Looking through Sherlock's eyes

van Marlen, Tim; van Wermeskerken, Margot; Jarodzka, Halszka; Raijmakers, Maartje; van Gog, Tamara

published in
Journal of Computer Assisted Learning
2022

DOI (link to publisher)
10.1111/jcal.12712

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

Received: 9 September 2021 | Revised: 18 April 2022 | Accepted: 15 May 2022

DOI: 10.1111/jcal.12712

ARTICLE

Journal of Computer Assisted Learning WILEY

Looking through Sherlock's eyes: Effects of eye movement modelling examples with and without verbal explanations on deductive reasoning

Tim van Marlen[1] | Margot van Wermeskerken[1] | Halszka Jarodzka[2] | Maartje Raijmakers[3,4] | Tamara van Gog[1]

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

[2]Welten Institute, Open University of the Netherlands, Heerlen, The Netherlands

[3]Developmental Psychology, Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

[4]Educational Studies, Faculty of Behavioral and Movement Sciences, Free University Amsterdam, Amsterdam, The Netherlands

Correspondence
Tim van Marlen, Department of Education, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands.
Email: tvavanmarlen@hotmail.com

Funding information
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 452-11-006

Abstract

Background: Eye movement modelling examples (EMME) are demonstrations in which learners' not only see a model's (e.g., a teacher's) task performance on a computer screen (as in regular video examples) but also the model's eye movements (represented as moving coloured dots overlaid on the screen). Thereby EMME help guide learners' attention towards the relevant information and can model cognitive strategies which are otherwise unobservable for learners.

Objectives: This study investigated whether EMME can help to learn deductive reasoning strategies and how the presence/absence of a teacher's verbal explanation affects learning from EMME.

Methods: Secondary education students ($N = 137$) were randomly assigned to study video examples under one of four conditions in a 2 (EMME: yes/no) x 2 (verbal explanations: yes/no) between-subjects design.

Results and Conclusions: Results revealed only a beneficial effect of the presence of verbal explanations on performance on the practice problems, but no pretest-to-posttest learning gains.

Implications: Seeing the teacher's eye movements does not appear to enhance learning of deductive reasoning. The presence/absence of the teacher's verbal explanation does not seem to affect learning deductive reasoning.

KEYWORDS
attention cueing, example-based learning, eye movement modelling examples, eye tracking

1 | INTRODUCTION

Studying video modelling examples is an effective way to learn new skills (Van Gog & Rummel, 2010). In video modelling examples a model, for instance a teacher or instructor, explains and demonstrates step by step how to perform a task. One of the possible advantages of studying video modelling examples is that relevant information can be presented to the learner both visually and verbally. The multimedia principle (Mayer, 2014) states that this can enhance learning. Especially when combined with the modality principle (Mousavi et al., 1995), which states that by presenting information in multiple modalities (e.g., visualizations with narrated rather than written verbal explanations) the learner is less likely to be cognitively overloaded, by spreading information processing across the visual and phonological working memory channels (Baddeley, 1992), and information integration is facilitated, leading to better learning outcomes. Observational learning from multimedia materials like video modelling examples has been shown to be effective for learning (Van Gog, Rummel, et al., 2019). However, the effectiveness can be affected by how the learning materials are designed. A lot of research within the field of multimedia learning has focused on which design characteristics are important to take into account when designing learning content. Some examples of design characteristics are: the use of static or dynamic visualizations (Höffler & Leutner, 2007), the visual presence or absence of the person who demonstrates and explains the task (Van Wermeskerken et al., 2018; Wang & Antonenko, 2017), the presence or absence of visual (Richter et al., 2016) and social cues (Mayer, 2014), or whether the video example is recorded in first-person perspective or in third-person perspective (Fiorella et al., 2017).

The current study focuses on one specific type of video modelling examples, referred to as ‘eye movement modeling examples’ (EMME; Van Gog, Jarodzka, et al., 2009), which consists of a screen recording that shows the learner a demonstration by a model (e.g. a teacher or an expert) of how to perform a task, while simultaneously showing where the model was looking during this task demonstration (e.g. as a coloured dot or circle). The aim of the current study was to investigate whether EMME can foster the learning of cognitive problem solving strategies and whether this depends on the absence or presence of the model's verbal explanation.

2 | EYE MOVEMENT MODELLING EXAMPLES

As stated above, EMME are video examples in which you see a screen recording of a model solving a task while also seeing where the model was looking at during task performance. EMME can serve two functions. The first function of EMME is that they align the learner's attention with that of the model, which then can help the learner to select and integrate the relevant information of the task demonstration. Compared to regular modelling examples (ME; i.e., video examples of screen recordings without the model's eye movements superimposed) in EMME the learner's attention is directed and synchronized with that of the model, thus creating a state of *joint attention* (i.e., the phenomenon of automatically attending an object someone else is attending; Brennan et al., 2008; Frischen et al., 2007). The creation of joint attention between the model and the student is important as information within video examples are often transient (i.e., information is only temporarily available) meaning that students can miss out on important information when they do not attend to the right information at the right time, which can negatively affect students' learning (Ayres & Paas, 2007). By offering attentional guidance by means of EMME, the risk of not attending important information can be reduced. In this way, the learner's processing of visual information (i.e. the on screen learning material with visible interactions—e.g. clicks, drags, typing- of the model with the material) and visual-verbal information (i.e. on screen learning material with the model's verbal explanation) can be facilitated.

