Evaluating multimodal literacy: Academic and professional interactions around student-produced instructional video tutorials

Dacia Dressen-Hammouda, Ciara R. Wigham

To cite this article:
Dressen-Hammouda, D., & Wigham, C.R. (2022). Evaluating multimodal literacy: Academic and professional interactions around student-produced instructional video tutorials. *System*, 10.1016/j.system.2022.102727

Evaluating multimodal literacy: Academic and professional interactions around student-produced instructional video tutorials

Abstract: This paper offers a reflection on how academic and professional interactions can help guide best practices for constructing viable evaluation grids to assess multimodal literacy. Preparing English-language learners for today's digitally and culturally complex workplace environment is a central concern in English as a second language (L2) teaching environments. It requires meeting specific teaching goals, such as supporting traditional print and multimodal literacies as well as increasing learners' English-language fluency and appropriateness. Our study focuses on an underexplored professional multimodal genre – instructional video tutorials – and proposes a multimodal evaluation grid incorporating theoretical concepts and empirical results from multimodal linguistics and multimedia learning. We examine how four video communication professionals use the grid to measure the effectiveness of students' video tutorials and identify areas for improvement. We present results for three areas for which the experts considered students did not meet expected professional standards: information organization, timing, and L2 spoken language narration. Our findings suggest possibilities for introducing appropriate forms of action or intervention into teaching multimodal design projects to better prepare L2 English students to meet workplace multimodal literacy requirements.

Keywords: workplace-based multimodal literacies, multimodal evaluation grids, instructional video tutorials, collaborative needs-based analysis, technical communication, EMI (English as Medium of Instruction) teaching

1. Introduction

Increasingly, students in higher education must develop domain competence in a foreign language (L2) to accommodate face-to-face and computer-mediated workplace interactions to navigate the complex technology-mediated environments they are training for (Hewett & Bourelle, 2020). Accordingly, language students' pedagogical needs extend beyond school-based literacy for L2 writing and speaking instruction, toward harnessing the multiliteracy skills more appropriate to current workplace environments. To become competent contributors in their workplace cultures, students need to develop 'language-techno cultural competence' (Sauro & Chappelle, 2017) whereby they appropriate the modes of communication, ways of doing and professional identities specific to the workplace culture, in two or more languages.

In response, the L2 teaching focus has moved beyond grammar and vocabulary toward a literacies approach (Ware, 2017), which posits that students are better served if their L2 courses are designed around workplace-based and academic literacy practices integrating a variety of literacy types: computer, information, visual, multimodal. In this sense, the print-dominant view of literacy has shifted to a 'multiliteracies' perspective (Cope & Kalantzis, 2000, 2015) emphasizing "a combination of one or more elements of digital, multimodal, communicative and multilingual practices" (Ware, 2017: 267).

A crucial component of multiliteracies is multimodal literacy, a combination of skills which allows for "reading, viewing, understanding, responding to and producing and interacting with multimedia and digital texts" (Walsh, 2010: 213). As Ware (2017) observes, demonstrating multimodal literacy involves producing texts that blend together various combinations of semiotic resources (e.g., language, speech, sound, graphics, animation. Such texts' multimodal composition challenges traditional L2 competency assessment practices (Crawford-Camiciottoli & Campoy-Cubillo, 2018).

This paper explores the challenges posed by professional multimodal literacies in the context of an English medium of instruction (EMI) graduate program in technical communication, a field which requires a diverse range of skills. Europe-based professional technical communicators must possess good writing skills in two or more languages, including English. They must also demonstrate the

ability to coordinate and repurpose several semiotic resources – linguistic, visual, spatial, aural, gestural – across numerous professional genres including print-based user-guides, online help and instructional video tutorials. Preparing students for such multilingual and multimodal work environments can be challenging.

This paper seeks to fill a gap in the literature: despite growing interest (Hafner, 2014, 2018; Walsh, 2010), there is limited guidance on teaching three-dimensional multimodal professional genres such as instructional video tutorials, i.e., professional 'how-to' videos. Our study proposes a framework teachers may use to help students better master the multimodal skills required for the workplace, particularly with regard to video tutorials. Students may also use the framework for self-evaluation. To this end, we devised an evaluation grid created specifically for teaching and assessing this professional genre. It combines a multimodal (Cope & Kalantzis, 2000; Kress, 2010; Kress & van Leeuwen, 1996) and a multimedia-learning approach (Mayer, 2014a; van der Meij & van der Meij, 2013) to ensure video tutorial effectiveness for a professional setting. While combining two very different epistemologies – multimodality and multimedia learning – within a single evaluative framework may raise concern, we argue that they are complementary and mutually reinforcing for the purposes of teaching and assessing instructional video tutorials. In addition, our evaluative framework integrates procedural genre-based features (Ganier, 2004; Steehouder & van der Meij, 2005) and L2 spoken language criteria from the Common European Frame of Reference (CEFR). Although the CEFR references many of the above issues, research juxtaposing all these points is rare.

We first describe instructional video tutorials and establish our theoretical framework: multimodal linguistics, social semiotics, and multimedia learning. We then present our multimodal evaluation grid before addressing two research aims: (1) to determine which criteria are most problematic for L2 students when designing video tutorials in English, according to domain professionals; (2) to present a methodology for designing and validating future evaluative frameworks through collaboration between academics and professionals. We tested the evaluation grid with four video communication professionals for usability as an assessment tool, i.e., whether it describes actual professional practice. We propose the grid can be used for student self-evaluation of video tutorial productions, and by instructors for training and assessment purposes. In conclusion, we discuss the implications of our findings for designing and validating evaluative frameworks for multimodal student projects.

2. Theoretical background

2.1 Instructional video tutorials

Over the past 20 years, video has become a primary means of delivering 'how-to' information, thus contributing to the emergence of hybrid forms of specialized genres. In education, the number of how-to educational videos has progressed steadily (Bateman & Schmidt-Borcherding, 2018; Bateman et al. 2021; de Koning et al., 2018). Industry and the professions are no exception to this trend. Traditional print-based procedural genres (e.g., user manuals) have given way to instructional video tutorials in response to user demands for greater interactivity (informality, entertainment, encouragement, support) and exposure to richer channels of communication including spoken narration, animation and sound (van der Meij & van der Meij, 2013).

Instructional video tutorials are a dynamic "three-dimensional" (Lotherington & Jenson, 2011) multimodal genre. Given the video media's affordances, designers must pay careful attention to viewers' engagement (Mayer, 2014b; Mayer et al., 2004), working memory and cognitive load (Paas & Sweller, 2014). For example, because viewers often find video tutorials on the Web, they can click off at any time if video content or accessibility does not suit their purposes. Professionally-produced video tutorials thus need to convince viewers they will efficiently demonstrate how to carry out a task while inciting them to accept future interactions with the company. To devise an evaluation grid for teaching and assessing this professional multimodal genre, we have drawn on relevant literature summarized as follows.

2.2 Multimodal linguistics and social semiotics

Multimodal linguistics investigates the ways in which semiotic resources interact. It draws on semiotics, "the systematic study of systems of signs" (Lemke, 1990: 183), a field which traces back to Saussure's (1916) and Peirce's (1867-1871) work on signs and social meaning (Jewitt et al., 2016; Tan et al., 2020). One of its present-day corollaries, social semiotics (Halliday, 1978; Kress & Hodge, 1988), builds on the premise that meaning-making is organized and codified as sociocultural practices. In this perspective, all meaning is considered to materialize and be interpreted through the realization of semiotic resources (Kress, 2010), which van Leeuwen (2005: 3) defines as "the actions and artefacts we use to communicate", ranging from the vocal tract, to the muscles allowing for facial expression and gestures, to technologies, etc.

Scholars in this field consider that the semiotic resources within a system comprise options chosen among to make meaning. Although language has received the most attention, scholars have also investigated how other systems of semiotic resources – visual, aural, gestural, spatial – similarly make meaning. As Kress et al. (2000: 44) note, "visual communication, gesture, and action have evolved through their social usage into articulated or partially articulated semiotic systems in the same way that language has." They are thus resources for meaning-making in their own right, and depend on situational context and culture for salience and relevance (van Leeuwen, 2011).

The multiplicity of semiotic resources and the ways in which they combine to make meaning has been termed 'multimodality' (Kress, 2010), as reflected in various social semiotic approaches. These approaches emphasize different aspects of meaning-making, e.g., focusing more on grammar as in O'Halloran & Lim's (2014) approach to systemic functional multimodal discourse analysis, or like Kress (2010) and Kress and van Leeuwen (1996), attending to a more general social semiotic stance (see also Jewitt et al., 2016).

In our multimodal evaluation grid, we adopt the New London Group's (Cope & Kalantzis, 2000) five-part modal classification scheme, reproduced in Figure 1. As explained in the following sections, we have adapted this scheme slightly: we do not include the gestural mode given the focus of this paper. Likewise, we include the temporal mode to better account for the specific features of instructional video tutorials.

