Received: 12 April 2022 | Revised: 3 February 2023 | Accepted: 26 February 2023

DOI: 10.1111/jcal.12803

ARTICLE

Journal of Computer Assisted Learning WILEY

Do social cues in instructional videos affect attention allocation, perceived cognitive load, and learning outcomes under different visual complexity conditions?

Julius Meier[1,2] | Bastian de Jong[1,3] | Dorien Preusterink van Montfort[1] | Anouk Verdonschot[1] | Margot van Wermeskerken[1] | Tamara van Gog[1]

[1]Department of Education, Utrecht University, The Netherlands

[2]Department of Psychology, University of Freiburg, Germany

[3]Institute of Child Development and Education, University of Amsterdam, The Netherlands

Correspondence
Julius Meier, Department of Psychology, University of Freiburg, Germany.
Email: julius.meier@psychologie.uni-freiburg.de

Abstract

Background: There are only few guidelines on how instructional videos should be designed to optimize learning. Recently, the effects of social cues on attention allocation and learning in instructional videos have been investigated. Due to inconsistent results, it has been suggested that the visual complexity of a video influences the effect of social cues on learning.

Objectives: Therefore, this study compared the effects of social cues (i.e., gaze & gesture) in low and high visual complexity videos on attention, perceived cognitive load, and learning outcomes.

Methods: Participants ($N = 71$) were allocated to a social cue or no social cue condition and watched both a low and a high visual complexity video. After each video, participants completed a knowledge test.

Results and Conclusions: Results showed that participants looked faster at referenced information and had higher learning outcomes in the low visual complexity condition. Social cues did not affect any of the dependent variables, except when including prior knowledge in the analysis: In this exploratory analysis, the inclusion of gaze and gesture cues in the videos did lead to better learning outcomes.

Takeaways: Our results show that the visual complexity of instructional videos and prior knowledge are important to take into account in future research on attention and learning from instructional videos.

KEYWORDS

attention, cognitive load, eye tracking, gaze cues, gesture cues, learning

Instructional videos have become increasingly popular in education. They are used to enhance traditional classroom courses, blended learning (i.e., the combination of online and traditional classroom learning), and fully online courses (Van Wermeskerken et al., 2018). A widely used form of instructional video is lecture-style videos, in which an instructor gives a verbal explanation about information that is displayed on a screen next to or behind the instructor (Fiorella & Mayer, 2016; Hoogerheide et al., 2016; Ouwehand et al., 2015). Recently, researchers have started to investigate the use of social cues such as gaze and gesture cues in instructional videos, that could guide learners' attention to the relevant content at the moment it is discussed by the instructor (e.g., Fiorella & Mayer, 2016b; Ouwehand et al., 2015; Stull et al., 2018; Töpper et al., 2014; Valenzeno et al., 2003). By guiding attention to the right information at the right time, these cues could have a beneficial effect on the processing of the relevant video content. More precisely, the integration of the verbal explanation with the visual content might be improved, which can be expected to result in better learning outcomes (Mayer, 2021; Van Gog, 2021).

Although several studies have shown that gaze and gesture cues direct attention to the relevant content of instructional videos (e.g., Singer & Goldin-Meadow, 2005; Theeuwes & Van der Stigchel, 2006), there is mixed evidence for their effect on learning outcomes (Cutica & Bucciarelli, 2008; Fiorella & Mayer, 2016b; Ouwehand et al., 2015; Stull et al., 2018; Valenzeno et al., 2003). In addition, it is unclear under which conditions these social cues foster learning; findings from research on visual cues (e.g., arrows, highlighting, colour coding) show, for instance, that the visual complexity (i.e., the number of objects, clutter, openness, symmetry, organization, and variety of colours; Olivia et al., 2004) of the video and the prior knowledge of the learner are important factors in the effectiveness of cueing on attention and learning (Richter et al., 2016; Van Gog, 2021). Therefore, the present study aimed to compare the effectiveness of social cues in high visual complexity versus low visual complexity videos on attention, cognitive load, and learning outcomes, and to explore the effect of prior knowledge.

1 | LEARNING FROM INSTRUCTIONAL VIDEOS

To learn from instructional videos, learners have to actively select, organize and integrate the visual information that is displayed on the screen with the verbal explanation provided by the instructor (see SOI model; Fiorella & Mayer, 2016a). An important first step in this learning process is that learners need to focus their attention on the visual information that the instructor is referring to in the verbal explanation timely (i.e., when it is being mentioned, Mayer, 2021). If much visual search is required before learners find the relevant visual information in the video, the verbal explanation will have progressed, and learners might not have enough time to connect the verbal explanation with the corresponding visual information. In other words, when the selection of relevant information is hampered, this information is not available for organization and integration, and learning will be impaired (Fiorella & Mayer, 2016a; Mayer, 2021; Van Gog, 2021).

Thus, to facilitate learning, instructional videos should be designed in such a way that first of all, the process of selection is optimized. One of the factors that may affect how quickly learners can locate visual information is the visual complexity of the learning material (Van Marlen et al., 2018; Van Marlen et al., 2019). Visual complexity can be determined by the perceptual dimensions of the number of objects, clutter, openness, symmetry, organization, and variety of colours. Visually complex displays or videos contain more objects and colours, have a more cluttered composition of objects, contain less open space, are less symmetrical, and are less organized compared to visually low complex displays (Olivia et al., 2004).

When the visual complexity of an instructional video is high, there is a lot of information for learners to process. This visual search process becomes even more demanding, when learners have to find the relevant information at the exact moment an instructor is referring to it (Van Marlen et al., 2019). Consequently, the visual search process imposes a high working memory load or cognitive load (see also Cognitive Load Theory, Sweller et al., 2011). However, not only visual complexity and temporal dependency influence the visual search and the resulting working memory load. In addition, learners' prior knowledge is important. Learners learn by building on prior knowledge and skills. As such, it is important, to design educational activities in such a way that they build on learners' prior knowledge (Schwartz et al., 2007). If learners already have prior knowledge on a topic (i.e., background knowledge about the content of the instructional video), it will be easier for them to select the verbally referenced information at the right time because they know where to look or they know for what kind of visual information to search for (Van Marlen et al., 2018). Accordingly, the overall working memory load is lower, which is beneficial for learning. Vice versa, with little prior knowledge, learners will find it harder to select the correct information at the right time, making them prone to lose track of the instructors' demonstration or explanation (Van Marlen et al., 2018, 2019).

Hence, it can be hypothesized that the selection (and subsequent organization and integration) of information can be optimized by decreasing the extent of visual search that is required to find the verbally referenced information, and that this would be especially helpful under conditions of high visual complexity and/or low prior knowledge (Van Marlen et al., 2018, 2019). One way to decrease visual search in instructional videos would be by having the instructor gaze and point at the verbally referenced information. Such deictic gestures are also referred to as social cues (Chandler & Sweller, 1991; Ouwehand et al., 2015; Theeuwes & Van der Stigchel, 2006). Social cues could guide the learners' attention to the right content at the right time because humans tend to automatically follow other people's eye movements and gestures to gain additional information about their intentions (e.g., Langton et al., 2000). Because of these cues, learners have to search less for the relevant content (Ouwehand et al., 2015; Theeuwes & Van der Stigchel, 2006), and their visual search becomes more efficient (Wolfe, 1994), and accurate (Davis et al., 2003; Wolfe et al., 2002).

2 | EFFECTS OF SOCIAL CUES ON ATTENTION AND LEARNING

Outcomes of several (eye-tracking) studies have supported the hypothesis that gazes and/or gestures in instructional videos help to (faster) direct the attention of the learner to the information that is verbally referred to (e.g., Fiorella & Mayer, 2016b; Ouwehand et al., 2015; Pi et al., 2019; Pi et al., 2021; Singer & Goldin-Meadow, 2005; Stull et al., 2018; Theeuwes & Van der Stigchel, 2006). However, even though it could be expected that cues that help learners attend to the right information at the right time would foster learning outcomes, the findings regarding the effects of these cues on learning are inconclusive.

Fiorella and Mayer (2016a) did not find that using gesture cues in an instructional video was more beneficial for learning outcomes among undergraduate students in comparison to no gesture cues. Moreover, Stull et al. (2018) did not find better learning outcomes among undergraduate students when gaze cues were used in a lecture-style video than when no gaze cues were used in the video. Ouwehand et al. (2015) assessed the effect of both gaze and gesture cues on learning outcomes in three experimental conditions: a gaze and gesture condition, a gaze only condition, and a no social cue condition (i.e., the model looked straight into the camera). No effects of gaze and gesture cues on learning outcomes or cognitive load were found (Ouwehand et al., 2015).