That the attentional guidance provided by EMME affects the learner's visual attention, aligning it more with the model's visual attention, and also enhances learning, is supported by research. Several studies have compared the effects of EMME with ME on attention allocation (by measuring the learner's eye movements during example study) and learning outcomes (i.e. post-test performance; Jarodzka et al., 2012, 2013; Van Marlen et al., 2016, 2018). For instance, the studies by Jarodzka et al. (2012, 2013) demonstrated that EMME compared to ME enhanced learning to perform classification tasks. In the study by Jarodzka et al. (2012) EMME were used to help students learn how to classify the symptoms of epileptic seizures of infants by showing an expert's eye movements while the expert watched case videos of patients and additionally provided verbal explanations. In the control condition, participants saw the same case videos with the verbal explanation but without the expert's superimposed eye-movements. In the study of Jarodzka et al. (2013), EMME were used to help students to learn to classify the locomotion patterns of fish. Again, the eye-movements and verbal explanations of an expert were recorded and shown to participants. Besides the learning benefits, the studies of Jarodzka et al. (2012, 2013) also showed that the learner's eye movements while watching the video examples in the EMME conditions were more similar to those of the model than in the ME condition, as evidenced by a higher scanpath similarity (Jarodzka et al., 2013) or smaller Euclidean distances between the model's gaze position and that of the learner (Jarodzka et al., 2012). In addition, Van Marlen et al. (2018) have recently demonstrated that college students watching EMME fixated on the verbally referred visual task elements more often and faster than students watching ME.

The second function of EMME is that they make it possible to visualize perceptual and cognitive strategies that would otherwise remain unobservable for learners. This can be done either in the presence or absence of the model's verbal explanation. EMME *without* the model's verbal explanation have been shown to enhance study strategies for digital hyperlinked texts (Salmerón & Llorens, 2018) and illustrated texts (in seventh grade students: Mason et al., 2015, 2016, 2017; in college students: Scheiter et al., 2018). More specifically, the studies by Mason et al. (2015, 2016, 2017) examined whether observing EMME prior to studying an illustrated text would enhance text picture integration during study. In the EMME, the model demonstrated how to integrate information from the text and picture by making transitions between certain terms in the text and the corresponding part of the picture. Compared to students who did not observe EMME (i.e., participants in the passive control group were only presented with the text without any other instructions), students in the EMME condition showed better text-picture integration while studying the (new) illustrated text, and also performed better on a text comprehension test.

That EMME without verbal explanations can also model perceptual strategies was demonstrated in studies showing that EMME enhance visual search performance when people had to search for errors in software code (Stein & Brennan, 2004), errors on printed circuit boards (Nalanagula et al., 2006) or lung-nodules on X-ray scans (Litchfield et al., 2010). The study of Stein and Brennan (2004) used a within-subject design in which for half of the software bugs participants saw EMME and for the other half of the bugs the participants did not receive an example video. In the study of Nalanagula et al. (2006) several types of EMME visualizations were compared with a control condition in which participants were presented with the printed circuit boards without any examples, whereas Litchfield et al.

(2010) compared several types of EMME's showing different eye movement patterns (that were either deliberately recorded/selected or originated from models with different levels of expertise).

Other research has shown that EMME *with* verbal explanations can be used to model inspection strategies aimed at enhancing to learn to classify in classification tasks (Jarodzka et al, 2012, 2013, Vitak et al., 2012) and can model procedural problem-solving strategies to enhance problem solving (Van Marlen et al., 2018, Experiment 2). For instance, in the study of Van Marlen et al. (2018, Experiment 2) EMME demonstrated procedural problem-solving strategies about solving geometry problems. In this study students watched EMME in which they saw the model making transitions between elements of the problem (i.e. the different angles that had to be solved) while verbally explaining the underlying principles to solve the angles.

In contrast, there is also research in which EMME demonstrated procedural problem-solving strategies but were not effective to enhance learning even though the model's verbal explanations were included (Van Gog et al., 2009). For instance, in the study by Van Gog et al. (2009) students watched EMME or ME about how to solve a procedural puzzle problem with or without the model's verbal explanations. The procedural puzzle problem was an isomorph of the Tower of Hanoi problem (Newell & Simon, 1972) in which puzzle pieces (frogs) had to be moved in a specific order from one side to the other side. The model in the video examples verbally explained from the start till the end of the problem how to solve this problem. Results revealed no positive effect of EMME on learning. However, on the transfer task it was even found that students who watched EMME including the model's verbalization had lower performance than students who watched the EMME without the model's verbalization. An alternative explanation given by the authors for these findings was that perhaps the verbal explanations were already sufficient for the students to guide their attention. One possible limitation of the study was that the verbal explanation in combination with the model's eye movements may have been redundant as the model also used the mouse cursor to perform problem-solving steps, and this also guided students' attention. In this context it is possible that the verbal explanation in combination with the visible interaction of the model with the puzzle environment made the EMME redundant. Research on this so-called *redundancy effect* has shown that the addition of redundant information hampers learning instead of facilitating learning (Kalyuga & Sweller, 2014). This suggests that the model's verbalizations might play an important role in learning from EMME.

In sum, some studies (Jarodzka et al., 2012, 2013; Van Marlen et al., 2018; Vitak et al., 2012) found positive effects of EMME conveying perceptual/cognitive strategies on learning if the EMME included verbal explanations, whereas other studies found a negative effect (Van Gog et al., 2009) or no effect (Van Marlen et al., 2016) of EMME with verbal explanations on learning. Drawing conclusions about the role of the model's verbalizations on learning from EMME remains difficult because, with the exception of Van Gog et al. (2009), none of the discussed studies have manipulated the presence or absence of verbal explanations. Therefore, in the current study we aimed to examine whether the presence or absence of verbal explanations affects learning deductive reasoning strategies from EMME. In the present study, possible redundancy of the verbal explanations should not be an issue, as we used a deductive reasoning task called Mastermind (see more details below) in which several sources of visual information need to be integrated in order to solve the problem. Thus, the model's visualized eye-movements indicated which sources needed to be integrated, and the model's verbal explanation indicated how they needed to be integrated.