Instructional video tutorials

Modes	Linguistic/ alphabetic mode	Aural mode	Visual mode	Spatial mode	*Temporal mode*
Semiotic resources	- written and spoken words - word choice - grammar - sentences - paragraphs - text	- spoken words (i.e., narration) - sound - music - volume - rhythm - delivery speed - pitch - tone - voice	- images (still) - images (moving) - color - visual layout - design - font - size - formatting - symbols - visual data - graphic animation	- physical arrangement - proximity - spacing - organization - direction - distance between elements	- *length* - *speed* - *tempo* - *pacing* - *pausing* - *temporal cueing*

Figure 1. Modes and resources for designing video tutorials (based on Cope & Kalantzis, 2000)

2.3 Empirical guidelines for designing video tutorials

A multimodality framework accounts for the modes and semiotic resources that comprise video tutorials. Studies drawing on multimedia learning principles, however, provide empirically-driven specifications for how to marshal specific semiotic resources for effectiveness in video tutorials.

Multimedia learning principles, developed in cognitive psychology (Mayer, 2014a), offer general guidance on how to reduce cognitive load in working memory and improve learning when instructional messages combine linguistic and visual modes, like in video tutorials. Focusing on the operational limitations of human learning and memory, Mayer (2014a) describes these principles, e.g., extraneous, non-semantically related material should be avoided (*coherence principle*); information related spatially or temporally should be presented near one another (*contiguity principle*); visual and linguistic cues should draw attention to relevant information (*signaling principle*); information should not be repeated using resources from the same mode, such as writing and speech (*redundancy principle*); information should be integrated to avoid competing areas of attention (*split attention principle*); combining animation and spoken language helps avoid split attention (*modality principle*); in dynamic formats, information must be learner-paced (*segmentation principle*); in narration, informal conversational styles are preferred (*personalization principle*). Another effect not explicitly described in Mayer (2014a), but of relevance to our study on video tutorials, relates to coherence. Mayer and Moreno (2000) state that irrelevant sounds including music should be excluded from the aural mode to avoid split attention and support coherence, thus reducing cognitive load. In sum, multimedia learning research examines how these principles impact people's ability to comprehend and remember information, e.g., in video tutorials, fast-paced, loud music can interfere with viewers' ability to concentrate on narration, and thus learn to carry out a task.

Such research serves as a basis for the detailed guidelines on software training videos proposed by van der Meij and van der Meij (2013, 2016). Incorporating research on usability and minimalism (e.g., Lazonder & van der Meij, 1993; van der Meij et al., 2009; van der Meij & Carroll, 1995), their eight guidelines for video tutorial design and demonstration-based training summarize criteria for designing effective video tutorials: facilitating ease of access, combining narration and animation, enabling viewer interactivity with the video, previewing tasks, and allowing users to practice and review content. They also emphasize the need to present procedural rather than conceptual information, make tasks clear and simple, and keep videos short. Detailed heuristics, based on an extensive literature review, provide research-based explanations to help designers achieve these guidelines, and describe how to determine what makes effective titles, task-based content, screen interfaces, pacing, narration, animation, and designer-viewer interaction.

2.4 Multimodal evaluation grid

We integrate the above into our multimodal evaluation grid, organized around five modes – linguistic, aural, visual, spatial, temporal – and their semiotic resources (cf. Figure 1). We did not include the gestural mode in the present version of the grid; because narrators' on-screen presence in video tutorials is optional, our teaching did not focus on gestures. Also, as far as we are aware, time has not been singled out as a specific design mode in multimodal analysis, although it is a crucial element of video tutorials. Without claiming to modify the New London Group's (Cope & Kalantzis, 2000) five-mode schema, we have opted for time's inclusion as a mode in our grid inasmuch as it is a social construct through which meaning is organized and created in mode-specific ways (Andersen et al., 2015; Jewitt, 2008; Kress, 2010). In video tutorials, time and its semiotic resources – length, speed, tempo, narration pacing, narration pauses, temporal cueing – are of utmost importance. Designers must balance meaningful sequences of simultaneously occurring resources (e.g., animation, spatial and visual content, spoken narration) within manageable segments of information delimited by length to keep viewers interested while avoiding cognitive overload (Paas & Sweller, 2014). Given the importance of time in video tutorials, we have subsequently turned to research on multimedia learning principles for further indications about relevant temporal resources.

Multimedia learning-based studies fill this gap, for example by providing research-based recommendations about suitable video length. While Plaisant and Schneiderman (2005) propose 15-60 seconds and Chan et al. (2010) three minutes, van der Meij and van der Meij (2013: 221) underscore that length should be decided in terms of information organization: "arbitrary time limits … [are] hardly satisfactory. What matters more is that the user perceives a video as having a clear beginning and end." They suggest designers look for physical screen changes, indicating goal or sub-goal

completion, as meaningful moments for segmentation, or that they divide "the stream of information into smaller units with identifiable beginning and end points" (p. 221). They further recommend using narration pauses and temporal cues (e.g., then, next, now) to help viewers identify meaningful segments.

Other temporal resources identified in the video tutorial literature include speed, organized by tempo and narration pacing which allow the designer to match the "scenario of the unfolding instructional events" with the viewer's cognitive "resources and capabilities" (van der Meij & van der Meij, 2013: 212). Tempo, for example, is controlled by synchronizing actions through graphic animation and narration. Determining "the right speed for the user" (p. 212) is also linked to narration pacing: extending natural pauses and adopting a "conversational pace". In turn, narration pace can be controlled through verbal-pausing interactions, rate of speech and prosody.

Getting a sense of these complex interactions is a significant challenge for novice designers. Considering these observations, we ascribe to Alhadeff-Jones' (2017: 2) view that the interpretation of time, whether "fast/slow, early/late, retarded/advanced, or mature/immature... remains socially constructed" and is contingent upon gaining professional experience through extensive contact with viewers' genre expectations (Schriver, 2013). With regard to video tutorials, we consider that time is a socially-situated and learned concept, whose communicative purpose is to divide the video into segments and balance the amount of information by using temporal resources.

We transpose five modes, including the temporal mode, into a multimodal evaluation grid comprising 28 *modal competency criteria* (Figure 2), or statements about how to render semiotic modes and resources (cf. Cope & Kalantzis, 2000) specifically operational for instructional video tutorials, following multimedia learning principles including Mayer (2014a), van der Meij and van der Meij (2013, 2016), and others. For example, to minimize cognitive load and motivate viewers, videos should orally narrate demonstrated tasks rather than provide on-screen written text (Ayres & Sweller, 2014). Furthermore, human rather than machine narrators should address viewers using simple but appealing language (Mayer, 2014b). Narration should be delivered using a "conversational pace" (Morain & Swarts, 2010) to allow for appropriate video tempo and speed. Also, ensuring sound quality and avoiding extraneous auditory information (e.g. music, sound effects) is essential (Mayer & Moreno, 2000). Animations' spatial content and task sequences should be displayed as viewers would see them (Tversky et al., 2002). Additionally, designers must create congruence between screen capture animation and real-life task execution, help viewers perceive temporal changes, and provide easy-to-follow models enabling viewers to mimic observed actions (Bétrancourt, 2005; van der Meij & van der Meij, 2014). Similarly, cropping the window to show only content relevant to the immediate message helps avoid cognitive overload (van der Meij & van der Meij, 2013).

Given our objective of devising a multimodal evaluation grid for teaching and assessing student-produced video tutorials in an EMI technical communication graduate program, our grid also integrates L2 spoken language as a subset of the aural mode using CEFR spoken language criteria, targeting B2+ level descriptors (Council of Europe, 2018).[1] This choice is motivated by the observation that narration is a crucial part of video tutorials. Effective narration includes appropriate word choice and syntax but also phonological control, accent and prosodic features such as word stress, rhythm, and intonation. These features pose particular challenges to L2 speakers, requiring careful attention in an EMI teaching context. Finally, given video tutorials' inherently instructional nature, we also incorporate features of instructional-procedural genres (Ganier, 2004; Steehouder & van der Meij, 2005). We present our evaluation framework in Figure 2.

Although it may be expected for L2 spoken language criteria to be included in the linguistic mode, we have opted to incorporate them under the aural mode following Cope and Kalantzis' (2000) categorization schema, to highlight their role in creating effective narration.

[1] The CEFR criteria were adapted for use by non-language-teaching specialists, and were intended to reflect native-speaker intuition.

