How generative drawing affects the learning process: An eye-tracking analysis

Johannes Hellenbrand1 | Richard E. Mayer2 | Maria Opfermann3 | Annett Schmeck4 | Detlev Leutner1

1 Department of Instructional Psychology, University of Duisburg-Essen, Essen, Germany
2 Department of Psychological and Brain Sciences, University of California, Santa Barbara
3 Department of Educational Science, Ruhr-University Bochum, Bochum, Germany
4 Stiftung Mercator Essen, Essen, Germany

Correspondence
Johannes Hellenbrand, Department of Instructional Psychology, Faculty of Educational Sciences, University of Duisburg-Essen, Universitätsstr. 2, 45141 Essen, Germany.
Email: johannes.hellenbrand@uni-due.de

Funding Information
Interdisciplinary Centre of Educational Research, University of Duisburg-Essen

Summary
Generative drawing is a learning strategy in which students draw illustrations while reading a text to depict the content of the lesson. In two experiments, students were asked to generate drawings as they read a scientific text or read the same text on influenza with author-provided illustrations (Experiment 1) or to generate drawings or write verbal summaries as they read (Experiment 2). An examination of students' eye movements during learning showed that students who engaged in generative drawing displayed more rereadings of words, higher proportion of fixations on the important words, higher rate of transitions between words and workspace, and higher proportion of transitions between important words and workspace than students given a text lesson with author-generated illustrations (Experiment 1) or students who were asked to write a summary (Experiment 2). These findings contribute new evidence to guide theories for explaining how generative drawing affects learning processes.

KEYWORDS
eye tracking, generative drawing, generative learning activities, learning processes, multimedia learning

1 INTRODUCTION

Generative drawing is a learning strategy in which students draw illustrations while reading a text to depict the content of the lesson. There is substantial evidence that students perform better on tests of learning outcome (such as retention and transfer tests) when they are asked to draw their own illustrations as they read a scientific passage rather than simply read (Fiorella & Mayer, 2015; Fiorella & Zhang, 2018; Leutner & Schmeck, 2014). The goal of the present study is to better understand the learning processes (as measured by eye-tracking methodology) that students engage in when they learn by drawing.

By asking students to create drawings while reading a scientific text, they are encouraged to engage actively in generative processing such as selecting key elements and relations, organizing them into descriptive and depictive representations, and integrating the mental representations with each other and with prior knowledge into a coherent mental model. The present study uses eye-tracking technology to identify the degree to which students engage in generative learning processes while learning by drawing.

In a recent review of research on learning by drawing, Fiorella and Zhang (2018) note that "evidence for drawing is often based on comparisons with weak control conditions, such as students who only read the text without provided illustrations," (p. 1115) and they call for research that compares learning by drawing with text-based generative learning strategies such as summarizing or with including instructor-provided illustrations. We take up this call in the present study.

Reviews of learning by drawing (Fiorella & Mayer, 2015; Fiorella & Zhang, 2018; Leutner & Schmeck, 2014) also show that evidence for drawing generally focuses on learning outcome measures such as performance on retention tests or transfer tests. In the present study,
we shift the focus to learning process measures based on eye-tracking methodology (Holmqvist et al., 2011; Mayer, 2010), in order to shed light on the learning process primed by drawing as compared with other study aids.

2 WHAT IS THE THEORETICAL FRAMEWORK BEHIND GENERATIVE DRAWING?

According to the cognitive theory of multimedia learning (CTML; Mayer, 2009, 2014), generative cognitive processing includes three key cognitive processes: selecting, which refers to attending to the important material in an instructional message; organizing, which refers to arranging incoming information into a coherent cognitive structure; and integrating, which refers to mentally connecting verbal and pictorial representations with each other and with relevant knowledge activated from long-term memory. Learning by drawing is intended to prime these cognitive processes during learning.

In applying Mayer’s theoretical accounts of multimedia learning to drawing, van Meter and Garner (2005) developed the generative theory of drawing construction (GTDC). The GTDC defines drawings as intentionally learner-generated pictorial representations intended to depict represented objects accurately that are described in a text. The GTDC states that when learners are asked to generate external visual representations that represent the main aspects of an instructional text, they are no longer passive consumers of information but actively involved in cognitive processes of selection, organization, and integration. The cognitive model of drawing construction (CMDC; van Meter & Firetto, 2013) is a revised version of the initial GTDC, which retains its basic principles, “namely, the selection and organization of descriptive elements and the forced integration of the descriptive and depictive representations” (van Meter & Firetto, 2013, pp. 255–256), but interprets these processes in terms of self-regulated learning processes.

The CMDC shares structural similarities with Mayer’s (2009, 2014) CTML but is shaped by the theoretical principles described in Schnotz’s (2005, 2014; see also Schnotz & Bannert, 2003) integrated model of text and picture comprehension, in which there is a distinction between descriptive and depictive representations. According to the CMDC, learners form a surface representation by selecting relevant words from the instructional material. The organization of these words and their relations leads to a propositional representation, and by integrating the semantic information with visuospatial information and prior knowledge from long-term memory, a depictive coherent mental model is derived. The integration itself “is forced as the verbal representation is the foundation for the nonverbal representation” (van Meter & Garner, 2005, p. 318). The mental model has then to be translated into a perceptual image, which is a depictive surface feature representation that learners can externalize as a drawing on paper. Prior knowledge has a crucial impact on the propositional representation and on mental model construction—for example, when an object to be drawn is described in the text as concave shaped, students have to consult their memory in order to determine how the word “concave” can be translated into visual form.