3 | THE PRESENT STUDY

The aim of the present study was to examine whether EMME would be effective for fostering learning of cognitive strategies (more specifically: deductive reasoning) and whether this would be affected by the presence or absence of the model's verbal explanations. Secondary education students were presented with video modelling examples in which the model demonstrated how to break the code in a Deductive Mastermind task, either with or without verbal explanation. The Deductive Mastermind task (cf. Gierasimczuk et al., 2013) is an adapted version of the classic board game Mastermind. A code has to be deduced from a set of code breaking attempts. In the Deductive Mastermind task, the code breaker is provided with an image depicting several code-breaking attempts and corresponding feedback, which the code breaker must use to deduce the correct code by systematically comparing the entered codes with its corresponding feedback. In the EMME, the cognitive strategies involved in systematically comparing the entered codes and feedback become visible through the depiction of the model's eye movements. In the present study, a 2×2 between-subject design was used in which the students received video examples with (EMME) or without (ME) visualizations of the model's eye movements and with or without the model's verbal explanation.

We expected that the students in the EMME conditions would benefit from seeing the model's problem-solving strategies and would therefore show greater learning gains from pretest to posttest and the near transfer problems than students in the ME conditions. Based on the multimedia principle (Mayer, 2014) and the modality principle (Mousavi et al., 1995), we expected students in the verbal explanation conditions to show higher learning gains than students in the no verbal explanation conditions. Even though Van Gog et al. (2009) found that the presence of a verbal explanation presumably made the attention guidance provided by the EMME redundant, we expected that the presence of verbal explanations would further enhance learning from EMME in the present study, as the attention guidance provided by the EMME would complement the strategy information (i.e. the model demonstrated multiple deductive reasoning strategies which had to be combined to solve the problems) conveyed in the verbal explanations.

4 | METHOD

4.1 | Participants

One hundred and forty Dutch secondary education students in their first year of pre-university education (the track that prepares students for enrollment at a university, and has a six-year duration) were recruited out of five classrooms all from the same school. Three students did not provide informed consent to use their data and were therefore excluded from the study. The final sample consisted of 137 students ($M_{age} = 12.66$, $SD = 0.51$, 64 male). Participants were randomly assigned to one of four conditions resulting from a 2 (modelling example: EMME vs. ME) $\times$ 2 (verbal explanation: present vs. absent) between-subjects design: EMME verbal explanation present ($n = 34$), EMME verbal explanation absent ($n = 35$), ME verbal explanation present ($n = 33$), ME verbal explanation absent ($n = 35$).

5 | MATERIALS

5.1 | Deductive mastermind task

The deductive reasoning task used in the current study is an adaptation of the board game Mastermind. This Deductive Mastermind task is part of the popular online learning platform called 'Math Garden' (in Dutch 'de Rekentuin' www.mathgarden.com; Gierasimczuk et al., 2013). In the digital Deductive Mastermind game, the learner plays the role of a code breaker. The task for the learner (code breaker) is to unravel the correct code. To do so, the learner is shown an image depicting previously entered codes with the corresponding feedback (see Figure 1). The learner must use the feedback of all previous code breaking attempts, to enter the correct code, and code breaker is given only one chance to break the code. Hence the term 'Deductive Mastermind', as the learner has to deduce the correct code by processing all the feedback of earlier code breaking attempts. In the current study, the coloured pins of the deductive Mastermind task were represented as coloured flowers. In addition, the feedback pins were represented as coloured dots: a green dot for correct flower and location, an orange dot for correct flower only (i.e. wrong location), and a red dot for stating that a flower does not occur in the code.

5.2 | Mastermind test problems

The Mastermind problems used in the present experiment differed in terms of the number of flowers in the code (two, three, or four flowers) and the number of code-breaking attempts being shown (i.e., this ranged between two to five attempts). The pretest comprised of six problems in which the code consisted of two flowers and six problems in which the code consisted of three flowers resulting in a total of 12 problems. Based on difficulty ratings obtained from the database of 'Math Garden' (Gierasimczuk et al., 2013) we selected problems of which half of the two flower codes and half of the three flower codes were rated as easy and the remaining half as difficult.

For the two video examples we used Mastermind problems that consisted of three flower codes, each showing three code breaking attempts. After each video example students were presented with the opportunity to practice applying the modelled strategy themselves on an isomorphic practice problem (created by replacing the flowers with different flowers so that the task looked different, but it was structurally identical).

Twelve post-test problems were created by replacing the flowers of the pretest problems with different types of flowers, so that the posttest problems looked different from the pretest problems but were otherwise identical. See Figure 1 for an example.

Finally, six near transfer problems were created, which were more complex than the pretest/posttest problems and consisted of four flower codes. Within the four-flower code category half of the near transfer problems were rated as easy and the other half as difficult based on the 'Math Garden' database.