Modes	Modal competency criteria
Linguistic	1 - Video title contains a verb and object, and tells viewers the goal 2 - Video title allows viewers to locate and identify content using a search engine 3 - Video contains identifiable parts: introduction, preview, demonstration, conclusion 4 - Video is organized around a clear goal and sub-goals 5 - Transitions between sub-tasks are clearly demarcated using slides and narration 6 - Includes checkpoints in task completion for viewers to monitor progress
Aural: Narration, Sound & Music	7 - Narrates video using spoken human voice, not written text 8 - Narrational style is conversational, not formal 9 - Uses basic command verbs, short simple sentences 10 - Respects working memory, limiting information 'bursts' to four related actions 11 - Introduces new information 'just in time' (e.g., critical vocabulary, hints) 12 - Avoids extraneous verbal content 13 - Avoids extraneous auditory information, like music and sound effects
Aural: L2 Spoken Language	14 - Narrator expresses her/himself fluently. The flow of language is smooth 15 - Narrator clearly expresses her/himself in an appropriate style 16 - Consistently maintains high degree of grammatical accuracy. Errors are rare and hard to spot 17 - Pronunciation is accurate and does not hinder comprehension 18 - Intonation is appropriate
Visual	19 - Reflects the actual interface (what viewers see or do) 20 - Avoids extraneous visual information (e.g., window is cropped for relevant information) 21 - Animations show interconnection of viewer actions and system reactions
Spatial	22 - Key information is highlighted with visual pointing (mouse, circles, arrows, zooms, speed) 23 - Shows where the action happens, and what it looks like
Time	24 - Video length is meaningful and well-adapted to task: neither too much nor too little 25 - Animated actions and voice are synched; actions announced just before being demonstrated 26 - Narration uses a conversational pace, and extends natural pauses with 2-5 second pauses 27 - Video pace and tempo are the 'right' speed for viewers' working memory 28 - Viewers' control of content is facilitated (ex. tempo and pacing allow them to pause video)

Figure 2. Multimodal evaluation grid with modal competency criteria

To address our study's two research aims, in what follows, we examine which of the aforementioned modal competency criteria, and interactions between them, students had most difficulty with, according to domain professionals. We also test the validity of our multimodal evaluation grid using a small-scale reception study with experts in the field.

3. Study approach and methods

3.1 Research design

In our reception study (Tardy & Matsuda, 2009), the effectiveness of student-produced video tutorials was assessed by professional technical communicators using an online multimodal evaluation survey, which replicates the five modes and 28 categories shown in Figure 2. Insight into the professionals' thinking during the assessment was collected using online written feedback and stimulated recall interviews. Additionally, the interviews were used to gauge the professionals' observations about the evaluation grid's viability as a professional design tool.

Many applied linguists and writing researchers use reception evidence to improve analysis validity (Ceccarelli, 2005; Paul et al., 2001; Tardy & Matsuda, 2009), thereby adopting the fundamental assumption that the meaning of a message – including all forms of media – is not pre-given but interpreted by its recipient; meaning emerges from the context-dependent interaction between a text and an interpretative reader (Fiske, 1987). To determine the socio-cognitive relevance of a text for a target audience, reception studies incorporate empirical and social scientific methods, including quantitative assessments (e.g., citation histories) and qualitative methods, e.g., peer commentary and observational studies of reader response using think-aloud protocols, where readers are recorded as they verbalize their reactions to what they are reading. Subsequently, our study employs written feedback and stimulated recall interviews, in addition to an evaluative survey, as a means of eliciting

professionals' reception of the student-produced videos. Stimulated recall interviews were chosen as an introspective research procedure to invite participants to recall their concurrent thinking during the event (Mackey, cited in Lyle 2002), allowing participants to explain their decision-making while allowing for fairly unstructured responses (Lyle, 2002).

3.2 Participants

Four student project teams participated in the study. Of the ten student participants (8 female, 2 male), seven were L1 French speakers, two were L1 Mandarin Chinese speakers and one L1 Cantonese. Their level of English ranged from CEFR B1+ to C1[2] (Council of Europe, 2018).

Four L1 English-speaking technical communication professionals (P1-P4) evaluated the student-produced videos' effectiveness. All informants worked in Europe with 15-to-30+ years job experience. Aged 50-72, one informant was female, three were male. All held Masters' degrees in technical communication or journalism, had significant experience in producing and/or evaluating video tutorials, and worked daily with non-native English speakers. Participant names have been changed.

3.3 Materials

To address our two research aims, we employed one type of learning material (student-produced video tutorials) and created three measurement tools for the reception study (online multimodal evaluation survey, online written feedback form, stimulated recall interview guide). This section details each.

3.3.1 Student-produced instructional video tutorials

Prior to designing the videos, the students participated in a 12-week workshop during a third semester EMI graduate-level program at a French university. The 80-hour project-based workshop was designed to help them assimilate knowledge and skills about technical communication. Working in self-elected groups of 2-4 students to create a set of user-support genres for the presentation software Prezi, the project teams produced, in English, a paper-based user-guide, an online help system and a set of instructional video tutorials. Each of the four teams determined video content employing a user analysis conducted for their paper-based user-guide. Because the evaluation grid described in this study was developed subsequent to the workshop, it was not available to guide the teams' video designs.

After identifying a theme common to all four teams' video productions ('adding animations'), we selected three teams' productions for analysis; one video, produced by the L1 Chinese team, was excluded due to incomparable length (over 8 minutes).[3] The remaining three videos ran between 1'15 to 2 minutes, and were produced by L1 French teams (Appendix B).

3.3.2 Online multimodal evaluation survey

To address our first research aim of determining which modal competency criteria are most problematic for L2 students when designing video tutorials, an online multimodal evaluation survey reproduced the 28 criteria from Figure 2 to gauge the informants' immediate reactions after viewing the video tutorials. For each criteria, the domain professionals indicated their (dis)agreement using a six-point Likert scale.

3.3.3 Online written feedback form

The 28 modal competency criteria (Figure 2) were also reproduced in an online written feedback form, used by informants to comment on the three selected video tutorials.

---

[2] CEFR B1 level is the equivalent to intermediate level, TOEFL results of 57-86 or IELTS level 4 and CEFR C1 level is the equivalent to advanced English, TOEFL results of 110-120 and IELTS level 7.

[3] Although eliminated for this study, the inclusion of the video in future analysis may help us to determine whether conditions related to length or L1 may influence effective video tutorial design, which in turn may help us refine our evaluation grid.

3.3.4 Stimulated recall interview guide

Stimulated recall interviews were held with the informants post-evaluation. An interview guide (Appendix A) elicited information about training and professional background, experience with video tutorials, and working language(s). In addition to gaining insight into informants' assessments of whether the videos met professional workplace standards, it also sought their opinions about the evaluation grid's relevance and ease-of-use. This addressed our second research aim, presenting a methodology for designing and validating future evaluative frameworks through collaboration between academics and professionals.

3.4 Data collection procedures and analysis

Before viewing the three videos, informants were asked to preview the 28 modal competency criteria in the written feedback form. As we were interested in their appreciation of the grid's usability as a professional assessment tool, no prior training was provided to avoid influencing their reaction. After previewing the criteria, informants viewed the video tutorials, completed the survey and provided written feedback using the online form. Informants were contacted within 24 hours of completing the evaluation to organize the stimulated recall interviews, which occurred between 6 and 30 days later and were conducted via videoconferencing and recorded. Interviews lasted between 31 and 49 minutes and were automatically transcribed using Otter.ai.

The numbers obtained in the survey were merged to produce average values for each modal competency criteria.[4] We targeted the most negatively rated criteria for analysis, seeking insight from specific comments in the written feedback form. These comments and the stimulated recall interviews were analyzed thematically to gain a deeper understanding of the professionals' reasoning. Our analysis draws together the data sources.

4. Results and discussion

This section describes the four professionals' assessment of students' preparedness for meeting professional expectations regarding workplace-based multimodal literacies and English-language fluency.

Here we target three modes for analysis: the linguistic mode, the temporal mode, and L2 spoken language within the aural mode. We chose to focus on the first two modes because students' enactment of them was rated most negatively in the online survey. Moreover, although informants rated the L2 spoken language positively, written feedback and interviews pointed to underlying issues with students' spoken proficiency in English in a professional setting, which could impact the overall perceived quality of videos produced in the workplace.

4.1 Linguistic mode

Below, we discuss the scores obtained from the 28-criteria survey in terms of how students managed the linguistic mode (i.e., Criteria 1-6).

[4] Each criteria was rated on a six-point Likert-scale: 1=strongly disagree, 2=disagree, 3=somewhat disagree, 4=somewhat agree, 5=agree, 6=strongly agree.

