leading to a four-stage framework based on the transforming process. It shares three features with the previous studies: composing the video script, filming the video, and editing the video. However, what distinguishes the framework here is the role of a target audience in a multimodal literacy project. A real target audience's involvement in the (re-)composing process enhances the overall effectiveness of the meaning-making and communication outcome. The multimodal orchestration process, in the meantime, provided the EFL learners a *capstone* learning experience beyond language per se, which may be examined by the framework of "*Learning by Design*" proposed by Kalantzis and Cope (2005).

5.1. The role of a target audience in a multimodal literacy project

Previous studies examining multimodal literacies in terms of composing (e.g., Campbell & Cox, 2018; Jiang, 2018) do not seem to specify a target audience, except for Hafner (2014) and Yeh (2018). While both Hafner and Yeh mentioned the target audience in their studies, the specific role(s) the audiences played in the "reading" stage of Hafner's model and the "composing stage" of Yeh's are not clear. In our study the authentic target audience offered their preferences for what they wished to see and for how the video should be made by way of a questionnaire based on the five modes in the multimodality framework and by elaborating on the results of the questionnaire in in-depth interviews. In short, our target audience had a direct impact on the process, as their choices and preferences shaped the content and format of our video.

Prior to the second stage, the research team had been well informed by applied linguistic studies that ample differences between oral and written language do exist (see Stage II); in addition, our target audience as key informants told us what type of oral language style (e.g., simple, fact-based oral language) they wished to see and hear in the video. Hafner (2014) also noticed that different people may draw on different "discursive forms (genres, registers, and styles)" (p. 657) to serve different purposes (e.g., written language in a lab report for an academic audience versus oral language in a Facebook update for friends). However, it was not clear whether the target audience(s) played a role in determining how language should be shaped to suit their interests and needs. Similarly, Yeh (2018) mentioned that her students "[used] English to introduce Taiwanese people, customs, cultural values, or architectural history" (p. 30) to "a global audience" (p. 36). There was not much elaboration on how their English was used in relation to the global audience in mind, e.g., genre, style, word choice. Our study shows a deliberate effort to draw our target audience into the preferred type of oral language for the video script.

Our target audience also served as critical reviewers in the final stage of video creation, providing key insights into the video-editing skills required for effective video production. Without the target audience reviews of the two versions of the video, it would have been difficult to ensure the effectiveness of the video production. Instead of seeking feedback from the target audience, Hafner (2014) asked his students to share their final videos online and obtain feedback from peers after the sharing session, while Yeh (2018) and her research team graded her students' video clips. In contrast, our target audience's ratings on and insights into the differences between the two versions of the video enabled a critical examination that helped understand what and how multimodal elements could be re-orchestrated to expand a viewing experience that our target audience had not expected.

5.2. The knowledge processes in multimodal orchestration

In terms of multimodal orchestration, the studies previously reviewed and the current one all presented an evolving process along which an effective video was produced. However, what made this case study stand out was the evaluation of how such orchestration processes also facilitated the knowledge construction processes conducted with an analytic framework including four elements: *experiencing the known and the new,*

conceptualizing by naming and theorizing, analyzing functionally and critically, and applying appropriately and creatively (Kalantzis & Cope, 2005).

The students embarked on this project by reviewing the magazine they had produced in an earlier course. The content was about a place of which the people and culture were already known to them. They had read and written about the place incorporating three modes (linguistic, visual and spatial) and presented it to their imaginary target audience in a magazine format as a case of print media. When enrolling in this capstone course, they determined to venture into *the new* by transforming the magazine into a video in order to reach a larger potential population which led to a series of, but not necessarily linear, processes of constructing knowledge.

Bringing their initial understanding and experience on multimodality and the concept of target audience, the students thus began to reach a group of authentic target audiences to collect detailed information for the later scripting, filming, and (re)composing processes. These processes helped students conceptualize and concretize their learned knowledge. First of all, the survey questionnaire was designed based on the five modes that have been theorized and applied widely in earlier research. Second, converting the discourse style from textual to oral language when scripting in terms of formality, depth, personal reference, word choice, and syntax was a new experience of practicing pragmatics for a specific target audience. This practice broadened and deepened their language learning in a new context. Next, in addition to applying the knowledge about language, multimodality, and target audience innovatively, they also experienced an interdisciplinary collaboration with a professional in filming which greatly expanded their perception of how the various multimodal semiotic resources should be taken into consideration when composing and recomposing the video.