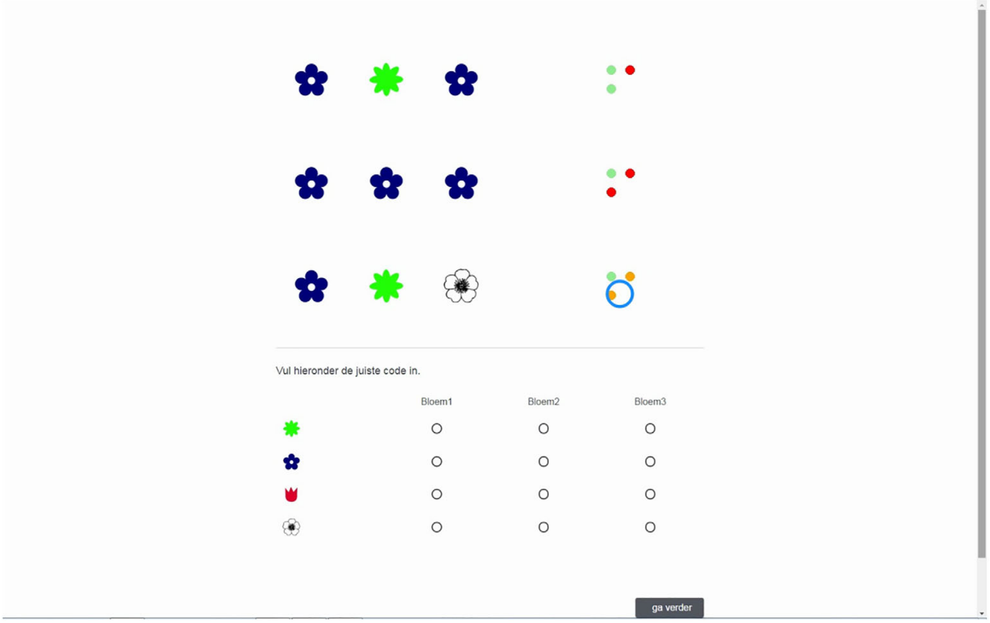


FIGURE 1 A screenshot of an eye movement modelling example depicting the mastermind task. The blue circle represents the gaze location of the model, in this case inspecting the feedback of the third code-breaking attempt. On top the three rows with the flower codes along with the corresponding feedback on the right are displayed. Underneath the codes the answer selection pane with the different possible flower types are displayed

5.3 | Eye movement modelling examples

A SMI 250 Hz remote eye tracker (SensoMotoric Instruments, GmbH) was used to create the two EMME videos. SMI Experiment Center 3.7.60 software to display the tasks and iViewX 2.8 software was used for recording the model's eye movements while the model completed the tasks. After recording, SMI BeGaze 3.7 software was used to visualize the model's eye movements in the videos of the screen-recording of the task. The eye movements were represented as a moving blue coloured circle with a diameter of 30 pixels and a line width of 3 pixels.

In the EMME videos, the male model started by inspecting the flower codes and the corresponding feedback from top to bottom. Once all the codes and feedback were inspected, the model made transitions between two different lines of code and the corresponding feedback, hereby implying that the comparison of these two lines of code and feedback would provide information regarding the correct code. In the conditions in which verbal explanations were also present, the model was then explaining strategies regarding what could be deduced from these lines of code and feedback. In the video examples, three strategies were demonstrated. The first strategy is the least difference strategy. In this strategy, two rows of code are compared in which only one flower in the code differs from the other code. By looking at the corresponding feedback, you can deduce whether this single difference in the code indicates whether a particular flower belongs in the code. The second strategy being demonstrated in the video examples is the integration of knowledge. By this we mean that the knowledge of a part of the code is further used to deduce later part(s) of the code. For example, knowing that the leftmost flower of the code must be a green flower instead of a blue flower might enable you to deduce that the blue flower must be placed on a different location in the code. The third strategy concerns the usefulness of feedback. If the feedback of code-breaking attempt consists of only red pins, then you can deduce that no flower in that code is present in the final code. The other way around, if the feedback pins consist of green and orange pins, then you can deduce that at least you know which type of flowers must be present in the final code. One video example demonstrated all strategies with the main focus on the usefulness of feedback strategy (the occurrence of two orange feedback pins indicating that the correct flowers of the code was present but not yet placed in the correct order) and the other video example demonstrated the least difference strategy and the integration of knowledge strategy. The least difference strategy was needed in nine pretest/posttest problems and in five near transfer problems, the integration of knowledge strategy was needed in seven pretest/posttest problems and in three near transfer problems, and the usefulness of feedback strategy was needed in seven pretest/posttest problems and in five near transfer problems.

In the EMME, the model made comparisons (i.e. comparisons between lines of code and corresponding feedback) until the full code was deduced. Every solved part of the code, or sub step of the problem, was entered immediately by clicking on the correct answer option underneath the problem and was visible in the screen recording of the task. The model behaved didactically, so the eye movements between the code and feedback were very deliberate. The length of the two EMME videos was 135 and 139 s, respectively. For the regular modelling example conditions (ME conditions), the screen recordings were exported without the eye movements superimposed. The screen recordings were exported with or without the verbal explanations. Therefore, the videos across conditions were equal regarding the screen recording and length and only differed regarding the presence/absence of the model's eye movements and presence/absence of the verbal explanation. See Appendix for the translated transcript of the verbal explanation about how to solve the Mastermind problem depicted in Figure 1.

5.4 | Experimental measures

5.4.1 | Proportion correct mastermind problems

One point was given for each correctly solved Mastermind problem. In total, students could earn 12 points for the pretest, 12 points for the posttest, 2 points for the isomorphic problems and 6 points for the near transfer problems. For the pretest, isomorphic problems and near transfer problems the proportion of correctly solved problems was calculated by dividing the number of points by the maximum obtainable points.

5.4.2 | Pre-test to post-test gain score

To measure the student's progression from pretest to posttest, we calculated a gain score. For each participant a gain score was calculated as the number of points earned in the posttest minus the number of points earned in the pretest (post-test–pre-test).

5.5 | Procedure

The experiment was conducted during a math lesson that lasted approximately 50 min. The tables in the classroom were separated to ensure the students would not collaborate or look at each other's laptop screens. On each table a sheet of paper was placed stating the student name, participant number, version number, headphone (if a student was assigned to a condition including verbal explanations) and a website URL. As the students entered the classroom they were instructed to find their table and to take out their laptop. Once every student was seated, the experimenter gave general instructions about the experiment and answered practical questions. After these instructions the students were asked to type in the URL, which opened the online questionnaire used for the experiment hosted by Qualtrics (www.qualtrics.com) and to type in the given participant number and experiment version. Subsequently, the students were instructed to start with the online program at their own pace. The program then started by asking the informed consent followed by demographic questions (age and gender). Students then received instructions about the Mastermind task stating that the students had to break the code by using the feedback provided for each code-breaking attempt. The students were explained the meaning of the different types of feedback (i.e. what the different coloured pins meant). However, how the students should use the feedback to unravel the code was not explained. After the Mastermind instructions, students worked on the 12 pretest problems. Then the students were presented with the video examples in random order. They received the instruction that they were about to see how someone else solved a Mastermind problem, and only students in the EMME condition were additionally instructed that the blue circle showed them where the person had looked while solving the problem. Students in the conditions with verbal explanations in the examples were asked to check whether the headphone was connected properly and the volume was on. Then the students watched the video example, followed by the corresponding isomorphic problem, and the second video example followed by the corresponding isomorphic problem. Subsequently, students received the 12 post-test problems followed by the six near transfer problems. The order of the problems in the pre-test, the video examples, the problems of the post-test, and the near transfer problems was random. The answer options selected by the students were registered by the Qualtrics software.