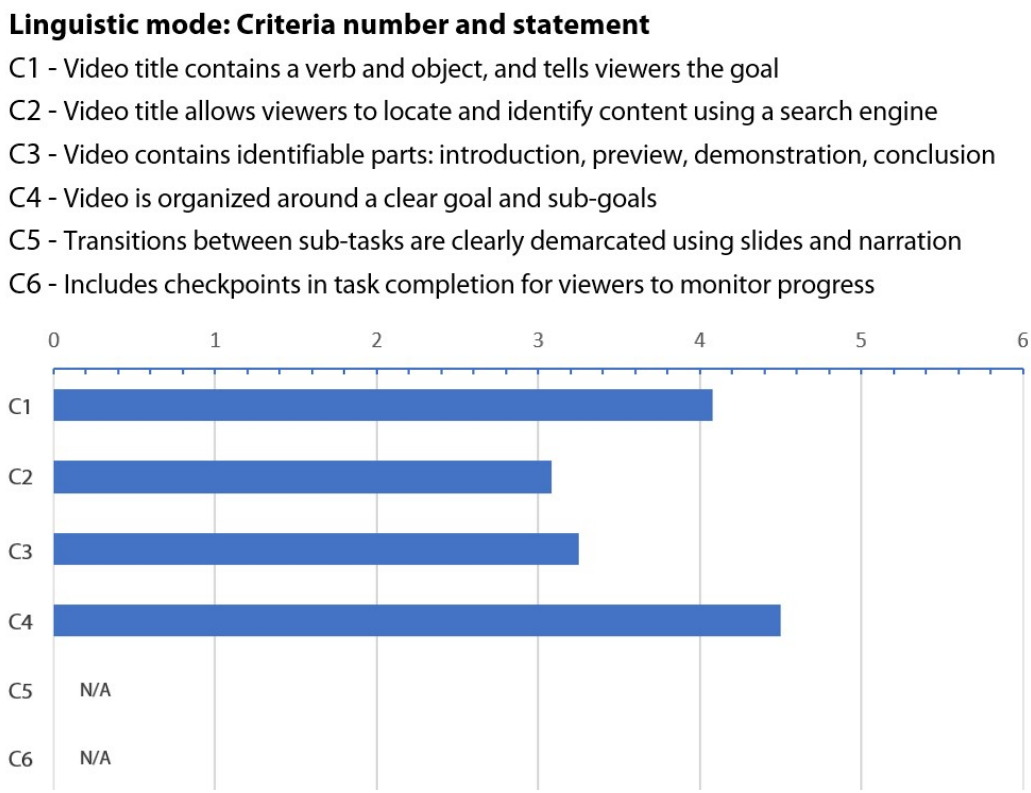


Figure 3: Survey results: Linguistic mode (C=Criteria) [4]

Figure 3 indicates that while informants appreciated the student videos' title presentation around an overall goal (C1=4.08), they felt the video titles needed more clarity (C2=3.08) and that the videos' sub-parts should be better identified (C3=3.25). Similarly, although the videos were segmented around a clear goal and sub-goal (C4=4.5), they lacked a navigational layer (C5=0, C6=0), e.g., displaying segmentation visually in the video timeline to help the audience better navigate within the video and view progress. This would have helped students better adhere to the principles of segmentation and contiguity (Mayer, 2014a).

Specifically, designers must keep in mind Mayer's (2014a) principle of spatial contiguity: the title and opening phrases will immediately impact viewers' decision to watch the video or not. However, informants reported that the students' video titles were too generic and lacked a distinctive search term identifiable by Web search engines. In videos 2 and 3, for example, the YouTube titles did not mention the software Prezi by name (Figure 4), nor did the students state the software's name in their introduction. Although video 2 (Figure 4) visually displays the software name, the meaning established through proximity between the word 'Prezi', the logo, and the opening-screen title occurs only after viewers access and begin watching the video.

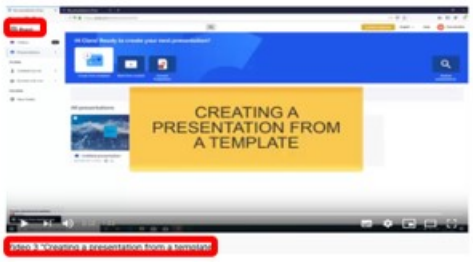
Timestamp	Visual mode	Aural mode (narration)
Video 2 0:00-0:07 Video title on YouTube		now that you've visually customized your presentation (00;13) you can add animations to it (02;03)
Video 3 0:00 - 0:03		(silence)

Figure 4. Disconnect between viewer search strategies and video findability

Tutorial viewers are not a captive audience, however; they can click off and look for answers elsewhere. This affordance reflects the need to create congruence between the tutorials and real-life tasks (Tversky et al., 2002). Professional informants focused on this aspect during stimulated recall interviews:

> P2: "[In] the first five seconds, if you realize that, oh, you didn't understand the title correctly, or that's not what you were looking to do, you can go off and find something else. [It's] really important, the first five or ten seconds. … [T]he word Prezi doesn't appear anywhere in the title .... So if I was searching, I'd be specifically looking for something on Prezi."

Similarly, informants commented that the narration did not contextually preview the video's purpose, necessary for convincing viewers that the video tutorial is relevant to their needs. Instead, students skipped directly to a demonstration of steps. This is illustrated in Figure 5 (lines 1-4), which shows the students' use of the linguistic mode (word choice, text) across visual and aural modes; the aural mode is further broken down into narration and sound/voice.

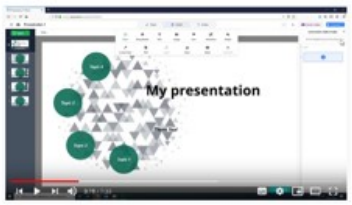
Line # and Timestamp	Visual mode	Aural mode (narration)	Aural mode (sound/voice)
1 0:00-0:07	 *Static screen*	now that you've visually customized your presentation (00;13) you can add animations to it (02;03)	Instrumental music plays softly
2 0:07-0:12	 *Mouse moves up to center of screen to show insert button*	(01;00) click on insert (02;31)	[0:10-0:12] Voice shouts "I feel good!"
3 0:12-0:16	 *Dialogue box opens, mouse moves to show animation button*	(01;01) click on animation (03;06)	Instrumental music plays more loudly
4 0:16-0:21	 *Right window appears, mouse moves right to show choice options*	(00;97) choose where you want to add your animation (02;00)	

Figure 5. Lack of appropriate task preview and contextualization in video 2

As professional informant P2 commented:

> P2: "They just launched into it. … I would like to see a little bit more structure in terms of, this is the feature of the product. This is what it does. Here's an example of what you can create with it. And this is how you go about doing that. None of them actually had that structure."

Indeed, all informants emphasized how much video tutorial viewers rely on the contextual setting, which must 'hook into' whatever moment viewers are watching the video:

> P2: "Video 1 starts off with, today we will be presenting to you x y z, right. So, given you don't know when this video is going to be viewed, it's not today when you're recording it that someone's going to be watching it, you just have to disassociate it with any particular moment in time because it's supposedly ever-green."

Two other issues raised for the linguistic mode include the need for a clear task-progression throughout the video, and to conclude the video by reviewing the steps, as described in van der Meij and van der Meij's (2013) guidelines:

> P1: "There is no, what do you do next? So the end of the video is kind of like 'Thanks for watching'. And, you know, that was great, but I need to know what to do next."
>
> P2: "The instruction bit should be broken down into one, two, three or four actions and then some kind of feedback loop on what those actions have accomplished. … And that's one thing I felt was missing, like one of the videos just seemed to be a random collection of three or four different things you can do with the software, but there was no progression towards the goal…."

4.2 Temporal mode

Another mode flagged by professional informants is the temporal mode, in terms of length, speed, pacing and narration pauses (Criteria 24-28).