On the other hand, Wang et al. (2019) did find that undergraduate students who watched a video with gaze cues performed better on the posttest than students in a no gaze condition. Likewise, Pi et al. (2019) found that undergraduate and graduate students performed better on the posttest when they watched a video using both gaze and gesture cues or only gesture cues than students that were in the no social cue condition. Moreover, Stull et al. (2021) found that students scored higher on the posttest when they watched a video with an instructor that gazed between the whiteboard and camera in comparison to students that watched a video with an instructor that only gazed at the whiteboard. Also, Valenzeno et al. (2003) found that preschool children who watched a video containing gaze and gesture cues performed better on the posttest than children who did not.

2.1 | The role of visual complexity

These mixed effects of social cues on learning outcomes could be explained by differences in the videos' visual complexity across studies. The more visually complex a video is, the more difficult it may become to locate the right information at the right moment because more visual search is necessary (Van Marlen et al., 2019). Moreover, the visual complexity of videos does not only affect cognitive load during video study (Van Marlen et al., 2018, 2019) but may also interact with the learners' prior knowledge of the instructed topic. As mentioned earlier, learners with little prior knowledge may not know where in the video they can find the relevant information and therefore will take longer to do so (if they can do so at all), and this problem will be aggravated when a video is more visually complex (Van Marlen et al., 2019).

2.2 | Effects of prior knowledge

Prior knowledge also affects the available working memory resources in general (Sweller et al., 2011). Moreover, a meta-analysis on visual cues (e.g., arrows, highlighting, colour coding) showed that these cues improved learning for students with low prior knowledge, but effects on learning outcomes were smaller or absent for learners with higher prior knowledge (Richter et al., 2016). Moreover, Fiorella and Mayer (2016) found that gesture cues were more effective for learning outcomes for low prior knowledge students than for high prior knowledge students. Social cues might be redundant for high prior knowledge learners as they already have existing knowledge about the content of the videos (Kalyuga, 2009) which should allow them to locate the referenced information easier and faster in comparison to low prior knowledge learners who would be less able to find referenced information rapidly (Johnson et al., 2015). However, how the effects of gaze and gesture cues on attention, cognitive load, and learning are affected by different visual complexity conditions is an open question (Van Wermeskerken et al., 2018). So far, only Pi et al. (2022) showed in their study that the visual complexity of instructional videos affects the effectiveness of head nods and beat gestures on learning outcomes. Consequently, we also expect visual complexity to mediate the effects of gaze and gesture cues on attention, cognitive load, and learning.

2.3 | The present study

In short, little is known about the effect of gaze and gesture cues on attention, cognitive load, and learning outcomes under different visual complexity conditions (Van Wermeskerken et al., 2018). Furthermore, only Fiorella and Mayer (2016a) investigated the effects of prior knowledge on gesture cues but not on both gaze and gesture cues. Therefore, the present study compares the effectiveness of social cues versus no social cues (between-subjects factor) in instructional videos with high versus low visual complexity (within-subjects factor) on attention, cognitive load,[1] and learning outcomes, and explores the effects of prior knowledge. Learners' attention allocation will mainly be investigated with eye tracking.

For this study, we formulated the following hypotheses regarding the effects of social cues and visual complexity on learners' quality and speed of attention allocation, perceived cognitive load, and learning outcomes: When watching low visual complexity videos, learners will find more relevant content (H1a; quality of attention allocation), will find relevant content in the video more quickly (H2a; speed of attention allocation), will perceive a lower cognitive load (H3a), and have higher learning outcomes (H4a) than when watching high visual complexity videos. Moreover, learners watching videos with social cues will find more relevant content (H1b), will find relevant content in the video more quickly (H2b), will perceive a lower cognitive load (H3b) and have higher learning outcomes (H4b) than learners

[1]Since it is sometimes debated in research to what extent learners are able to indicate how much mental effort they actually had to invest (Seeber, 2013), we also attempted to investigate changes in pupil size as a physiological measure of cognitive load (Van Gerven et al., 2004). When the mental effort is higher, the diameter of the pupil usually increases (Pomplun & Sunkara, 2003; Van Gerven et al., 2004). However, our data did not show any meaningfully interpretable effects of the experimental conditions on pupil size, probably due to the dynamic nature of the stimuli, which have a strong influence on pupil size due to changing light conditions (i.e., we did find effects on pupil size during the calibration procedure, where participants had to do mental calculations). Therefore, we did not include those data in the manuscript. Further information on the calibration procedure, the data analysis and results can be requested from the corresponding authors.

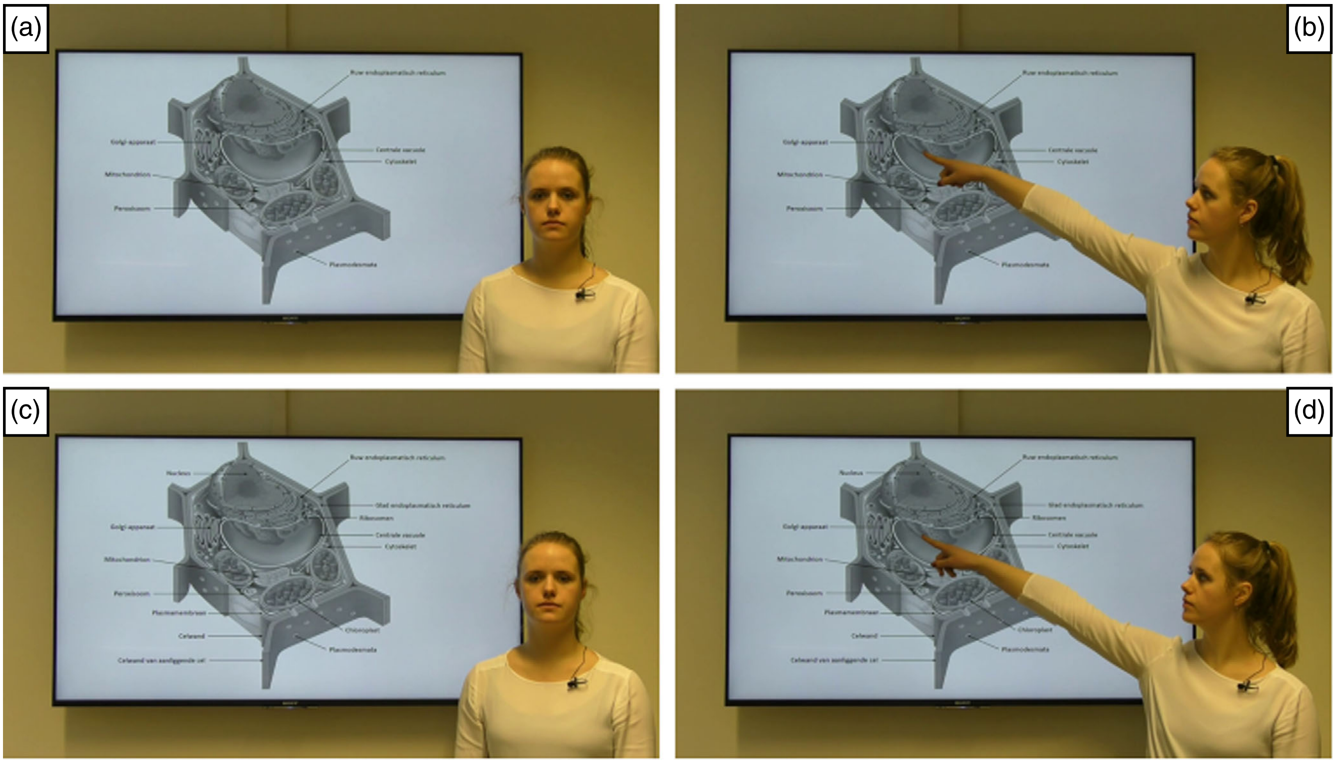

FIGURE 1 Screenshots of the four videos explaining the structure of a plant cell. Note: The top left screenshot (a) shows a low visual complexity video (LC) and no social cues (NSC); the top right screenshot (b) shows a low visual complexity video (LC) with social cues (SC); the bottom left screenshot (c) shows a high visual complexity video (HC) and no social cues (NSC); the bottom right screenshot (d) shows a high complexity video (HC) with social cues (SC).

watching videos without social cues. Finally, we expect interaction effects between social cues and visual complexity, indicating that social cues are more beneficial for high visual complexity videos on quality (H1c) and speed (H2c) of attention allocation, cognitive load (H3c), and learning outcomes (H4c) than for videos of low visual complexity. Furthermore, we explore the effects of prior knowledge, as one would expect that social cues have larger beneficial effects on attention, cognitive load and learning outcomes for low prior knowledge students than high prior knowledge students since the latter group needs to use less working memory capacity to comprehend the content (Sweller et al., 2011).

3 | METHOD

3.1 | Participants and design

Participants were 71 Dutch adults (26 male; $M_{age} = 22.9$ years, $SD = 4.2$; age range 18–50 years) who were either enrolled in higher education or had completed a higher education degree (i.e., Bachelor or Master). Only participants without epilepsy and with normal or corrected-to-normal vision were recruited. 36 of the participants indicated they took final exams in biology in secondary education (meaning they had 5–6 years of biology classes) the remaining 35 participants did not (meaning they had max. 3 years of biology classes). Participants received a small gift or course credit for their participation.