5.6 | Data analysis

One student was unable to finish the pre-test due to technical problems, and was therefore excluded from all analyses. In addition, eight students were unable to finish the isomorphic practice problems in time, 25 additional students were unable to finish the post-test in time, three students were considered outliers on the post-test (i.e. absolute *z*-score larger than 2.5), and an additional 14 students did not finish the near transfer problems in time. Table 1 shows the total number of participants per condition included in the different types of analyses.

The data were analysed with 2 (modelling example: EMME vs. ME) x 2 (verbal explanation: present vs. absent) ANOVAs and partial eta squared is reported as a measure of effect size, with $\eta_p^2 = 0.01$, $\eta_p^2 = 0.06$, $\eta_p^2 = 0.14$, representing small, medium, and large effects respectively (Cohen, 1988). Due to the violation of normality assumption regarding all outcome measures, we conducted additional non-parametric Kruskal-Wallis tests to check whether the results of the main effects of the ANOVAs would hold. These non-parametric analyses revealed the same pattern of results as found with the ANOVAs. Non-parametric Mann–Whitney *U* tests were conducted to check for possible interactions (EMME verbal vs. ME verbal, and EMME no verbal vs. ME no verbal), however like the results of the ANOVAs no significant interactions were found. Additionally, Bayesian ANOVAs were conducted with JASP (version 0.8.6; jasp-stats.org; JASP team, 2019; Wagenmakers et al., 2018). The advantage of Bayesian analyses is that instead of simply rejecting the null hypothesis, Bayesian analyses provide an estimate of how much more

TABLE 1 Mean (and SD) and median (and min–max) for the pretest, the gain score (posttest–pretest), and the near transfer problems for the four conditions: Modelling example (EMME vs. ME) × verbal explanations (verbal explanations vs. no-verbal explanations)

	EMME		ME	
	Verbal explanations *M* (*SD*) | *Mdn* (min–max)	No-verbal explanations *M* (*SD*) | *Mdn* (min–max)	Verbal explanations *M* (*SD*) | *Mdn* (min–max)	No-verbal explanations *M* (*SD*) | *Mdn* (min–max)
Proportion Correct				
Pretest ($n = 136$)[a]	0.43 (0.29) | 0.33 (0.00–0.83)	0.40 (0.32) | 0.33 (0.00–1.00)	0.39 (0.30) | 0.42 (0.00–0.92)	0.34 (0.26) | 0.25 (0.00–0.83)
Isomorphic ($n = 128$)[b]	0.65 (0.42) | 1.00 (0.00–1.00)	0.34 (0.37) | 0.50 (0.00–1.00)	0.55 (0.42) | 0.50 (0.00–1.00)	0.32 (0.38) | 0.00 (0.00–1.00)
Posttest ($n = 100$)[c]	0.51 (0.26) | 0.58 (0.00–0.92)	0.34 (0.28) | 0.25 (0.00–0.92)	0.47 (0.28) | 0.46 (0.00–0.83)	0.38 (0.29) | 0.38 (0.00–0.83)
Near Transfer ($n = 86$)[d]	0.23 (0.22) | 0.17 (0.00–0.67)	0.15 (0.17) | 0.17 (0.00–0.50)	0.14 (0.19) | 0.00 (0.00–0.50)	0.20 (0.21) | 0.17 (0.00–0.67)
Gain score ($n = 100$)[c]	0.63 (1.95) | 0.00 (−2.00–4.00)	0.46 (2.49) | 0.00 (−4.00–6.00)	0.46 (1.77) | 0.00 (−3.00–4.00)	0.50 (2.30) | 0.50 (−5.00–6.00)

[a]The number of participants per condition for the Pretest are: EMME Verbal ($n = 34$), EMME No-Verbal ($n = 35$), ME Verbal ($n = 33$), and ME No-Verbal ($n = 34$).
[b]The number of participants per condition for the Isomorphic problems are: EMME Verbal ($n = 34$), EMME No-Verbal ($n = 31$), ME Verbal ($n = 31$), and ME No-Verbal ($n = 32$).
[c]The number of participants per condition for the Posttest and Gain score are: EMME Verbal ($n = 24$), EMME No-Verbal ($n = 26$), ME Verbal ($n = 24$), and ME No-Verbal ($n = 26$).
[d]The number of participants per condition for the Near Transfer problems are: EMME Verbal ($n = 21$), EMME No-Verbal ($n = 19$), ME Verbal ($n = 21$), and ME No-Verbal ($n = 25$).

likely the alternative or null hypothesis is compared to the other hypothesis given the obtained data. The inclusion Bayes Factor (BF_{inc}), which is an estimate of the likelihood of the model if it contains the effect, were reported for the main analyses. For example, a $BF_{inc} = 20.00$ when reporting a main effect of verbal explanation would indicate that the model is 20 times more likely than the null model without an effect of verbal explanation.

6 | RESULTS

The performance data of the pretest problems, isomorphic problems, near transfer problems, and the gain scores are presented in Table 1. We first examined whether the prior knowledge measured as the performance on the pretest was equal across conditions. Results of this 2 × 2 ANOVA with the proportion correct of the pretest as dependent variable indicated no main effect of modelling example, $F(1, 132) < 1.00$, $p = 0.341$, no main effect of verbal explanation, $F(1, 132) < 1.00$, $p = 0.380$, and no interaction, $F(1, 132) < 1.00$, $p = 0.829$.