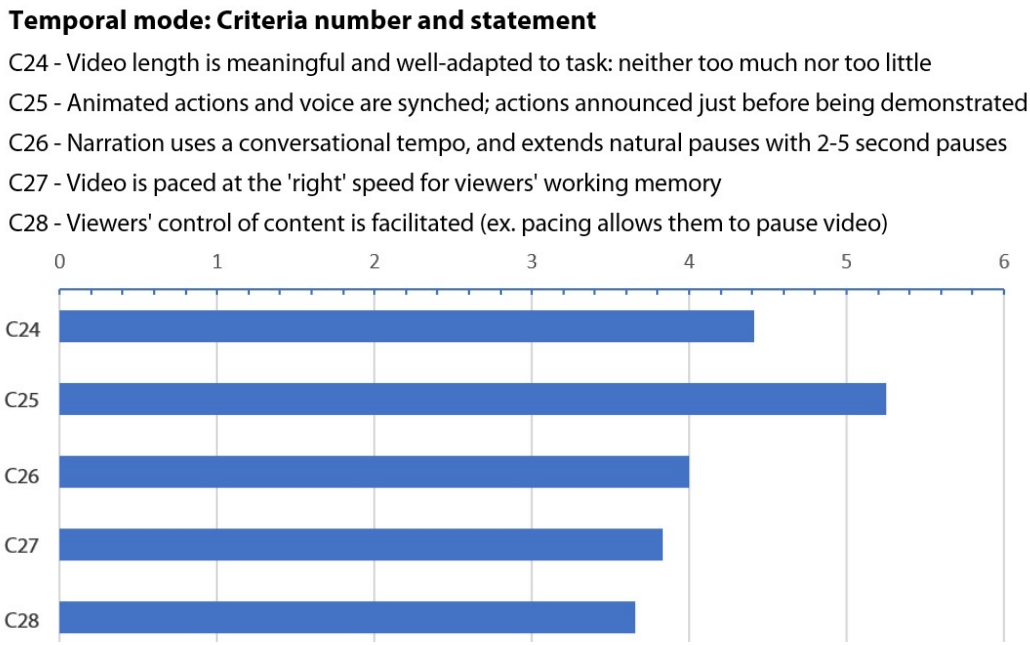
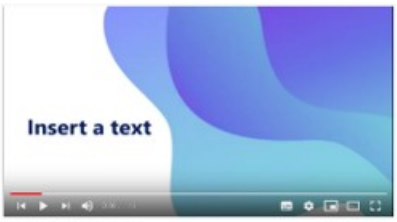
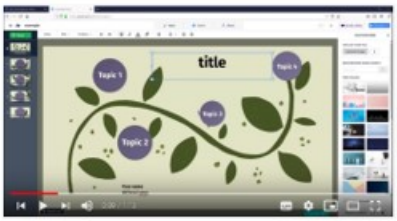
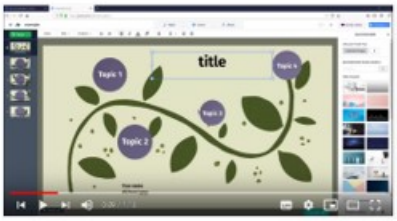
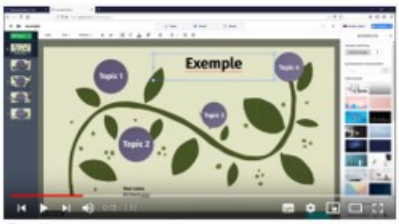


Figure 6. Survey results: Temporal mode (C=Criteria) [4]

Figure 6 shows that informants favorably evaluated students' synchronization of animated actions and voice (C25=5.25), a feature of tempo used to manage speed and enhance comprehension in multimedia presentations (Mayer, 2014b). However, students' skill in managing other temporal resources was less productive. Video tutorial speed (C27=3.83) and viewer control of content (C28=3.66) raised concern. Moreover, informant P4 highlighted the need to pause after each tutorial step (C26=4.0) so that viewers can assimilate the information and/or stop the video to execute the step.

Informants were sensitive to conversational pace (Morain & Swarts, 2010). They considered that video 2's (cf. Figure 5) overall speed, pacing and tempo could have been faster. Conversely, video 1 felt rushed; its speed was too fast to allow viewers to follow the steps. Figure 7 shows the difficulties students had managing the temporal mode in video 1, as illustrated by the short timestamp and narration pauses, and fast-paced music.

Line # and Timestamp	Visual mode	Aural mode (narration)	Aural mode (sound/voice)
1 0:00-0:05	 *Static screen*	today we are presenting you a video about how to modify the features of your Prezi presentation (00;13)	Fast-paced, instrumental music plays throughout
2 0:05-0:08	 *Static screen*	first we'll explain to you how to modify a text (00;13)	
3 0:09-0:11	 *Static screen capture*	choose the text you want to modify (00;28)	
4 0:11-0:13	 *No screen change from line 3*	double click on the area (00;25)	
5 0:13-0:18	 *Static screen capture*	highlight the text and type on your new text (03;14)	

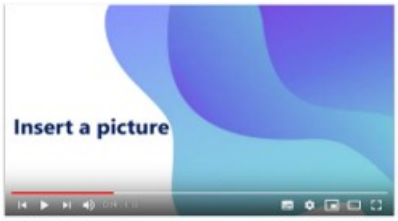
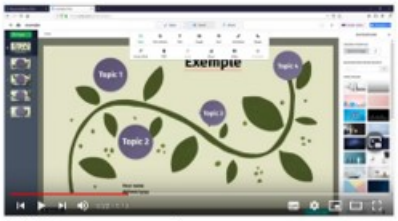
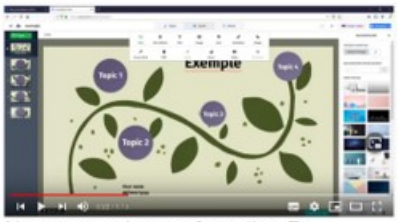
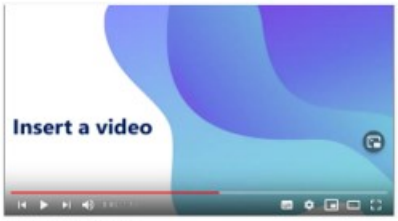
6 0:19-0:22	 *Static screen*	then (00;04) let's see how to insert an image (01;15)	
7 0:22-0:26	 *Static screen capture*	to do so click on insert on the upper menu (00;19)	
8 0:26-0:37	 *No screen change from line 7*	choose image (01;00) on the right you'll have the choice between choosing the free images from the web or downloading an images from your computer	
9 0:37-0:40	 *Static screen capture*	(silence)	
10 0:40-0:42	 *Static screen*	now (01;00) you may want to insert a video (02;00)	

Figure 7. Difficulties in pacing and tempo (video 1)

Contrary to Tversky et al.'s (2002) results underlying the need to make temporal changes obvious, video 1's narration provided almost no pauses between segments, temporal cues were used only to introduce new tasks (Figure 7: lines 2, 6, 10), and contrary to recommendations by Mayer and Moreno (2000), fast-paced music occurred throughout:

> P3: "In the first video […] there's no breathing space. It just jumps from one [task] to the other. And you're not sure why suddenly the subject is different. And it would have been perfectly fine to have just a beat of time in between."

Accentuating narration pauses are important for not overburdening viewers' cognitive load (Paas & Sweller, 2014) and in allowing for viewers' understanding to emerge:

> P4: "When you're doing video and particularly a tutorial, you have to slow down and take

what we would feel like are unnatural pauses, but for the viewer, they're necessary. Anything you're doing on video, you have to over articulate and over pause. ... [Timing] is very important, even though it may feel unnatural, it doesn't feel unnatural necessarily when you're watching it. It also gives the user the opportunity to hit the pause button. And it's difficult to hit the pause button when someone jumps from one step to the next as though it's one continuous sentence."

P3: "Time is critical, and people need time to absorb. … You have to assume that people are only going to get one exposure. So you have to make sure that the density of your information is just right."

The temporal mode was challenging for students, in terms of choosing relevant length, managing information density by segmenting tasks and sub-tasks through pausing and temporal cues, as well as controlling speed and narration pacing. Students appeared to systematically apply stringent length criteria (Plaisant & Schneiderman, 2005; Chan et al. 2010), rather than construct it using adaptive temporal resources (van der Meij & van der Meij, 2013). These difficulties also impacted their L2 spoken language, as discussed in the next section.

4.3 L2 spoken language within the aural mode

In this section, we identify specific L2 spoken language concerns from the professional informants' point of view (Criteria 14-18). Our decision to focus on this set of modal competency criteria is motivated by comments informants made in written feedback and during stimulated recall interviews.

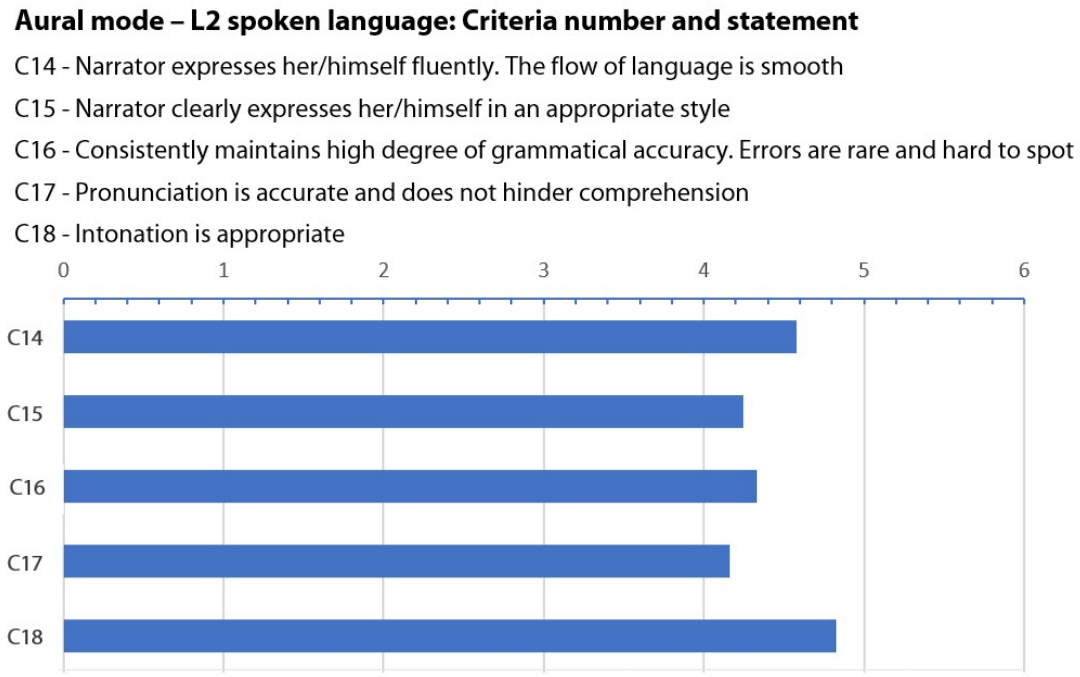


Figure 8. Survey results: Aural mode L2 spoken language (C=Criteria) [4]