A 2x2 mixed factors design was used, with social cues (yes/no) as between-subjects factor and visual complexity (low/high) as within-subjects factor. Participants were randomly assigned to either the social cues (SC; $n = 35$) or no social cues (NSC; $n = 36$) condition. All participants studied a high visual complexity (HC) and a low visual complexity (LC) video (with gaze and gesture cues in the SC condition, without those cues in the NSC condition). While studying, participants' eye movements were recorded. The instructional videos were from the domain of biology and focused on the structure of plant cells and leaves. The topics and the order in which the visual complexity videos were presented were counterbalanced (i.e., some participants studied a LC video on plant cell structure, others on leaf structure and some participants watched the LC video first, others the HC video first) to prevent a confound between topic or order and visual complexity of the video.

4 | MATERIALS AND INSTRUMENTS

4.1 | Prior knowledge test

As a measure of general prior knowledge, participants indicated if they took final exams in biology at the end of secondary education

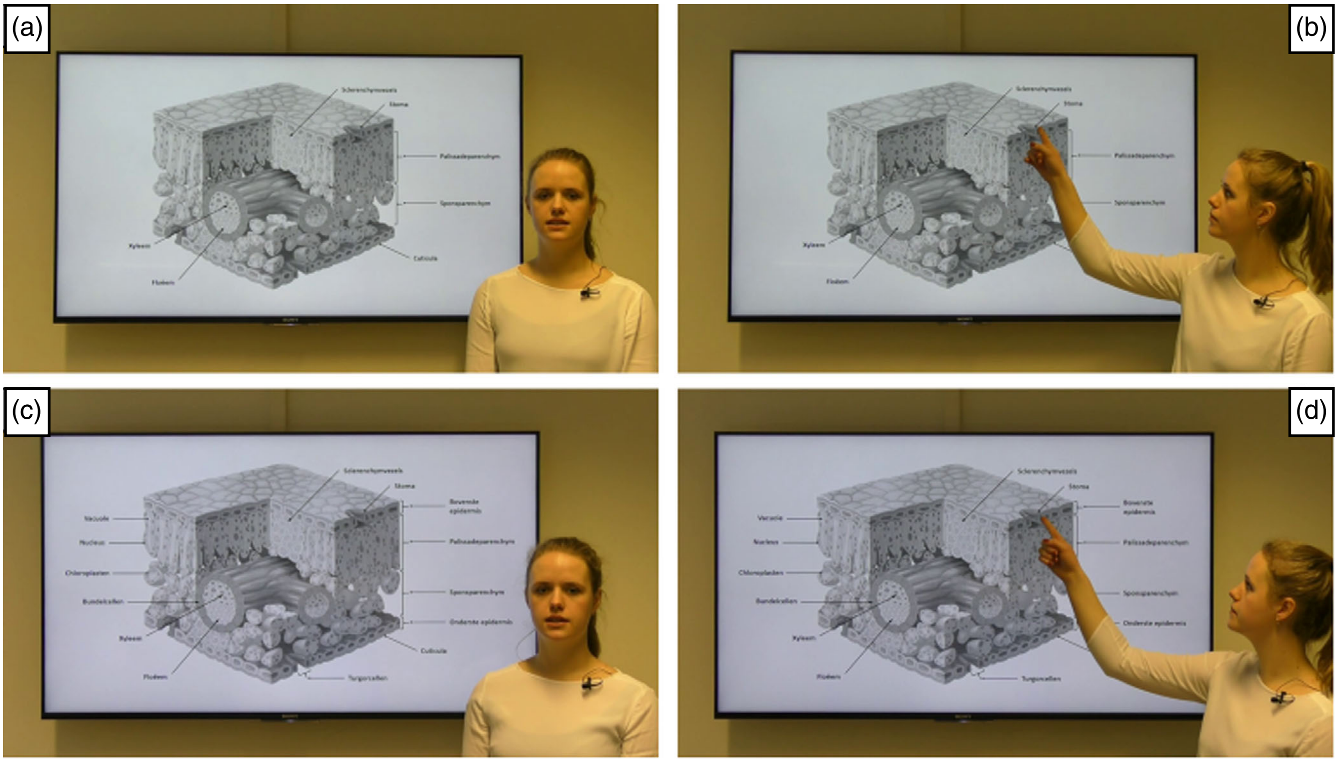

FIGURE 2 Screenshots of the four videos explaining the structure of a leaf. Note: The top left screenshot (a) shows a low visual complexity video (LC) and no social cues (NSC); the top right screenshot (b) shows a low visual complexity video (LC) with social cues (SC); the bottom left screenshot (c) shows a high visual complexity video (HC) and no social cues (NSC); the bottom right screenshot (d) shows a high complexity video (HC) with social cues (SC).

(yes/no). In addition, two open (free recall) questions were used to test participants' specific prior knowledge about plant cells and leaf layers (e.g., write down all the parts of a plant cell that you can think of).[2]

4.2 | Instructional videos

Eight videos (resolution 1280x720px) showing a female instructor next to a screen were recorded. In the videos, the instructor either explained the structure of a plant cell (four videos; Figure 1) or a leaf (four videos; Figure 2). Visual complexity was defined by the number of labelled parts. In the LC condition, only the seven parts of the plant cell or leaf that the verbal instructions focused on were labelled. In the HC condition, seven additional parts were labelled. The presence of social (i.e., gaze and gesture) cues was manipulated by having the instructor either look straight into the camera while delivering the lecture on the cell or the leaf (i.e., no social cues; NSC) or having the instructor gaze and gesture towards the seven parts she was referring to (i.e., social cues; SC). Hence, in the SC conditions, the instructor gave seven social cues by looking and gazing at the referred information, which were not present in the NSC condition. The same instructor recorded all videos. The lectures were scripted and consisted of the same number of words for both topics. The total duration for all videos was 79 s.

4.3 | Eye-tracking equipment

To investigate participants' speed and quality of attention allocation, their eye movements were recorded while watching the videos. A SMI RED250 (SensoMotoric Instruments GmbH, Teltow, Germany) eye-tracker (250 Hz) was used. SMI Experiment Center (Version 3.7.68) was used to present the videos. The instructional videos were presented on a 22-inch monitor with a resolution of 1680 x 1050px. The distance between the participants' eyes and the monitor's center was approximately 60 cm. A forehead-and-chinrest was used to reduce head movements. The eye-tracking system was calibrated with a thirteen-point animated cross procedure. This procedure was repeated up to four times until a calibration accuracy below 0.6° was achieved in both the x and y direction. If that was not the case after four calibrations, the best

[2]Originally, we planned to use the score in the open prior knowledge tests as a covariate in the analyses. Participants' answers were scored by the authors two and four with the scoring scheme presented in Table A1. However, the answers to the open prior knowledge questions suggested that none of the participants had any specific prior knowledge of leaf layers (all participants scored 0; interrater agreement was 100%) and that specific prior knowledge about plant cells was extremely low (interrater agreement was 89%). Consequently, this prior knowledge score was not suitable as covariate. Instead, in an additional exploratory analyses, we used the general prior knowledge measure (i.e., whether participants took final exams in biology in school) as additional factor.

calibration was used. The eye-movement data were analysed using SMI BeGaze software (version 3.7.59).

4.4 | Perceived cognitive load

Perceived cognitive load was measured with the 9-point mental effort rating scale developed by Paas (1992; also used in e.g., Hoogerheide et al., 2016; Ouwehand et al., 2015) which asked participants 'How much effort did you invest while studying the video?' with answer options ranging from (1) very, very low mental effort to (9) very, very high mental effort.

4.5 | Learning outcomes

To assess participants' learning outcomes after watching one of the videos, one posttest was designed for the video on the plant cell and one for the video on the leaf structure. In both posttests, participants were provided with the pictures from the HC videos. The fourteen labels of the HC pictures had been replaced by letters. For each of the seven parts (of a cell or a leaf) that were mentioned in the videos, participants were then asked to name the letter representing a particular part (e.g., Which letter represents the xylem?). Thus, this knowledge test was a retention test. For each correct letter, participants received one point. Hence, in both posttests, a maximum of seven points could be achieved.[3]

4.6 | Procedure

Each participant was tested individually in a laboratory. After giving informed consent, participants were seated in front of the computer monitor, with their heads positioned in a forehead-and-chinrest, after which headphones were put on them. Table 1 provides an overview of the experimental procedure.