6.1 | Performance isomorphic practice problems

To test our hypothesis that students would learn more from an EMME than a regular modelling example and whether learning was affected by the presence/absence of verbal explanations, a 2 × 2 ANOVA with the proportion correctly solved isomorphic problems as the dependent variable was conducted. The results revealed no main effect of modelling example, $F(1, 124) < 1.00$, $p = 0.394$, $\eta_p^2 < 0.01$, $BF_{inc} = 0.234$, a main effect of verbal explanation, $F(1, 124) = 15.04$, $p < 0.001$, $\eta_p^2 = 0.10$, $BF_{inc} = 81.114$, indicating that students presented with video examples that also included verbal explanations outperformed students who did not receive verbal explanations. There was no significant interaction, $F(1, 124) < 1.00$, $p = 0.581$, $\eta_p^2 < 0.01$, $BF_{inc} = 0.268$.

6.2 | Pretest to post-test gain score

To test whether students with EMME showed larger learning benefits than students with the regular video examples and to examine whether this is influenced by the presence/absence of verbal explanations, a 2 × 2 ANOVA[1] with the gain score as the dependent variable was conducted. The results revealed no main effect of modelling example, $F(1, 96) < 1.00$, $p = 0.882$, $\eta_p^2 < 0.01$, $BF_{inc} = 0.148$, no main effect of verbal explanation, $F(1, 96) < 1.00$, $p = 0.888$, $\eta_p^2 < 0.01$, $BF_{inc} = 0.148$, and no significant interaction, $F(1, 96) < 1.00$, $p = 0.813$, $\eta_p^2 < 0.01$, $BF_{inc} = 0.034$.

6.3 | Performance near transfer problems

A 2 × 2 ANOVA with the proportion correctly solved near transfer problems as dependent measure was conducted to examine whether students in the EMME conditions outperformed students in the ME conditions on the near transfer problems and whether this was influenced by the verbal explanations. The results revealed no main effect of modelling example, $F(1, 82) < 1.00$, $p = 0.674$, $\eta_p^2 < 0.01$, $BF_{inc} = 0.188$, no main effect of verbal explanation, $F(1, 82) < 1.00$, $p = 0.783$, $\eta_p^2 < 0.01$, $BF_{inc} = 0.178$, and no significant interaction, $F(1, 82) = 2.56$, $p = 0.114$, $\eta_p^2 = 0.03$, $BF_{inc} = 0.124$.

To summarize, with the exception of an effect of verbal explanation on performance on the isomorphic practice problems, the experimental conditions did not significantly affect any other outcome measures.

7 | DISCUSSION

The aim of the present study was to examine whether EMME would be effective for fostering learning of cognitive strategies and whether this would be affected by the presence or absence of the model's verbal explanations. Thus, the present study set out to conceptually replicate Van Gog et al. (2009), with a deductive reasoning task. We hypothesized that the students in the EMME conditions would show greater learning gains from pretest to posttest than students in the regular modelling example conditions (i.e. video modelling example without the model's eye movements superimposed). In addition, we expected that EMME with verbal explanations would be more effective than all other conditions (i.e., interaction effect) as the attention guidance provided by the EMME would complement the strategy information conveyed in the verbal explanations.

In contrast to our hypothesis, we found no effect of attention guidance: there were no differences in learning outcomes between the EMME and ME conditions. The only exception is the finding that students performed better on the isomorphic practice problems if the modelling example included verbal explanations. However, this benefit of having heard the model's verbal explanations during the video examples did not translate into a higher learning gain score or better performance on the near transfer problems. Also, in contrast to our hypothesis we found no interaction between the type of modelling example (EMME vs. ME) and the presence/absence of verbal explanations. Thus, although we did not find a negative effect of EMME with verbal explanations on learning as found by Van Gog et al. (2009), we also did not find the positive effects of EMME with verbal explanations often found with classification tasks (Jarodzka et al., 2012, 2013; Vitak et al., 2012).

[1]Upon request of a reviewer we additionally conducted a mixed ANOVA in order to test to what extent participants improved in solving the Mastermind problems from pretest to posttest. The mixed ANOVA with the pretest and posttest scores as repeated measure and the experimental conditions as fixed effects, revealed a main effect of test, $F(1, 96) = 5.61$, $p = 0.020$, $\eta_p^2 = 0.06$, indicating that regardless of any experimental conditions participants slightly improved from pretest to posttest. There were no further significant interactions between the repeated measure and the fixed factors. Regarding the between-subject effects, the only significant effect was a main effect of verbal explanation, $F(1, 96) = 5.78$, $p = 0.018$, $\eta_p^2 = 0.06$, indicating that participants whom received a verbal explanation vs. no verbal explanation slightly improved from pretest to posttest.

A possible explanation for the lack of differences in learning outcomes between the conditions might be due to the difficulty of the task. Support for the assumption that students experienced the task to be difficult stems from the fact that across conditions the gain score was below one. This means that from pretest to posttest, students hardly improved in solving the mastermind problems even though on average, students in the verbal explanation conditions solved more than half of the isomorphic problems. This suggests that students were able to apply the strategies that were just demonstrated in the example, on an isomorphic practice problem, but could not apply what they had learned to solve the posttest problems. This is further underlined by the low performance on the near transfer problems. It should be noted that the statistical power to detect possible differences between conditions might have been low due to the limited number of participants being able to finish making the near transfer problems. However, the results of the Bayesian analyses suggest that it is unlikely that a slightly larger sample would have drastically changed the results.