Finally, the re-composition process was conducted due to a critical internal analysis by the students themselves with the help of a professional and an evaluative comparison between the two versions of the video given by the target audience. For our students, their multimodal orchestration competence built along the project enabled them to self-criticize the quality of the video, hence an improved version for the video was recomposed. For the target audience, they were able to reconsider their original requests (no sound and special effects) and accepted the new video as a better version. The new feedback given by the target audience, in turn, resulted in a new discernment in our students in that the target audience's preferences should be flexibly adapted and adjusted.

6. Conclusion and implications

Being situated in the 21st multimodal global society, literacy educators are confronted with the demand to equip their students not only with fundamental literacy skills (reading and writing) but also with multimodal literacies in order to achieve communicative purposes effectively. Hence, this case study investigated how a printed text (i.e., a magazine with still text-image compositions), as a common mode in the majority of EFL classrooms, can be transformed into a multimodal one (i.e., a video with moving text-image compositions) when involving a specified target audience. The results show that the transformation from a paper format to a digital video can be enhanced through the four-stage digital composing process, and the orchestration of multiple modes in a video can be best evaluated via the target audience's authentic comments and feedback. The findings also reveal that the EFL learners, as composers of the video in this case study, whose knowledge about digital composing process and multiliteracies could be (re-)conceptualized, expanded, and renewed when applying the multimodality framework. The whole process reported in this case study, thus, exemplified the knowledge processes proposed by Kalantzis and Cope (2005) and this *Learning by Design* framework also helped us examine and confirm such multimodal practices indeed facilitate knowledge construction and uplift a multiliteracies capacity in EFL learners. Three important implications for multimodal literacies are generated based on the major findings and discussion.

First, from a theoretical perspective, the results indicate that a well-rounded video composing process should not overlook the critical roles that a target audience can play throughout the different stages of the process. Thus, it implies that a target audience should be specified and integrated into the multiliteracies curriculum so that learners can work with the target audience to determine what to include, how to produce a video, and to what extent their video production is effective from the target audience's perspective. Second, the target audience's significantly different ratings on the two different versions of the video highlight the importance of multimodal competencies and skills. As language majors, who tend to have a better control of the linguistic mode, may not have the capabilities to integrate other modes. A multiliteracies curriculum should create a mechanism for interdisciplinary collaboration, for example, allowing design majors to take multiliteracies courses with language majors to facilitate such collaboration and to enhance the multimodal process and product. Lastly, the study indicates that the knowledge and skills to combine, orchestrate, or mix different modes should not be viewed as a

given even with a curriculum component featuring interdisciplinary collaboration. Ample multimodal learning opportunities should be designed and provided to engage EFL learners in multiliteracies activities so as to equip them with knowledge and competences to experience, conceptualize, analyze, and apply from the known to create new possibilities for multiliteracies education.

7. Limitations and further studies

This study was limited in three ways. First, the knowledge processes would have been more complete if a third version of the video had been created based on the target audience's ratings and feedback on the two versions of the video. Further research can investigate whether or not a third version of the video can provide an even more satisfying viewing experience and create more in-depth knowledge processes. Second, the professional invited did not involve in the four-stage transforming process. It will be interesting to examine if a different transforming process would emerge when a professional is invited to participate in the very beginning of the project, or when the research team is involved in a different classroom structure, such as working with class/teammates with multimodal expertise. Lastly, the four-stage model was developed based on a group of international students designated as the specified target audience in our study. The preferences for the five modes may be limited to the perspective of this specified target audience. It is therefore suggested that more than one group of target audience be involved in a case study to examine whether or not different preferences among different groups of target audience can be observed and how the differences impact each stage of the video composing process.

Acknowledgement

References

Alexander, J., & Rhodes, J. (Eds.). (2014). *On multimodality: New media in composition studies*. Urbana, IL: Conference on College Composition and Communication-NCTE.

Campbell, L. O., & Cox, T. D. (2018). Digital video as a personalized learning assignment: A Qualitative study of student authored video using the ICSDR model. *Journal of the Scholarship of Teaching and Learning, 18*(1), 11-24.

Cho, Y., & Choi, I. (2018). Writing from sources: Does audience matter? *Assessing Writing, 37*, 25-38.

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). A Case study approach. *BMC Medical Research Methodology, 11*, 1-9.