First, participants provided demographical information and answered the prior knowledge questions. Then, they were informed about the experimental procedure and asked not to move while being in the forehead-and-chinrest. Subsequently, the eye tracker was calibrated, after which the first instructional video was presented. When the video finished, participants could take their heads out of the forehead-and-chinrest and they were instructed to complete the mental effort rating and the posttest on the topic of the first video. Before moving on to the second video, participants placed their heads in the forehead-and-chinrest again and the calibration of the eye tracker was checked using a 5-point validation procedure. If necessary, a recalibration took place. Then, the second video was presented, and participants again rated invested mental effort and completed the posttest. The entire procedure lasted approximately 30 min. Except for the videos presented with SMI Experiment Center (version 3.7.68) software, all other steps of the experiment (e.g., demographics and posttest) were performed with the online survey tool Qualtrics.

TABLE 1 Overview of the experimental procedure

Step	Performed procedure
Demographics and prior knowledge	Participants provided demographical information and answered the prior knowledge questions.
Calibration	The eye tracker was calibrated.
First video	Participants watched the first video.
Mental effort and posttest	Participants rated their mental effort and completed a posttest regarding the first video.
Validation	The eye tracker calibration was validated.
Second video	Participants watched the second video.
Mental effort and posttest	Participants rated their mental effort and completed a posttest regarding the second video.

4.7 | Data analysis

Because of an experimenter error during the experiment, one participant (from the SC condition, who did take final exams in biology) had to be excluded from all data analyses, leaving us with $n = 70$ participants (NSC: $n = 36$; SC: $n = 34$). For the exploratory analysis with the additional between-subjects factor *prior knowledge* (PK; i.e., whether participants had taken final exams in biology: *yes* or *no*), the subsample sizes of high and low PK groups were also well balanced (i.e., SC condition: $n = 17$ in both the high and low PK group; NSC condition: $n = 18$ in both the high and low PK group).

Data were analysed with mixed 2x2 ANOVAs with social cue condition (SC vs. NSC) as between-subjects factor and visual complexity (HC vs. LC) as within-subjects factor. Furthermore, we also performed exploratory 2x2x2 repeated measures ANOVAs with between-subjects factors PK (high/low) and social cues (yes/no) and within-subjects factor visual complexity (low/high). A significance level of $\alpha = 0.05$ was used. Partial eta squared is reported as a measure of effect size, with $\eta_p{}^2 = 0.01$, $\eta_p{}^2 = 0.06$, and $\eta_p{}^2 = 0.14$ indicating small, medium, and large effects, respectively (Cohen, 1988).

Before each analysis, for each of the dependent variables the assumptions of absence of outliers, normal distribution, and homogeneous variances were checked. For most analyses, the assumptions were met, except for a violation of normality in the analysis of

[3]Note that the posttest also included questions that asked participants to name the part (of the cell or leaf) represented by a letter (7 items; e.g., Which part does the letter h represent?). However, we decided to not analyse the answers on these questions. A preliminary screening of the answers revealed great variety in spelling and completeness of answers (e.g., a participant provided the answer *palissade* instead of *palissadeparenchym*), making it impossible to reliably score these data.
We also asked participants to describe the function of components as mentioned in the video lecture (e.g., What is the function of the component represented by letter h?). These questions aimed to prevent participants from rote learning the labels (i.e., as the posttest was completed immediately after each video, participants would know what to expect after the first posttest which could influence their study behaviour on the second video). However, these answers were not scored and analysed.

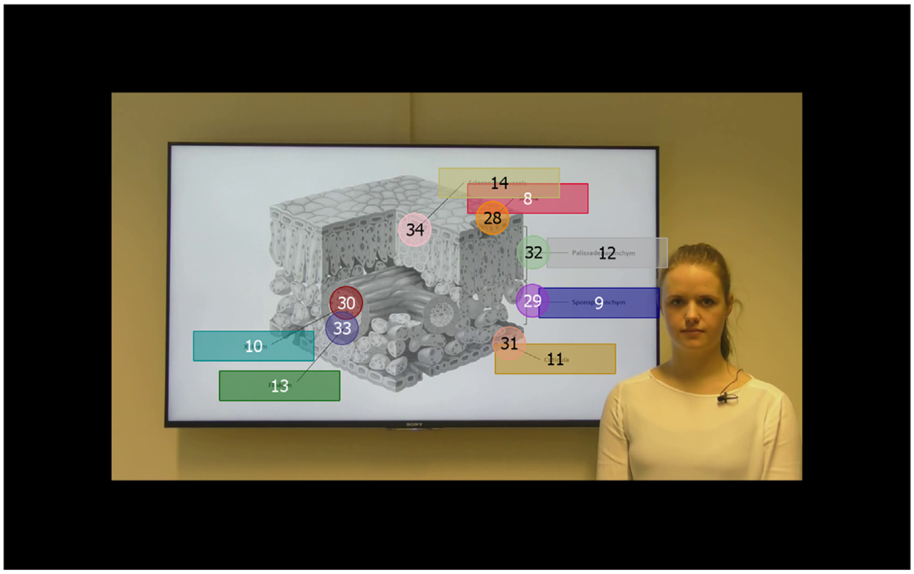

FIGURE 3 Screenshot of a low complexity video of a leaf layer without social cues with highlighted AOIs.

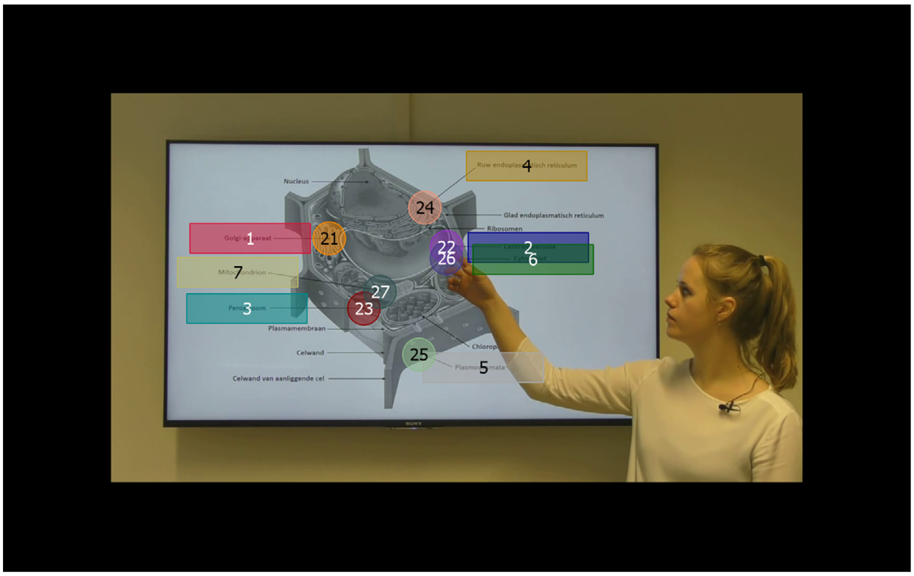

FIGURE 4 Screenshot of a high complexity video of a plant cell with social cues with highlighted AOIs.

perceived cognitive load. The different dependent variables and their definition will be described in the following two sections.

4.8 | Attention

To investigate participants' quality and speed of attention allocation (i.e., whether and how fast they looked at the part of the plant cell or leaf picture that was verbally referenced by the instructor) two Areas of Interest (AOIs) were created for each referenced label (see Figures 3 and 4). Each one AOI covered the label (seven rectangular AOIs, 223px x 55px) and one covered the head of the arrow (seven circular AOIs, diameter of 60px).

Subsequently, we determined the onset (i.e., the moment at which the instructor started vocalizing) of each verbal referent (i.e., words referring to parts of the plant cell or leaf pictured in the video; e.g., 'palissadeparenchym'). For each video and each referent, we then determined whether the participants fixated on one of the two corresponding AOIs (i.e., label or head of the arrow) between 100 ms (i.e., the time needed for processing of the verbal information; cf. Altmann, 2011) and 3500 ms after the onset of the referent (cf. Van Wermeskerken et al., 2018[4]). Fixations were defined as lasting >50 ms and velocity ≤ 40°/s.

To investigate participants' quality of attention allocation (cf. H1a, H1b, and H1c), we summed up the occasions a participant did fixate

[4]Note that Van Wermeskerken et al. (2018) used a smaller time window of 2000 ms after the onset of the referent. However, we decided to use a longer time window as there were more distracting items and we used smaller AoIs in the present paper in comparison to Van Wermeskerken et al. (2018).

at least one of the two AOIs per referent (i.e., *number of fixated AOIs*), resulting in a number between 0 and 7 for each complexity condition.

As a first measure of speed of attention allocation (cf. H2a, H2b, and H2c), we assessed the time period between the onset of a referent and the moment in which a participant first fixated one of the two corresponding AOIs. This resulted in 0 to 7 values from which we then calculated the mean, resulting in one *average time to first fixation* value per complexity condition.

However, the average time to first fixation may be biased by the number of observers who actually fixate one of the AOIs (e.g., only 65% of the participants fixated the first referent in the LC condition). As a result, the average time to first fixation could be inaccurate and underestimated, since only those participants who locate the corresponding label are included in the calculation. Thus, as a second measure of speed of attention allocation, for each of the referents, the time by which 50% of the participants in each condition had fixated on one of the two AOIs was determined (i.e., *T50*, Hooge & Camps, 2013). In contrast to the average time to first fixation, this T50 measure also takes into account participants who do not find a label at all.