A limitation of these findings is that they make it difficult to draw firm conclusions regarding whether the effectiveness of EMME is moderated by the presence of verbal explanations. Taken as a whole, our data suggest that students were unable to abstract the cognitive strategies being demonstrated by the model, irrespective of whether they were accompanied by verbal explanations. Possibly, students would need more example-problem pairs in the learning phase to practice how to apply the strategies. The use of two example-problem pairs in the learning phase is similar to a previous EMME study that found EMME to enhance learning of geometry problem solving (Van Marlen et al., 2018). However, the current study differs with respect to the task being demonstrated. Perhaps the underlying cognitive strategies involved in the current task are too difficult to grasp with only two example-problem pairs, especially since the students did not receive feedback. An interesting research avenue could be to see whether students would learn more if the learning phase consisted of video examples with the corresponding isomorphic practice problems followed by feedback on their performance, for instance by means of a correct demonstration (i.e. video example) of the isomorphic practice problem.

The finding that students were unable to abstract and/or apply the modelled cognitive strategy for solving the posttest and near transfer problems, is at odds with research regarding enhancing text and picture integration with EMME (Mason et al., 2015, 2016, 2017). However, in these studies the modelled strategy was arguably less complex: the model in the EMME demonstrated how to read and process texts by making transitions between key concepts in the text and the corresponding elements in the picture (and this resulted in higher performance on text comprehension tests compared to a control condition). In the present study, the model in the EMME conditions emphasized how to compare and deduce the correct code by making transitions back and forth between specific parts of the code and the corresponding feedback. However, in the current study the function of EMME was not only to indicate which information sources needed to be integrated but also to convey the underlying cognitive strategies involved to solve the deductive reasoning problems. It is possible that the combination of both attending the video example while also trying to understand the underlying cognitive strategies was too demanding for the students.

A study by Scheiter et al. (2018) also illustrates that students may not be able to learn from EMME when these are too demanding. Their findings suggest that in order to fully benefit from the attentional guidance in EMME the student needs to have a minimal amount of cognitive prerequisites (i.e., more broad knowledge regarding scientific thinking enabling the student to understand instructions, thus not to be confused with specific prior knowledge). In their study, students either watched EMME demonstrating how to process multimedia learning materials regarding mitosis or students were given an equivalent amount of time to read the illustrated text in the control condition. Results showed that on learning outcomes that required more in-depth processing of the learning materials actually the stronger students (i.e., with higher cognitive prerequisites) benefitted more from having watched EMME compared to weaker students. Similar results were also found in an EMME study regarding medical image diagnosis in which radiologists vs. medical residents seemed to benefit more from having watched EMME (Gegenfurtner et al., 2017). Thus, perhaps the students in the present study did not have the necessary cognitive prerequisites to use the guidance in the EMME to their full potential. For future research it would therefore be interesting to test an older sample, and/or to take measures of cognitive abilities into account when investigating EMME.

Besides including broader measures of cognitive ability, the addition of eye tracking would also be informative for future research. Because the present study was conducted in the classroom, we were not able to measure the students' eye movements while they were watching the video examples. Although many studies regarding EMME found that the student's visual attention allocation was affected by EMME in terms of fixating relevant information more often, faster and longer (Jarodzka et al., 2013; Van Marlen et al., 2016, 2018; Mason et al., 2015; Scheiter et al., 2018) in the current study we cannot know for sure whether the EMME affected the visual attention allocation of the students. It is possible that they attempted to engage in solving the problem shown in the example themselves, without following the model's gaze, in which case the attentional guidance provided by EMME would not be very useful.

To conclude, EMME may not be effective for learning to solve deductive reasoning problems, regardless of whether or not the gaze guidance is combined with verbal explanations. However, before being able to draw a definitive conclusion on the usefulness of EMME for acquisition of deductive reasoning strategies, further research is needed that includes students with higher cognitive prerequisites, and measures their eye movements during example study.

ACKNOWLEDGMENTS

This research was funded by a Vidi grant (#452-11-006) from the Netherlands Organization for Scientific Research (NWO) awarded to Tamara van Gog. The authors would like to thank Susan Ravensbergen for assistance with data collection. Lastly, the authors thank all the staff members and students of Krimpenerwaard College for participating.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12712.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OSF at (link blinded for review).

ORCID

Tim van Marlen https://orcid.org/0000-0002-0894-989X

Margot van Wermeskerken https://orcid.org/0000-0003-3883-9875

Halszka Jarodzka https://orcid.org/0000-0003-2312-4703

REFERENCES

Ayres, P., & Paas, F. (2007). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology*, *21*, 811–820. https://doi.org/10.1002/acp.1351

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*, 1465–1477. https://doi.org/10.1016/j.cognition.2007.05.012

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

Fiorella, L., Van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, *109*, 653–665. https://doi.org/10.1037/edu0000161

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*, 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education*, *113*, 212–225. https://doi.org/10.1016/j.compedu.2017.06.001

Gierasimczuk, N., van der Maas, H. L. J., & Raijmakers, M. E. J. (2013). An analytic tableaux model for deductive mastermind empirically tested with a massively used online learning system. *Journal of Logic, Language and Information*, *22*, 297–314. https://doi.org/10.1007/s10849-013-9177-5

Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, *40*(5), 813–827. https://doi.org/10.1007/s11251-012-9218-5

Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, *25*, 62–70. https://doi.org/10.1016/j.learninstruc.2012.11.004

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 247–262). Cambridge University Press.