Figure 8, reporting survey results for L2 spoken language, shows that informants rated the five spoken language criteria favorably. Flow and intonation (C14=4.58, C18=4.83) were rated somewhat more positively than style and grammar (C15=4.25, C16=4.33). L2 accent, although detectable, was not considered overly distracting (C17=4.16):

P3: "The person speaking was saying 'template' [*templet]. Because it was repeated and because it was close enough, you get it after about the second or third time."

Although hesitant to criticize the students' language proficiency, all four informants revealed concerns

in written feedback and during interviews about how prosody interacts multimodally with other modal competency criteria: narration pace and pauses (temporal mode), video length and speed (temporal mode), animation (visual mode) and conversationality (aural mode).

Drawing on the features of conversational English to create a dialogue with viewers is important for establishing the video's legitimacy, as captured by the personalization principle (Mayer, 2014b, Mayer et al., 2004). It allows for a motivational relationship to emerge with viewers, who intuitively become more receptive to the message conveyed by the video's narrator. Prosody, which includes rhythm, intonation and word stress, plays a crucial role in this process. Because video tutorial narration should largely mirror natural speech, getting the "rhythms, inflections and intonations" of conversational English right was identified by informants as a key feature of effective video tutorials.

The genre frame appeared to be a source of difficulty in this regard. Professionally-produced video tutorials should be perfectly calibrated performances that seamlessly orchestrate a balance between verbal and visual content, timing, audience awareness, cognitive load and social interaction. Students found themselves having to project spoken narration for an audience they had difficulty imagining while trying to conform to procedural-genre expectations, such as being clear and brief (Ganier, 2004; Steehouder & van der Meij, 2005). Consequently, their efforts to strive for clarity (van der Meij & van der Meij, 2013) and not overload working memory (Paas & Sweller, 2014) seem to have caused them to lose control over some prosodic features through 'hyper-correct' narration (Figure 9, video 2).

In video 2, students applied themselves to making the narration easy to follow: the narrator spoke slowly, over-enunciated, and used short, simple sentences (Figure 9, showing the L2 spoken language focus in the aural mode). This hyper-correctness led informant P2 to state in written feedback that the prosody was inappropriate: the narration was "quite formal and monotone" and should have been "more conversational, faster, and expressive." As informant P4 explained:

> P4: "It's okay to have inflection and emotion. When you're doing a video tutorial, you don't have to have your voice really flat. [Being] passionate ... comes across in your voice [and] subconsciously makes [viewers] very interested in the subject."

This issue coincided with the narration pace being too slow and long pauses occurring between step demonstrations (Figure 9: lines 2-3). Students used the pauses to animate narrated steps, causing a perceived redundancy (Mayer, 2014a) between animation and narration:

> P1: "They would read the screen to me, they would read the options to me, they would say okay, click here and here are the options that you have and I go, I can see that, give me something else."

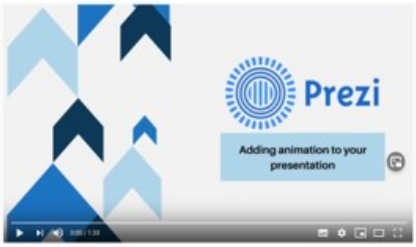
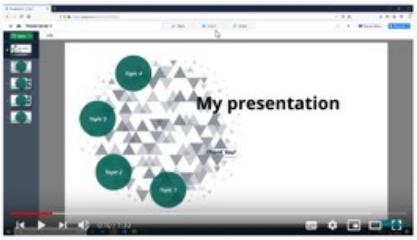
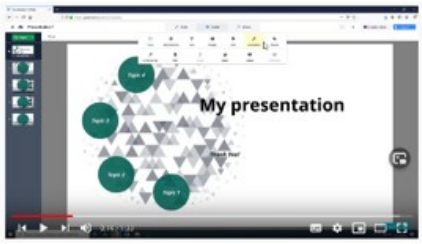
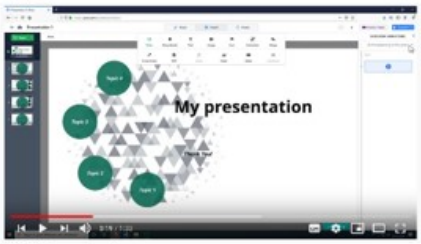
Line # and Timestamp	Visual mode	Aural mode (narration/ L2 spoken language)	Aural mode (sound/voice)
1 0:00-0:07	 *Static screen*	now/ that you've visually customized [*ku:s.tə.maɪzd] your presentation/ (00.13) you can add animations to it (02;03)	Instrumental music plays softly
2 0:07-0:12	 *Mouse moves up to top of screen to show insert button*	(01;00) click on insert\ (02;31)	[0:10-0:12] Voice shouts "I feel good!"
3 0:12-0:16	 *Dialogue box opens, mouse moves to show animation button*	(01;01) click on animation\ [*æˈn.ɪˈmeɪ.ʃən] (03;06)	Instrumental music plays more loudly
4 0:16-0:21	 *Right window appears, mouse moves right to show choice options*	(00,97) choose where you want to add your animation (02;00)	

Figure 9. Narration hyper-correctness affects intonation and prosodic features[5]

Similarly, over-enunciation occasionally accentuated the narrator's L1 French phonology (Figure 9: lines 1, 3). Finally, contrary to Mayer and Moreno (2000) who recommend not overloading the audio with extraneous information including music, the inclusion of a voice-based sound track disconcertingly drowns out the narrator's more subdued tone (Figure 9: lines 2-4).

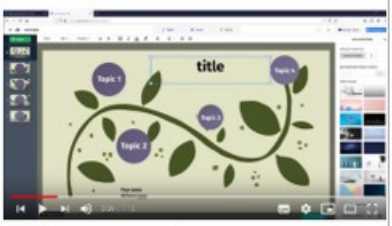
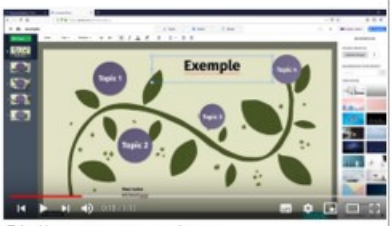
Keeping the video 'short' can also affect a narrator's control of rhythm, intonation and word stress. We see this in video 1:

P2: "[T]he narration goes too quickly. In one case, the last task starts without any break after

[5] Key: / rising intonation; \ falling intonation; * incorrect pronunciation or word stress shown in transcription; pause in (ss;x/30s)

the previous one finishes: you don't have time to absorb what words have been spoken."

The quick video speed caused significant problems for the narrator's word stress and intonation, which rose and fell in unexpected ways (Figure 10, video 1). Normally, a falling intonation signals the end of a turn or step, so viewers would expect a screen change. However, falling intonation occurred repeatedly mid-sentence without an accompanying step or screen change (Figure 10: lines 1, 5, 7, 8).