Participants with poor calibration accuracy (i.e., higher than 0.9) were excluded from all eye movement data analyses ($n = 18$; remaining $n = 52$; NSC: $n = 26$, of whom 13 had high PK and 13 low PK; SC: $n = 26$, of whom 13 had high PK and 13 low PK). The mean calibration accuracy of the remaining participants was $M_X = 0.53$ ($SD_X = 0.17$), $M_Y = 0.45$ ($SD_Y = 0.17$) for the first video and $M_X = 0.45$ ($SD_X = 0.15$) $M_Y = 0.43$ ($SD_Y = 0.17$) for the second video. The average tracking ratio was 97.51% ($SD = 1.73$%). Moreover, if participants did not fixate any of the relevant AOIs for one of the videos, also no average time to first fixation was available for this video or complexity condition. This was the case three times, leaving us with $n = 49$ participants (NSC: $n = 25$, of whom 13 had high PK and 12 low PK; SC: $n = 24$, of whom 11 had high PK and 13 low PK) for these analyses of average time to first fixation.

4.9 | Perceived cognitive load and learning outcomes

To investigate participants' rating of *perceived cognitive load* the 9-point mental effort rating scale was analysed. Because the assumption of normality was violated for perceived cognitive load, a non-parametric form of the mixed ANOVA was performed using the Rpackage nparLD (Noguchi et al., 2012). To analyse learning outcomes, we summed the correct scores on the questions asking the participants to name the letter that corresponded to a certain part. One point was assigned for each correct answer (i.e., the *posttest score* ranged from 0–7). For these analyses, data of all 70 participants were available.

5 | RESULTS

Descriptive statistics of attentional measures (number of fixated AOIs, average time to first fixation, T50), perceived cognitive load, and learning outcomes are displayed in Table 2. Descriptive statistics for the explorative analysis with PK as additional between-subjects factor are displayed in Table 3. The results of all analyses are displayed in Tables 4 and 5; below, significant results are described.

5.1 | Attention

The results of the mixed ANOVAs in Table 4 show significant main effects of the within-subjects factor visual complexity on average time to first fixation and T50 (both H2a). For both effects, the means in Table 2 indicate that participants found relevant labels faster when watching a low complexity video than when watching a high complexity video. Moreover, there was a significant interaction effect of visual complexity and social cues on the number of fixated AOIs (H1c; see Figure 5). When no social cues were present, participants found on average one label fewer when watching a high complexity video than when watching a low complexity video. When social cues were present, however, participants fixated the same number of labels, irrespective of the complexity of the video they were watching. Adding PK as additional between-subjects factor in the exploratory analysis did not affect these outcomes (see Tables 3 and 5).

5.2 | Perceived cognitive load

The non-parametric ANOVA did not reveal any significant main or interaction effects (see Table 4; H3a to H3c), implying that perceived

TABLE 2 Means and standard deviations of attentional measures, perceived cognitive load and posttest score displayed for low and high complexity videos without (NSC) and with social cues (SC)

	Low complexity		High complexity	
Variable	NSC	SC	NSC	SC
Number of fixated AOIs (of 7)	4.92 (1.90)	4.88 (1.82)	3.65 (1.62)	5.12 (1.73)
Average time to first fixation (ms)	1251.21 (311.46)	1352.27 (254.86)	1882.34(588.42)	1866.91 (346.92)
T50 (ms)	1132.14 (320.48)	1340.67 (318.84)	1951.49 (563.50)	1851.75 (532.41)
Perceived cognitive load (1–9)	5.53 (1.30)	5.53 (1.33)	5.53 (1.21)	5.50 (1.44)
Posttest score (0–7)	4.17 (1.98)	4.94 (1.98)	3.44 (2.10)	4.15 (1.81)

TABLE 3 Means and standard deviations for the attentional measures, perceived cognitive load and posttest score for the exploratory analyses with prior knowledge (PK) as additional between-subjects factor

	Low complexity				High complexity			
	NSC		SC		NSC		SC	
Variable	Low PK	High PK	Low PK	High PK	Low PK	High PK	Low PK	High PK
Number of fixated AOIs (of 7)	4.92 (2.25)	4.92 (1.55)	5.38 (1.12)	4.38 (2.26)	3.69 (2.02)	3.62 (1.19)	5.54 (1.33)	4.69 (2.02)
Average time to first fixation (ms)	1377.41 (300.37)	1134.71 (284.03)	1399.24 (282.29)	1296.76 (217.96)	1939.33 (708.32)	1829.73 (475.73)	1870.48 (358.62)	1862.69 (349.92)
T50 (ms)	1220.74 (309.16)	1043.53 (329.60)	1346.22 (347.99)	1335.12 (314.78)	2056.80 (526.46)	1846.18 (620.45)	1871.93 (509.56)	1831.57 (594.62)
Perceived cognitive load (1–9)	5.72 (1.32)	5.33 (1.28)	5.47 (1.50)	5.59 (1.18)	5.72 (1.23)	5.33 (1.19)	5.18 (1.74)	5.82 (1.01)
Posttest score (0–7)	3.28 (1.87)	5.06 (1.70)	4.12 (2.03)	5.76 (1.60)	2.22 (1.56)	4.67 (1.88)	3.24 (1.44)	5.06 (1.71)

cognitive load did not differ across conditions. Adding PK as additional between-subjects factor resulted in a marginally significant three-way interaction effect of visual complexity, social cues and prior knowledge on perceived cognitive load (see Tables 3 and 5). Figures 6 and 7 illustrate this effect: for participants with low prior knowledge, social cues reduced the perceived cognitive load. This effect was stronger for high complexity videos than for low complexity videos. For participants with high prior knowledge, however, social cues increased perceived cognitive load. Again, the effect was stronger for high complexity videos than for low complexity videos.

5.3 | Learning outcomes

The results of the mixed ANOVAs in Table 4 showed a significant main effect of visual complexity on learning outcomes (H4a). Participants performed better on the posttest after watching a low complexity video than after watching a high complexity video (see Table 2 for the reported means).

After adding PK as additional between-subjects factor in the exploratory analysis, however, there were significant main effects of the within-subjects factor visual complexity and the between-subjects factors social cues and prior knowledge (but no significant interaction effects; see Table 5). Participants performed better on the posttest after watching a low complexity video than after watching a high complexity video. Moreover, participants in the SC condition also performed better than participants in the NSC condition. Finally, participants with high PK (i.e., participants who took a final exam in biology) performed better on the posttest than participants with low PK (see Table 3 for the means).

6 | DISCUSSION

This study aimed to investigate the effect of gaze and gesture cues in high versus low visual complexity videos on attention, cognitive load, and learning outcomes and to explore the effects of prior knowledge. We expected a lower visual complexity to increase quality (H1a) and speed (H2a) of attention allocation, decrease cognitive load (H3a), and promote learning outcomes (H4a). Similarly, we expected social cues to increase quality (H1b) and speed (H2b) of attention allocation, decrease cognitive load (H3b), and promote learning outcomes (H4b). Eventually, we expected social cues to be more beneficial for high visual complexity videos for quality (H1c) and speed (H2c) of attention allocation, cognitive load (H3c), and learning outcomes (H4c) than for videos of low visual complexity. We found support for hypotheses H2a, H4a, and H1c, whereas all other hypotheses were rejected. These findings are discussed separately below for the three key dependent variables attention, cognitive load and learning outcomes.

6.1 | Attention

In line with H1c, participants found more relevant labels when studying a video with social cues in the high complexity condition. This finding suggests that in the low complexity conditions participants did not have to rely on social cues to find more relevant labels but that the cues improved their visual search under the high complexity condition. Also as expected, participants allocated their attention more quickly to an area that was verbally referred to by the instructor when they watched a video with low visual complexity content than when they watched a video with high visual complexity content (H2a). This is in line with previous studies that reported similar effects (Davis et al., 2003; Mayer & Moreno, 2003; Wolfe, 1994; Wolfe et al., 2002). Wolfe et al. (2002) and Mayer and Moreno (2003) argue that increased visual complexity resulted in less efficient visual search, hampering attending to the right information at the right time. We assume that this was also the case in our study.