Litchfield, D., Ball, L., Donovan, T., Manning, D., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied*, *16*, 251–262. https://doi.org/10.1037/a0020082

Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, *41*, 172–187. https://doi.org/10.1016/j.cedpsych.2015.01.004

Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*, *47*, 1083–1095. https://doi.org/10.1111/bjet.12271

Mason, L., Scheiter, K., & Tornatora, C. (2017). Using eye movements to model the sequence of textpicture processing for multimedia comprehension. *Journal of Computer Assisted Learning*, *33*, 443–460. https://doi.org/10.1111/jcal.12191

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd, rev. ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

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

Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics*, *36*, 289–300. https://doi.org/10.1016/j.ergon.2005.11.008

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

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

Salmerón, L., & Llorens, A. (2018). Instruction of digital reading strategies based on eye-movements modeling examples. *Journal of Educational Computing Research. Advance online publication.*, *57*, 343–359. https://doi.org/10.1177/0735633117751605

Scheiter, K., Schubert, C., & Schüler, A. (2018). Self-regulated learning from illustrated text: Eye movement modelling to support use and regulation of cognitive processes during learning from multimedia. *British Journal of Educational Technology*, *88*, 80–94. https://doi.org/10.1111/bjep.12175

Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. In *Proceedings of the 6th international conference on multimodal interfaces* (pp. 9–15). Penn State University, State College. https://doi.org/10.1145/1027933.1027936

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, *25*, 785–791. https://doi.org/10.1016/j.chb.2009.02.007

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*, 155–174. https://doi.org/10.1007/s10648-010-9134-7

Van Gog, T., Rummel, N., & Renkl, A. (2019). Learning how to solve problems by studying examples. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (pp. 183–208). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108235631.009

Van Marlen, T., Van Wermeskerken, M. M., Jarodzka, H., & Van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior*, *65*, 448–459. https://doi.org/10.1016/j.chb.2016.08.041

Van Marlen, T., Van Wermeskerken, M. 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*, 273–284. https://doi.org/10.1016/j.learninstruc.2018.07.005

Van Wermeskerken, M. 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

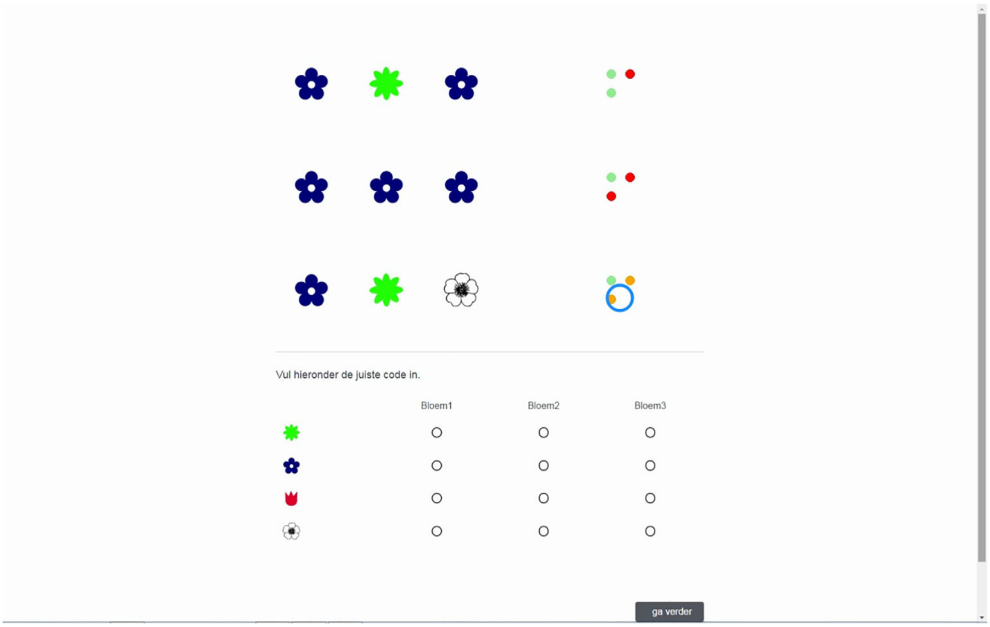
Vitak, S. A., Ingram, J.E., Duchowski, A.T., Ellis, S., & Gramopadhye, A.K. (2012). Gaze-augmented think-aloud as an aid to learning. *Proceedings of ACM CHI 2012 conference on human factors in computing systems*, 2991–3000, Austin, Texas, 2012. ACM Press.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*, 35–37. https://doi.org/10.3758/s13423-017-1343-3

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, *71*, 79–89. https://doi.org/10.1016/j.chb.2017.01.049

How to cite this article: van Marlen, T., van Wermeskerken, M., Jarodzka, H., Raijmakers, M., & van Gog, T. (2022). Looking through Sherlock's eyes: Effects of eye movement modelling examples with and without verbal explanations on deductive reasoning. *Journal of Computer Assisted Learning*, *38*(5), 1497–1506. https://doi.org/10.1111/jcal.12712

APPENDIX A: A SCREENSHOT OF THE EMME WITH THE TRANSLATED (FROM DUTCH) TRANSCRIPT OF THE VERBAL EXPLANATIONS



Here we see the Mastermind task and as you can see the code consists of three flowers. You see that three attempts have already been made and to the right of every attempt you also see the feedback. Let's start to crack this code. First thing that stands out is the third attempt. Here you see that the feedback consists of one green dot and two orange dots. The green dot means that one flower is in the correct location and the orange dots mean that there are two flowers that are in the code but are not yet placed in the correct location. Thus, this third attempt indicates which flowers must be present in the code. The second thing that stands out when looking at the first and second attempt is that only one flower changed, namely the middle green flower in the first attempt changes into a blue flower in the second attempt. Subsequently, when we look at the feedback, then you see that first there were two green dots and at the second attempt there is only one green dot. Thus, from two correct flowers we went to one correct flower. So, we can deduce that the middle flower must be a green flower. Now that we know this flower, we can figure out the rest with the third attempt. That is, there is one flower in the correct location and two are not yet placed in the correct location. But now we know that the middle flower is green, so this is the correct flower. This means that the blue flower and the white flower are not yet correctly placed. Thus, these have to be switched. So the first flower of the code should not be the blue flower but should be the white flower and the third flower of the code should not be the white flower but should be the blue flower. So finally, the code should be white flower, green flower, blue flower.