DeVito, J. A. (1965). Comprehension factors in oral and written discourse of skilled communicators. *Speech Monographs, 32*, 124-128.

Einhorn, L. (1978). Oral and written style: An Examination of differences. *Southern Speech Communication Journal, 43*(3), 302-311.

Gunel, M., Hand, B., & McDermott, M. A. (2009). Writing for different audiences: Effects on high-school students conceptual understanding of biology. *Learning and Instruction, 19*, 354-367.

Hafner, C. A. (2014). Embedding digital literacies in English language teaching: Students' digital video projects as multimodal ensembles. *TESOL Quarterly, 48*(4), 655-685.

Hung, S.-T. A. (2019). Creating digital stories: EFL learners' engagement, cognitive and metacognitive skills. *Educational Technology & Society, 22*(2), 26–37.

Jewitt, C. (2006). *Technology, literacy, and learning: A Multimodal approach*. London, UK: Routledge.

Jiang, L. (2017). The Affordances of digital multimodal composing for EFL learning. *ELT Journal, 71*(4), 413–422.

Jiang, L. (2018). Digital multimodal composing and investment change in learners' writing in English as a foreign language. *Journal of Second Language Writing, 40*, 60-72.

Kakh, S. Y., Mansor, W. F., & Zakaria, M. H. (2014). Rhetorical analysis tasks to develop audience awareness in thesis writing. *Procedia-Social and Behavioral Sciences, 98*, 806-813.

Kalantzis, M., & Cope, B. (2005). *Learning by design.* Melbourne, Australia: Victorian Schools Innovation Commission in association with Common Ground Publishing.

Kress, G. (2000). Multimodality: Challenges to thinking about language. *TESOL Quarterly, 34*(2), 337-340.

Kress, G. (2003). *Literacy in the new media age*. London, UK: Routledge.

Lin, T.-B., Li, J.-Y., Deng, F., & Lee, L. (2013). Understanding new media literacy: An Explorative theoretical framework. *Educational Technology & Society*, *16*(4), 160–170.

Martinez-Insua, A. E. (2019). Scientific writing and the contentfulness of subject themes. How science was explained to (lay) audiences. *Journal of Pragmatics, 139*, 216-230.

Mills, K. A. (2010). "Filming in progress": New spaces for multimodal design. *Linguistics and Education, 21*, 14-28.

Mills, K. A. (2011). *The Multiliteracies classroom*. New York, NY: Routledge.

Mills, K. A. (2015). Doing digital composition on the social web: Knowledge processes in literacy learning. In B. Cope & M. Kalantzis (Eds.), *A Pedagogy of multiliteracies: Learning by design* (pp. 172-185). London, UK: Palgrave Macmillan.

Monte Mǒr, W. (2015) *Learning by design*: Reconstructing knowledge processes in teaching and learning practices. In B. Cope & M. Kalantzis (Eds.), *A pedagogy of multiliteracies: Learning by design* (pp. 186-209). London: Palgrave Macmillan.

National Council of Teachers of English. (2013). *NCTE Framework for 21st Century Curriculum and Assessment*. Retrieved from https://secure.ncte.org/library/NCTEFiles/Resources/Positions/Framework_21stCent_Curr_Assessment.pdf

New London Group. (1996). A Pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60-92.

Pan, J. H., & Zhang, R. J. (2019). Preliminary investigation of the current situation and influencing factors of international students in Taiwan under the background of new southbound policy. *Taiwan Educational Review, 8*(2), 154-175.

Schallert, D. L., Kleiman, G. M., & Rubin, A. D. (1977). *Analyses of differences between written and oral language* (Report No. 29). Champaign, IL: University of Illinois at Urbana-Champaign.

Strauss, S., Feiz, P., & Xiang, X. (2018). Written versus spoken grammar. *The TESOL Encyclopedia of English Language Teaching, 1*, 1-8.

Wagennar, T. C. (1993). The Capstone course. *Teaching Sociology, 21*(3), 209-14.

Wilkinson, A. M. (1971). *The Foundation of language: Talking and reading in young children*. London, UK: Oxford University Press.

Wong, T. Y. (2005). Writers' mental representations of the intended audience and of the rhetorical purpose for writing and the strategies that they employed when they composed. *System, 33*(1), 29-47.

Yeh, H. C. (2018). Exploring the perceived benefits of the process of multimodal video making in developing multiliteracies. *Language Learning & Technology, 22*(2), 28-37.