Contrary to our expectations, the presence of social cues did not decrease the time participants needed to find and fixate on the relevant content either overall (H2b) or as a function of visual complexity (H2c). This is not in line with previous studies in which the use of gaze and gesture cues guided the learner towards the relevant content. (e.g., Fiorella & Mayer, 2016b; Ouwehand et al., 2015; Pi et al., 2019; Singer & Goldin-Meadow, 2005; Stull et al., 2018; Theeuwes & Van

TABLE 4 Two-way mixed ANOVAs for the effects of complexity and cue condition on the attentional measures, perceived cognitive load and posttest score

Quality of attention allocation: Number of fixated AOIs				
	Hypothesis	F(1, 50)	p	η_p^2
Complexity	H1a	3.25	0.08	0.06
Social cues	H1b	3.21	0.08	0.06
Social cues x Complexity	H1c	6.78	0.01*	0.12
Speed of attention allocation: Average time to first fixation				
	Hypothesis	F(1, 47)	p	η_p^2
Complexity	H2a	50.75	<0.001*	0.52
Social cues	H2b	0.28	0.60	0.01
Social cues x Complexity	H2c	0.53	0.47	0.01
Speed of attention allocation: T50				
	Hypothesis	F(1, 26)	p	η_p^2
Complexity	H2a	44.87	<0.001*	0.63
Social cues	H2b	0.16	0.70	0.01
Social cues x Complexity	H2c	2.41	0.13	0.09
Perceived cognitive load				
	Hypothesis	ATS(1)	p	η_p^2
Complexity	H3a	0.10	0.76	n/a
Social cues	H3b	0.26	0.61	n/a
Social cues x Complexity	H3c	0.00	0.98	n/a
Posttest score				
Source	Hypothesis	F(1, 68)	p	η_p^2
Complexity	H4a	7.37	0.01*	0.10
Social cues	H4b	3.76	0.06	0.05
Social cues x Complexity	H4c	0.02	0.90	0.00

Note: ATS = ANOVA Type Statistic of the non-parametric analysis.
*α < 0.05.

der Stigchel, 2006). As a result, the learner looked less often at the instructor explaining the content. Our finding also seems to contradict the support for H1c. Possibly, participants in the social cue condition waited for the instructor to gesture towards the relevant content and only then attended to it, as the instructor executed the social cues only after the respective verbal referent had been fully verbalized. Participants in the no social cue condition, in contrast, may have already started searching for the respective content when the instructor just started vocalizing the word and therefore may have found it as fast as participants who did see gaze and gesture cues.

6.2 | Perceived cognitive load

Against our expectations, there were no significant effects of visual complexity or social cues on perceived cognitive load (H3a, H3b, H3c). These outcomes are not in line with studies on other types of visual cueing that did report that cueing reduced perceived cognitive load (e.g., Kalyuga et al., 1999). Ouwehand et al. (2015), on the other hand, also did not find significant effects of gaze and gesture cues on perceived cognitive load while using the 9-point subjective scale by Paas (1992) that we also used in our study. It is possible that this measure was not sensitive enough to measure the effects of visual complexity and social cues on experienced cognitive load. The scale by Paas (1992) measures experienced cognitive load as an overall measure, but does not distinguish the three different types of cognitive load (intrinsic, germane and extraneous; Sweller, 2010). Regarding visual complexity, one would expect effects particularly on intrinsic cognitive load (the number of interacting information elements in a task) and extraneous cognitive load (increased visual search), whereas for social cues, one would expect effects particularly on extraneous cognitive load (counteracting negative effects of increased visual search and thereby reducing extraneous load; De Koning et al., 2009). It could be the case that we would have found effects on extraneous cognitive load if we used an instrument tailored to measure extraneous load (e.g., Klepsch et al., 2017; Klepsch & Seufert, 2020; Leppink et al., 2013). It would be interesting to investigate this further in future research on social cues.

When we added prior knowledge as additional between-subjects factor in the exploratory analysis, we found a marginally significant three-way interaction of visual complexity, social cues and prior knowledge on perceived cognitive load. This effect could be explained, at least in part, by the expertise reversal effect (Kalyuga & Renkl, 2010). This effect describes that instructional aids in learning materials, such as social cues, can be helpful for less experienced learners, but at the same time impair learning success for learners with more prior knowledge. We found that for learners with little prior knowledge, social cues reduced cognitive load. For these learners, social cues may have served as valuable guidance that supported schema construction and thereby reduced the load on working memory. This effect was even more pronounced for the complex videos, where the initial load on working memory was even greater. For learners with high prior knowledge, on the other hand, the videos in the different complexity conditions (high vs low) were equally demanding in terms of cognitive load. When learning with social cues, however, cognitive load increased. Possibly, high prior knowledge learners do not need the support of the cues because they can find the relevant items on their own due to their more sophisticated schemas in the domain of biology. The finding that this effect was again larger for the complex videos could mean that the complex videos produced such a cognitive load level that the learners could just barely cope with them. The additional, actually unnecessary social cues, therefore, led to a stronger increase in cognitive load for the complex videos than for the less complex videos, for which there was still a certain buffer in the learners' working memory. However, we could not find the same three-way interaction for either attentional measures or learning outcomes. To investigate this effect in more

TABLE 5 Exploratory three-way mixed ANOVA for the effects of complexity, cue condition and prior knowledge on the attentional measures, perceived cognitive load and posttest score

Quality of attention allocation: Number of fixated AOIs			
	$F(1, 48)$	p	η_p^2
Complexity	3.12	0.08	0.06
Social cues	3.26	0.08	0.06
PK	1.49	0.23	0.03
Social cues x Complexity	6.51	0.01*	0.12
Complexity x PK	0.00	0.95	0.00
Social cues x PK	1.26	0.27	0.03
Complexity x social cues x PK	0.04	0.85	0.00
Speed of attention allocation: Average time to first fixation			
	$F(1, 45)$	p	η_p^2
Complexity	49.03	<0.001*	0.52
Social cues	0.21	0.65	0.01
PK	2.09	0.16	0.04
Social cues x Complexity	0.45	0.51	0.01
Complexity x PK	0.48	0.49	0.01
Social cues x PK	0.57	0.45	0.01
Complexity x social cues x PK	0.014	0.91	0.00
Speed of attention allocation: T50			
	$F(1, 24)$	p	η_p^2
Complexity	41.46	<0.001*	0.63
Social cues	0.15	0.70	0.01
PK	0.61	0.44	0.03
Social cues x Complexity	2.23	0.15	0.09
Complexity x PK	0.23	0.88	<0.01
Social cues x PK	0.36	0.55	0.02
Complexity x social cues x PK	0.00	0.99	<0.01
Perceived cognitive load			
	$ATS(1)$	p	η_p^2
Complexity	0.07	0.79	n/a
Social cues	0.24	0.62	n/a
PK	0.00	0.96	n/a
Social cues x Complexity	0.00	0.95	n/a
Complexity x PK	0.26	0.61	n/a
Social cues x PK	0.04	0.85	n/a
Complexity x social cues x PK	3.72	0.05*	n/a
Posttest score			
	$F(1, 66)$	p	η_p^2
Complexity	7.24	0.01*	0.10
Social cues	5.90	0.02*	0.08
PK	40.01	<0.001*	0.39
Social cues x Complexity	0.02	0.90	<0.01
Complexity x PK	0.56	0.46	0.01
Social cues x PK	0.38	0.54	0.01
Complexity x social cues x PK	0.19	0.67	<0.01

Note: ATS = ANOVA Type Statistic of the non-parametric analysis.
*$\alpha < 0.05$.

detail in the future, it makes sense to use a detailed instrument for the measurement of cognitive load, as described above.

6.3 | Learning outcomes

Concerning learning outcomes, our findings showed that higher visual complexity resulted in lower learning outcomes, which is in line with our expectations (H4a). However, in contrast to our expectations, social cues were not beneficial for learning outcomes either overall (H4b) or as a function of visual complexity (H4c).

These findings align with the findings on learners' attention allocation, which showed that more visual search was required during the high complexity videos (see H2a), meaning that participants found and fixated on relevant content later. As a result, participants would have had less time to connect the verbal explanation with the corresponding visual information before the next item was mentioned (Van Gog, 2021). This impeded selection of information would subsequently also negatively affect the organization and integration of information (i.e., if it was not attended to it cannot be further processed; Davis et al., 2003; Harp & Mayer, 1998; Mayer, 2021; Mayer & Moreno, 2003; Van Gog, 2021), and thereby lead to lower learning outcomes (Atkinson, 2002; Valenzeno et al., 2003).

Since social cues did not improve speed of attention allocation either (see H2b) or in interaction with visual complexity (see H2c), it is not surprising that they did not affect learning outcomes. Interestingly, however, the exploratory analysis showed that when prior knowledge was included as an additional factor, both prior knowledge and social cues became significant predictors of learning outcomes. The effect of prior knowledge indicated that participants with high prior knowledge performed better on the posttest than low prior knowledge participants, presumably because they could integrate the new information presented in the videos with their prior knowledge retrieved from long-term memory (Richter et al., 2016), which facilitates learning (Van Kesteren et al., 2010). The fact that social cues did seem beneficial for learning when prior knowledge was taken into account as a factor in the analysis is interesting and relevant for future research as it may indicate that prior knowledge might play a role in measuring the effectivity of social cues on learning outcomes.