Line # and Timestamp	Visual mode	Aural mode (narration/ L2 spoken language)	Aural mode (sound/voice)
1 0:00-0:05	 *Static screen*	today/ we are presenting you a video\ [vɪ'd.i.əʊ/] about how to modify/ the features/ of your Prezi/ presentation\ (00;13)	Fast-paced, instrumental music plays throughout
2 0:05-0:08	 *Static screen*	first we'll explain to you/ how to modify/ a text\ (00;13)	
3 0:09-0:11	 *Static screen capture*	choose the text you want to modify\ (00;28)	
4 0:11-0:13	 *No screen change from line 3*	double click/ on the area\ (00;25)	
5 0:13-0:18	 *Static screen capture*	highlight the text\ and type on your/ new text\ (03;14)	

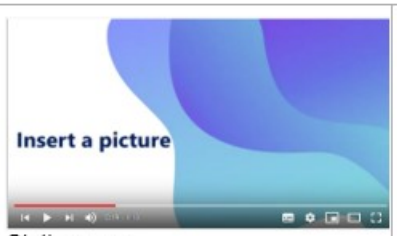
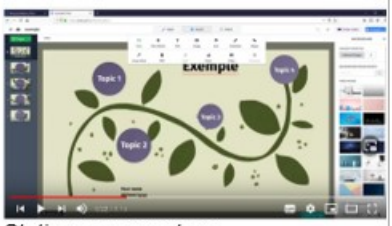
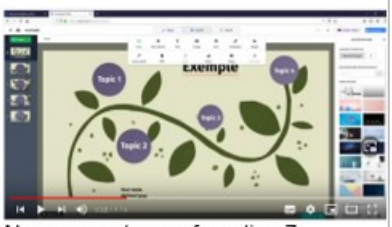
6 0:19-0:22	 *Static screen*	then/ (00;04) let's see how to insert [*ɪnsɜːˈt] an image\ [*/ɪmˈeɪdʒ/] (01;15)	
7 0:22-0:26	 *Static screen capture*	to do so/ click on insert\ on the upper menu (00;19)	
8 0:26-0:37	 *No screen change from line 7*	choose image\ [*/ɪmˈeɪdʒ/] (00;29) on the right/ you'll have the choice between choosing the free (*θriː) images[*/ɪmˈeɪdʌʒ/] from the web/ or downloading/ an images\ [*/ɪmˈeɪdʌʒ/] from your computer	

Figure 10. Attention to video length affects intonation and prosodic features[5]

The narrator's phonological control in English ("Accent is rather strong, it's sometimes difficult to capture what is being said", P3) further compounds these issues, potentially interfering with the conversational tone narrators are expected to use in video tutorials (Mayer, 2014b; Mayer et al. 2004). Informants stressed the importance of having a native English speaker do the narration for professional video tutorials, when possible. If not, the narrator must have a good sense of the language's 'music and rhythm':

> P3: "You have to get the tonic accent, right? You have to get the music of the language, right? And no, if you don't get that right, very often people won't understand what you're saying. And why make it hard for them? Especially since they can just click off your video anytime they want. I would always say if … you can get a native speaker to do the narration, that's always best. Second best is somebody who really gets the rhythm."

In sum, our informants' observations about how narration pace, including rate of speech, narration pauses and prosody (rhythm, intonation, word stress) affects tutorial viewers' perception of effectiveness align closely with research on social cues in multimedia presentations. Video designers can enhance their viewers' motivational commitment by using social cues, communicated through the narrator's attention to conversational style (Mayer, 2014b; Mayer et al., 2004). Therefore, attending to the key features of conversational English is of vital importance in video tutorials because it "primes the activation of a social response in [viewers], such as the commitment to try to make sense of what the speaker is saying" (Mayer, 2014b: 346).

5. Conclusion: Implications and practical applications

Our study establishes a set of research and workplace-based criteria to determine what students and teachers could attend to when designing or providing feedback on dynamic multimodal genres. Our purpose-made multimodal evaluation grid includes 28 modal competency criteria organized around five modes (linguistic, aural including L2 spoken language, visual, spatial, temporal), allowing for a

better targeting of critical design areas.

To address our first research aim, this paper has described professionals' assessment of student-produced video tutorials using the grid, and their identification of areas of difficulty for students. While prior work has identified the modes and resources (e.g., Cope & Kalantzis, 2000) which can be used to produce video tutorials, and other studies have provided specific design recommendations using cognitive principles and end-user appreciations of usability (Mayer, 2014a; van der Meij & van der Meij, 2013), our study combines these research areas to better examine the issues novice designers encounter when producing video tutorials. Results show students' difficulties in three areas. Regarding the linguistic mode, choosing identifiable titles, previewing the task, and specifying task progression and step review were problematic. Moreover, establishing appropriate length through tempo, pacing and narration pauses proved elusive. Spoken language was also a concern in terms of prosody, with a breakdown of intonation control due to perceived genre expectations.

While relying on empirical research to describe workplace-based genre expectations is essential, validating those expectations using a guided reception study with domain specialists is equally important. Thus addressing our second aim, our evaluation grid has been tested by field experts to engage a dialogue with the targeted professional community (Tardy & Matsuda, 2009) and foster the development of more authentic teaching materials, thereby improving pedagogical and design guidance for instructional video tutorials. Notably, the experts offered insight into how to make feedback meaningful for the professional workplace. They helped refine criteria wording and identify other criteria to include. Moreover, written feedback and interviews allowed us to understand some of the more sensitive issues at play, such as unspoken attitudes about L2 proficiency which professionals may be hesitant to go on record with when assessing student-produced video tutorials, but which affect overall quality nonetheless.

In terms of study limitations, we take pains to emphasize that, given our goal of showing how empirically-driven recommendations from multimedia learning can be integrated as modal competency criteria within a multimodal framework, and of making this information accessible to novice designers, the evaluation grid described in this paper is intentionally non-exhaustive in terms of the numerous semiotic resources that give form to instructional video tutorials. It has undoubtedly paid less attention to resources like grammar and word choice more common in systemic functional multimodal analysis (O'Halloran, 2008; O'Halloran & Lim, 2014) and to visual and spatial semiotic resources, such as color, proximity and visual cueing direction, as addressed in a social semiotics approach to multimodality (Kress & van Leeuwen, 1996). Further research could elaborate on these and other outstanding issues in criteria determination, including the temporal resources which strongly impact the enactment of other modal competency criteria.

Undoubtedly, the students from our study would have benefitted from an articulated set of guidelines to guide their design process (Goodrich Andrade, 2005; Brookhart & Chen, 2015). This and the foregoing observations lead us to consider appropriate forms of pedagogical action that could be introduced into course design to better prepare students to meet workplace requirements regarding multimodal literacies and L2 fluency, and better accompany their integration into the professional community by raising their awareness of professionals' standards and expectations.

In ongoing work (Dressen-Hammouda & Wigham, 2022a, 2022b), we reemploy the research data collected as pedagogical materials. Awareness-raising activities integrating our data include a course task in which students analyze the strengths and weaknesses of the student-produced videos reported on in this study, and confront their opinions with professionals' by showing them segments from stimulated recall interviews. We anticipate that this form of *a posteriori* peer review could help students-in-training become more conscious of the unspoken codes of professional practice learned mostly through on-the-job experience, in turn better allowing for a more critical awareness of multimodal literacies (Dressen-Hammouda & Wigham, 2022a).

In another study, we employ the revised multimodal evaluation grid as a self-assessment tool to guide students' video tutorial productions and support their capacity to produce professional-standard level

work. A replication study with the same informants compares student video tutorial productions with and without access to the grid during design (Dressen-Hammouda & Wigham, 2022b). Other research avenues could investigate how the grid may help teach students the differences between paper-based and related three-dimensional genres.

Our ultimate aim is to render the various skills involved in successful multimodal design more "teachable" by articulating them with more explicit accounts (Bateman, 2008), thus helping students build a clearer mental image of users and tasks. The ability to call on strong mental images is one of the foundations of information design expertise (Schriver, 2013), a learned skill for which students require varied types of pedagogical support.

References

Alhadeff-Jones, M. (2017). *Time and the Rhythms of Emancipatory Education. Rethinking the Temporal Complexity of Self and Society***.** Routledge.

Andersen, TH., Boeriis, M., Maagerø, E., & Tønnessen, E.S. (Eds.). (2015). *Social Semiotics: Key Figures, New Directions*. Routledge.

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R.L. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning,* 2nd edition (pp. 206-225). Cambridge University Press.

Bateman, J.A., & Schmidt-Borcherding, F. (2018). The communicative effectiveness of education videos: Towards an empirically-motivated multimodal account. *Multimodal Technologies and Interaction*, *2*, 27. https://doi.org/10.3390/mti2030059

Bateman, J.A., Thiele, L., & Akin, H. (2021). Explanation videos unravelled: Breaking the waves. *Journal of Pragmatics*, *175*, 112–128.

Bateman, J.A. (2008). *Multimodality and Genre. A Foundation for the Systematic Analysis of Multimodal Documents.* Palgrave Macmillan.

Bétrancourt, M. (2005). The animation and interactivity principles. In R.E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 287-296). Cambridge University Press.

Brookhart, S., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review, 67,* 343-368.

Ceccarelli, L. (2005). A hard look at ourselves: A reception study of rhetoric of science. *Technical Communication Quarterly, 14,* 257-265.

Chan, L.P., Patil, N.G., Chen, J.Y., Lam, J.M., Lau, C.S., & Ip, M.S.M. (2010). Advantages of video trigger in problem-based learning. *Medical Teacher, 32,* 760-765.

Cope, B., & Kalantzis, M. (Eds.) (2000). *Multiliteracies: Literacy Learning and the Design of Social Futures*. Routledge.