However, these findings should be interpreted with caution, as the analysis was exploratory and the sample size of the sub-groups was halved by adding prior knowledge as factor. This resulted in a small sample size in each condition, which lowers statistical power. Therefore, future studies with a larger sample size should investigate whether our outcomes could be replicated. We also recommend future studies to use a more elaborated measure of prior knowledge. We measured prior knowledge (i.e., whether the participants did an exam in biology or not in secondary education) with a single item. As such, we were unable to examine the internal consistency and validity of our prior knowledge measure. Using more items allows to more reliably and validly assess prior knowledge which can aid the investigation of the influence of prior knowledge on learning outcomes.

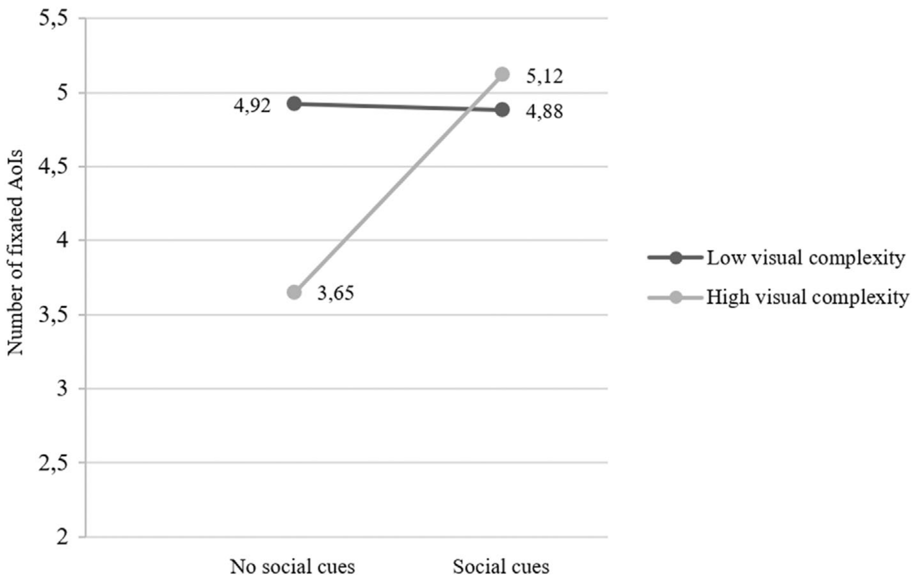

FIGURE 5 Number of fixated AOIs in both cue conditions and visual complexity conditions.

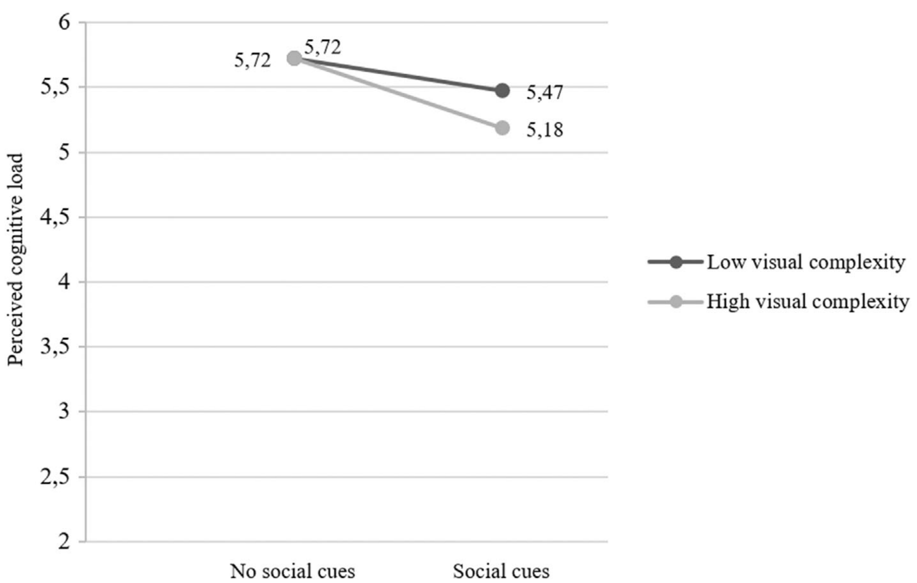

FIGURE 6 Perceived cognitive load in both cue conditions and visual complexity conditions for participants with low prior knowledge.

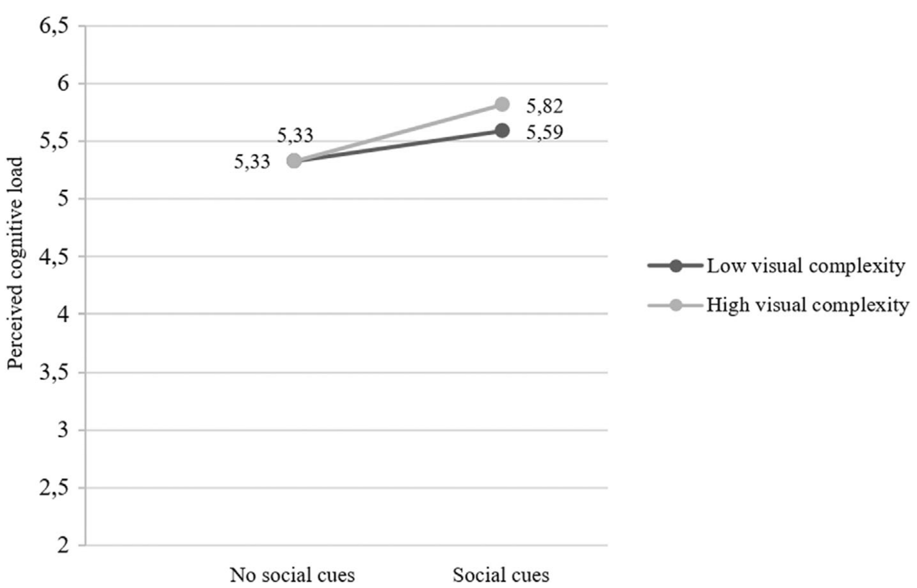

FIGURE 7 Perceived cognitive load in both cue conditions and visual complexity conditions for participants with high prior knowledge.

6.4 | Practical implications and directions for future research

Our study yielded two main findings that are also relevant for instructional designers seeking evidence-based guidelines for the design of instructional videos. First, our study underlines that the visual complexity of instructional videos significantly affected participants' attention allocation and learning outcomes. Under high visual complexity conditions, students were slower to find and fixate on the information that the instructor was referring to and (presumably as a consequence) had lower learning outcomes. As such, we recommend instructional practitioners to take into account the visual complexity when they design instructional videos. It is important to ensure that the visual complexity meets the capacities of learners to ensure that the video complexity does not impede learning.

Second, our study is the first that investigated the complex interplay between prior knowledge, gaze and gesture cues and visual complexity. In contrast to other studies, we did not find that social (i.e., gaze and gesture) cues affected learning outcomes, except for when we took prior knowledge into account as a factor in the analysis. Although our findings need to be further investigated with a larger sample, these preliminary findings do indicate the merits of using gaze and gesture cues in instructional videos. We therefore recommend practitioners to use gaze and gesture cues in instructional videos to enhance the learning process among learners. When doing so, practitioners need to take their learners' prior knowledge into account and tailor the instructional design of their instructional videos to the learners' specific requirements (Dochy et al., 1999). For example, practitioners could make decisions as to how often social cues need to be used and how visual complex the instructional video should be made.

In addition, we recommend researchers to continue studying the complex interplay of visual complexity and (social) cues in instructional videos. For example, it could be interesting to investigate the utility of using other social cues (e.g., first-person perspective, drawing; Mayer et al., 2020) under different complexity conditions. This could give us more insight into how effective social cues are for attention allocation, cognitive load and learning. Eventually, this can help establish refining design guidelines that will improve the quality of instructional videos and increase students' learning.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12803.

DATA AVAILABILITY STATEMENT

Data were generated at Utrecht University. Data supporting the findings of this study are available from the corresponding author on request.

ORCID

Julius Meier https://orcid.org/0000-0003-2671-4995
Bastian de Jong https://orcid.org/0000-0001-7848-1527
Margot van Wermeskerken https://orcid.org/0000-0003-3883-9875

REFERENCES

Altmann, G. T. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, *137*, 190–200. https://doi.org/10.1016/j.actpsy.2010.09.009

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, *94*, 416–427. https://doi.org/10.1037//0022-0663.94.2.416

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332. https://doi.org/10.1207/s1532690xci0804_2

Cohen, J. (1988). *Statistical power analysis the behavioral sciences* (2th ed.). Erlbaum.

Cutica, I., & Bucciarelli, M. (2008). The deep versus the shallow: Effects of co-speech gestures in learning from discourse. *Cognitive Science*, *32*, 921–935. https://doi.org/10.1080/03640210802222039

Davis, E. T., Shikano, T., Peterson, S. A., & Michel, R. K. (2003). Divided attention and visual search for simple versus complex features. *Vision Research*, *43*, 2213–2232. https://doi.org/10.1016/S0042-6989(03)00339-0

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, *21*, 113–140. https://doi.org/10.1007/s10648-009-9098-7

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, *69*(2), 145–186. https://doi.org/10.3102/00346543069002145

Fiorella, L., & Mayer, R. E. (2016a). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology*, *108*(4), 528–546. https://doi.org/10.1037/edu0000065

Fiorella, L., & Mayer, R. E. (2016b). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest science learning. *Journal of Educational Psychology*, *90*, 414–434. https://doi.org/10.1037/0022-6633.90.3.414

Hooge, I. T., & Camps, G. (2013). Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. *Frontiers in Psychology*, *4*, 1–10. https://doi.org/10.3389/fpsyg.2013.00996

Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M., & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, *44*, 22–30. https://doi.org/10.1016/j.learninstruc.2016.02.004

Johnson, A. M., Ozogul, G., & Reisslein, M. (2015). Supporting multimedia learning with visual signalling and animated pedagogical agent: Moderating effects of prior knowledge. *Journal of Computer Assisted Learning*, *31*(2), 97–115. https://doi.org/10.1111/jcal.12078

Kalyuga, S. (2009). Instructional designs for the development of transferable knowledge and skills: A cognitive load perspective. *Computers in Human Behavior*, *25*(2), 332–338. https://doi.org/10.1016/j.chb.2008.12.019

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*, 351–371. https://doi.org/10.1002/(sici)1099-0720(199908)13:4<351::aid-acp589>3.0.co;2-6

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, *38*(3), 209–215. https://doi.org/10.1007/s11251-009-9102-0

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*(1997), 1–18. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, *48*(1), 45–77. https://doi.org/10.1007/s11251-020-09502-9

Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59. https://doi.org/10.1016/S1364-6613(99)01436-9

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*, 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Mayer, R. E. (2021). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press. https://doi.org/10.1017/2F9781108894333.008

Mayer, R. E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development*, *68*(3), 837–852. https://doi.org/10.1007/s11423-020-09749-6

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52. https://doi.org/10.1207/S15326985EP3801_6

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, *50*(12), 1–23. https://www.jstatsoft.org/article/view/v050i12

Olivia, A., Mack, M. L., Shrestha, M., & Peeper, A. (2004). Identifying the perceptual dimensions of visual complexity of scenes. *Proceedings of the Cognitive Science Society*, *26*, 1040–1046 Retrieved from https://escholarship.org/content/qt-17s4h6w8/qt17s4h6w8.pdf

Ouwehand, K., Van Gog, T., & Paas, F. (2015). Designing effective video-based modeling examples using gaze and gesture cues. *Educational Technology & Society*, *18*, 78–88 Retrieved from https://ro.uow.edu.au/cgi/viewcontent.cgi?article=2823&context=sspapers

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Pi, Z., Zhang, Y., Zhu, F., Chen, L., Guo, X., & Yang, J. (2021). The mutual influence of an instructor's eye gaze and facial expression in video lectures. *Interactive Learning Environments*, *1-18*, 1–18. https://doi.org/10.1080/10494820.2021.1940213

Pi, Z., Zhang, Y., Zhu, F., Xu, K., Yang, J., & Hu, W. (2019). Instructors' pointing gestures improve learning regardless of their use of directed gaze in video lectures. *Computers & Education*, *128*, 345–352. https://doi.org/10.1016/j.compedu.2018.10.006

Pi, Z., Zhu, F., Zhang, Y., Chen, L., & Yang, J. (2022). Complexity of visual learning material moderates the effects of instructor's beat gestures and head nods in video lectures. *Learning and Instruction*, *77*, 101520. https://doi.org/10.1016/j.learninstruc.2021.101520

Pomplun, M., & Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in human-computer interaction. Proceedings of the international conference on HCI. Retrieved from https://www.cs.umb.edu/~marc/lab/publications/pomplun_hci2003.pdf

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, *17*, 19–36. https://doi.org/10.1016/j.edurev.2015.12.003

Schwartz, D. L., Sears, D., & Chang, J. (2007). Reconsidering prior knowledge. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 319–344). Lawrence Erlbaum Associates Publishers.

Seeber, K. G. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target. International Journal of Translation Studies*, *25*, 18–32. https://doi.org/10.1075/target.25.1.03see

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*, 85–89. https://doi.org/10.1111/j.0956-7976.2005.00786.x

Stull, A. T., Fiorella, L., & Mayer, R. E. (2018). An eye-tracking analysis of instructor presence in video lectures. *Computers in Human Behavior*, *88*, 263–272. https://doi.org/10.1016/j.chb.2018.07.019

Stull, A. T., Fiorella, L., & Mayer, R. E. (2021). The case for embodied instruction: The instructor as a source of attentional and social cues in video lectures. *Journal of Educational Psychology*, *113*(7), 1441–1453.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*, 657–665. https://doi.org/10.1080/13506280500410949

Töpper, J., Glaser, M., & Schwan, S. (2014). Extending social cue based principles of multimedia learning beyond their immediate effects. *Learning and Instruction*, *29*, 10–20. https://doi.org/10.1016/j.learninstruc.2013.07.002

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*, 187–204. https://doi.org/10.1016/S0361-476X(02)00007-3

Van Gerven, P. W., Paas, F., Van Merriënboer, J. J., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, *41*(2), 167–174. https://doi.org/10.1111/j.1469-8986.2003.00148.x

Van Gog, T. (2021). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 221–230). Cambridge University Press. https://doi.org/10.1017/9781108894333.022

Van Kesteren, M. T., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *Journal of Neuroscience*, *30*, 15888–15894. https://doi.org/10.1523/jneurosci.2674-10.2010

Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2018). Effectiveness of eye movement modeling examples in problem solving: The role of verbal ambiguity and prior knowledge. *Learning and Instruction*, *58*, 274–283. https://doi.org/10.1016/j.learninstruc.2018.07.005

Van Marlen, T., Van Wermeskerken, M., & Van Gog, T. (2019). Effects of visual complexity and ambiguity of verbal instructions on target identification. *Journal of Cognitive Psychology*, *31*(2), 206–214. https://doi.org/10.1080/20445911.2018.1552700

Van Wermeskerken, M., Ravensbergen, S., & Van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior*, *89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Wang, H., Pi, Z., & Hu, W. (2019). The instructor's gaze guidance in video lectures improves learning. *Journal of Computer Assisted Learning*, *35*(1), 42–50.

Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research*, *34*, 1187–1195. https://doi.org/10.1016/0042-6989(94)90300-X

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, *42*, 2985–3004. https://doi.org/10.1016/S0042-6989(02)00388-7

How to cite this article: Meier, J., de Jong, B., van Montfort, D. P., Verdonschot, A., van Wermeskerken, M., & van Gog, T. (2023). Do social cues in instructional videos affect attention allocation, perceived cognitive load, and learning outcomes under different visual complexity conditions? *Journal of Computer Assisted Learning*, *39*(4), 1339–1353. https://doi.org/10.1111/jcal.12803

APPENDIX A

Prior knowledge was rated by authors two and four using the prespecified scoring table below. Participants received one point for each correct label.

TABLE A1 The used prespecified scoring form to measure prior knowledge

Leaf layers	Plant cell
Meristematisch weefsel	Plasmodesma
Primair topmeristeem	Celwand
Secundair meristeem	Chloroplast
Intercalair meristeem	Thylakoïde
Permanent weefsel	Zetmeelkorrel
Primair weefsel	Vacuole
Sponsparenchym	Tonoplast
Collenchym	Mitochondrion
Sclerenchymvezels	Peroxisoom
Epidermis	Cytoplasma
Speciaal permanent weefsel, klierweefsel	Kleine vestikels
Xyleem	Ruw endoplasmatisch reticulum
Tracheïden	Celkern
Tracheëen	Kemporie
Houtvezels	Kernmembraan
Houtparenchym	Nucleolus
Floëem	Ribosoom
Zeefvaten	Glad endoplasmatisch reticulum
Zeefcellen	Golgivesikels
Begeleidende cellen	Golgiapparaat
Bastvezels	Cytoskelet
Bastparenchym	Celmembraan
Bundelcellen	Cistern
Turgorcellen	Cytosol
Cuticula	Dictyosoom
Onderste epidermis	Flagel
Palissadeparenchym	Golgicomplex
Bovenste epidermis	Lysosoom
Stoma	Mictrotubulus
Vacuole	Nucleus
Nucleus	Trilhaar
	Vesikel
	Zweepstaartje
	Amyloplast
	Bladgroenkorrel
	Chromoplast
	Elaioplast
	Etioplast
	Gerontoplast
	Leukoplast
	Nucleomorf
	Plastide
	Proplastide
	Proteïnoplast
	Statoliet
	Plasmamembraam
	Celwand van aanliggende cel
	Centrale vacuole