Cope, B., & Kalantzis, M. (Eds.) (2015). *A Pedagogy of Multiliteracies: Learning by Design*. Palgrave Macmillan.

Council of Europe (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Companion volume with new descriptors. Cambridge University Press.

Crawford Camiciottoli, B., & Campoy-Cubillo, M. (Eds.) (2018). The nexus of multimodality, multimodal literacy, and English language teaching in research and practice in higher education settings. *System, 77,* 1-9.

Dressen-Hammouda, D., & Wigham, C.R. (2022a). Challenges for ESP genre pedagogy in the digital age: Designing and evaluating critical pedagogical action. *In preparation.*

Dressen-Hammouda, D., & Wigham, C.R. (2022b). Aligning the teaching of multimodal literacies with students' professional needs. *In preparation*.

Fiske, J. (1987). *Television Culture*. London: Methuen.

Ganier, F. (2004). Factors affecting the processing of procedural instructions: Implications for document design. *IEEE Transactions on Professional Communication, 47*, 15-26.

Goodrich Andrade, H. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53,* 27-30.

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

Hafner, C.A. (2018). Genre innovation and multimodal expression in scholarly communication: Video methods articles in experimental biology. *Ibérica*, *36*, 15–41.

Halliday, M.K. (1978). *Language as Social Semiotic*. Edward Arnold.

Hewett, B.L., & Bourelle, T. (2020). *Professional Development in Online Teaching and Learning in Technical Communication: A Ten-Year Retrospective*. Routledge.

Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, *32*, 241–267.

Jewitt, C., Bezemer, J., & O'Halloran, K.L. (2016). *Introducing Multimodality*. Routledge.

de Koning, B. B., Hoogerheide, V., & Boucheix, J. M. (2018). Developments and trends in learning with instructional video. *Computers in Human Behavior*, *89,* 395–398.

Kress, G. (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge.

Kress, G., & Hodge, R. (1988). *Social Semiotics*. Polity Press.

Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2000). *Multimodal Teaching and Learning.* Continuum.

Kress, G. & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. Routledge.

Lazonder, A., & van der Meij, H. (1993). The minimal manual: Is less really more? *International Journal Man-Machine Studies, 39,* 729-752.

van Leeuwen, T. (2005). Multimodality, genre and design. In R.H. Jones & S. Norris (Eds.), *Discourse in Action: Introducing Mediated Discourse Analysis* (pp. 73–93). Routledge.

van Leeuwen, T. (2011). Rhythm and multimodal semiosis. In S. Dreyfus, S. Hood & M. Stenglin (Eds.), *Semiotic Margins: Meaning in Multimodalities* (pp. 168-176). Continuum.

Lemke, J.L. (1990). *Talking Science*. Ablex Publishing.

Lotherington, H., & Jenson, J. (2011). Teaching multimodal and digital literacy in L2 settings: New literacies, new basics, new pedagogies. *Annual Review of Applied Linguistics*, *31*, 226–246.

Lyle, J. (2002). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal, 29*, 861-878.

Mayer, R.E. (Ed.) (2014a). *Cambridge Handbook of Multimedia Learning,* 2nd edition. Cambridge University Press.

Mayer, R.E. (2014b). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R.L. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning,* 2nd edition (pp. 345-368). Cambridge University Press.

Mayer, R.E., & Moreno, R.E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92,* 117-125.

Mayer, R.E., Fennell, S., Farmer, L., Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*, 389-395.

van der Meij, H., & Carroll, J.M. (1995). Principles and heuristics for designing minimalist instruction. *Technical Communication, 42,* 243-261.

van der Meij, H., Karreman, J., & Steehouder, M. (2009). Three decades of research and professional practice on printed software tutorials for novices. *Technical Communication, 56,* 265-292.

van der Meij, H., & van der Meij, J. (2013). Eight guidelines for the design of instructional videos for software training. *Technical Communication, 60*, 205-228.

van der Meij, H., & van der Meij, J. (2014). A comparison of paper-based and video tutorials for software

learning. *Computers & Education*, *78*, 150–159.

van der Meij, H., & van der Meij, J. (2016). Demonstration-based training (DBT) in the design of a video tutorial for software training. *Instructional Science, 44,* 527-542.

Morain, M., & Swarts, J. (2012). YouTutorial: A framework for assessing instructional online video. *Technical Communication Quarterly, 21*, 6-24.

O'Halloran, K. (2008). Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery. *Visual Communication*, *7*, 443–475.

O'Halloran, K.L., & Lim, F.V. (2014). Systemic functional multimodal discourse analysis. In S. Morris & C. Maier (Eds.), *Texts, Images and Interactions: A Reader in Multimodality* (pp. 137–153). Mouton de Gruyter.

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R.L. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning,* 2nd edition (pp. 27-42). Cambridge University Press.

Paul, D., Charney, A., & Kendall, A. (2001). Moving beyond the moment: Reception studies in the rhetoric of science. *Journal of Business and Technical Communication*, *15,* 372-99.

Plaisant, C., & Shneiderman, B. (2005). Show me! Guidelines for recorded demonstration. Paper presented at the 2005 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/ HCC'05), Dallas, Texas. Retrieved from http:// www.cs.umd.edu/localphp/hcil/tech-reports-search.php?number=2005-02

Sauro, S., & Chappelle, C.A. (2017). Toward Langua-technocultural Competence. In C.A. Chappelle & S. Sauro (Eds), *The Handbook of Technology and Second Language Teaching and Learning* (pp.459-472). Wiley Blackwell.

Schriver, K. (2013). What do technical communicators need to know about information design? In J. Johnson-Eilola & S. Selber (Eds.), *Solving Problems in Technical Communication*. University of Chicago Press.

Steehouder, M.F., & van der Meij, H. (2005). Designing and evaluating procedural instructions with the four components model. In *IEEE International Professional Communication Conference Proceedings*, pp.797-801

Tan, S., O'Halloran, K.L., & Wignell, P. (2020). Multimodality. In A. De Fina, & A. Georgakopoulou (Eds.), *Handbook of Discourse Studies* (pp. 263-281). Cambridge University Press.

Tardy, C.M., & Matsuda, P. (2009). The construction of author voice by editorial board members. *Written Communication, 26,* 32-52.

Tversky, B., Bauer-Morrison, J., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57,* 247-262.

Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy, 3,* 211-223.

Ware, P. (2017). Technology, new literacies and language learners. In C.A. Chappelle & S. Sauro (Eds). *The Handbook of Technology and Second Language Teaching and Learning* (pp.265-277). Wiley Blackwell.

Appendix A. Stimulated Recall Interview Guide

Stage	Step	Questions
Introducing the research	Researcher introductions	Brief presentation (depending). Check technical features.
	Thank participant	Thanks for participating in our study and agreeing to a follow up interview online with me today. The purpose of this meeting is to get a little more background information about you and have a short exchange about your professional evaluation of the videos you watched.
	Permission to record	Would it be okay with you if I record our conversation from here on in using the screen capture software XXX? We'll make all information identifying you and the companies you've worked for anonymous so that in the final research study a reader would not be able to identify you.
	Any initial questions	Do you have any questions before we begin?
Professional background	Experience in field	Could you tell me how long you've worked in the field of technical communication and the different companies you've worked for and roles you've had?
	Experience development	In terms of how you've developed your experience, have you followed any formal training or certification programs? To what extent have you gained on-the-job experience?
	Language experience	What languages do you work in? Do you speak any other languages? Do you often work with non-native speakers of English? In what type of contexts do you work with them? What types of interactions does that mean you have with them?
	Experience with video tutorials	Coming back to the focus of our study, what is your experience in producing video tutorials? How often would you say you produce this type of document? Do you have any experience in evaluating or giving feedback on other people's videos?
	Other relevant information	Is there any other information regarding your professional background and tasks that you think might be relevant to us?
Discussion about videos and grid	Overall impressions about the videos	What was your overall feeling about how the videos met professional standards for tutorials of this type? Did one video in particular stand out? Was there anything specific that impressed you, either positively or negatively? Did you find that the videos had particular problems?
	Specific points from the evaluation	How did you find the evaluation grid? Do you think it covers the points students should think about? Is anything missing?

Finishing up	Thank participant. Turn-off screen recorder. Contribution to research	Thanks very much for agreeing to take part in our study – we really appreciate it as we feel it's so important to get professionals' opinions to help us better understand how university training programs are meeting workplace needs.
	Confidentiality	Just to reassure you about confidentiality, all names will be changed in the study as well as, for example, the names of businesses where you've worked.
	Participant questions	Do you have any questions about what we're planning to do with your participation in our study?
	Follow-up information	I don't think we will but if we are in need of any further information, would it be okay to contact you again via email / LinkedIn?

Appendix B. Supplemental data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2022.102727.