However, generative drawing is not a linear sequence: Learners will undergo many recursions through the steps of the CMDC, driving them back and forth between various internal and external representations in order to generate a drawing (van Meter & Firetto, 2013). To specify these recursive processes in the light of metacognitive awareness, the CMDC incorporates and adapts principles of Winne and Hadwin’s (1998, Winne & Perry, 2000) model of self-regulated learning. Thus, the metacognitive processes of generative drawing are described in the CMDC as a three-phased self-regulation cycle, which involves “(1) setting standards for performance, (2) applying (strategic) operations, and (3) monitoring goal progress” (van Meter & Firetto, 2013, p. 257).

In order to create a drawing, learners undergo the first phase by setting performance standards and deciding how many details they need to include and how to express relations among different parts. Therefore, the CMDC predicts that by using the drawing strategy, learners’ attention is directed towards key elements and their relations in the text (i.e., the process of selecting). The second phase refers to the cognitive processes of selecting and organizing in order to establish the propositional representation. The CMDC predicts that drawing prompts the usage of other known learning strategies (e.g., self-questioning and rereading) that support the processes of selecting and organizing and therefore facilitate the construction of a propositional network. According to the CMDC, the third phase occurs when learners monitor their progress by comparing their in-progress drawing to the standards set in the first phase, which triggers metacognitive control. If learners are unable to externalize the drawing or have problems representing certain elements and their specific spatial relations in the drawing, the process of integrating pictorial and verbal representations with each other has failed and forward movement is blocked. When this happens, learners know that the learning material might not been understood well enough. In order to fill the missing gaps and to complete the drawing, learners are forced to switch back to the text, reread and reselect relevant words, and make new meaningful connections between their work-in-progress drawing and corresponding text passages (forcing function of generative drawing; van Meter & Firetto, 2013). In line with the integrated model of text and picture comprehension, the CMDC assumes that the mental model receives now input from both the perceptual image and the propositional representation. As a consequence of this forcing function, the CMDC predicts that learners who generate drawings on their own engage in these self-monitoring and self-regulation processes more frequently than learners who do not use this strategy.

3 WHAT IS THE EMPIRICAL FRAMEWORK FOR GENERATIVE DRAWING?

Over the last 30 years, research on the effectiveness of drawing as a generative learning strategy has shown strong evidence that students
who engage in generative drawing learn more deeply from a scientific
text than students who, for example, learn only with the text (for an
overview, see Fiorella & Mayer, 2015; Leutner & Schmeck, 2014). In
particular, research shows that generative drawing is more likely
to develop its full potential when the drawing process itself is sup-
ported, for instance, with a legend showing all relevant elements
for drawing (cf., Schmeck, Mayer, Opfermann, Pfeiffer, & Leutner,
2014; Schwamborn, Mayer, Thillmann, Leopold, & Leutner, 2010).
Instructional support can reduce extraneous cognitive load that the
mechanics of drawing itself induce. Under this boundary condition,
Schwamborn et al. (2010; see also Leutner & Schmeck, 2014) pro-
posed the generative drawing principle, that is, “people learn better
from a science text when they are asked to draw illustrations
representing the main ideas of the text” (p. 878).

To clarify that the learning improvement is actually due to learners’
active engagement in drawing activities during reading and not due to
the mere application of the multimedia principle, which states that
“people learn better from words and pictures than from words alone”
(Mayer, 2009, p. 223), Schmeck et al. (2014) conducted an experiment,
in which a drawing group was not only compared with a reading only
control group but also compared with a group who received author-
generated pictures to the text. The drawing group performed signifi-
cantly better on a comprehension posttest than the other groups,
which did not differ significantly from each other. Based on the higher
learning outcomes of the drawing group, the authors concluded that
the instruction to draw pictures during reading fosters the engage-
ment in generative activities and is more effective than simply provid-
ing pictures to a text.

Another crucial factor for the success of the drawing strategy is
how accurately learners generate their drawings. Van Meter and
Garner (2005) defined drawing accuracy as “the degree to which
completed drawings resemble the represented object(s)” (p. 299).
Thus, drawings should be representational pictures (Alessandrini,
1984; Carney & Levin, 2002), rather than artistic expressions. Partic-
ularly suitable for creating representational drawings are texts that
convey how a scientific system works (Fiorella & Mayer, 2015). Several studies measured
the quality of learners’ drawings with regard to learning outcomes
(e.g., Greene, 1989; Hall, Bailey, & Tillman, 1997; Leopold & Leutner,
2015; Lesgold, de Good, & Levin, 1977; Lesgold, Levin, Shimron, &
Guttmann, 1975; Schmeck et al., 2014; Schmidgall, 2017;
Schwamborn et al., 2010; van Meter, 2001; van Meter, Aleksic,
Schwartz, & Garner, 2006). Results show that drawing accuracy is
positively related to learning outcomes: The higher the drawing accu-
arcy, the higher the score on a learning outcome posttest. Based on
these results, Schwamborn et al. (2010; see also Leutner & Schmeck,
2014) proposed the prognostic drawing principle, that is, “the quality of
learners’ drawings during learning predicts the quality of their learning
outcomes” (p. 878).

Thus, there is strong support for the generative drawing principle,
but there is no research so far on how generative drawing affects spe-
cific cognitive processes during learning such as focusing on and
selecting relevant information and integrating it. With the steady
developments in the field of eye-tracking research, it is now possible
to shed some light on these open questions.

4 | COGNITIVE PROCESSING AND EYE
MOVEMENTS

Eye-tracking methodology has proven to have an impact on multime-
dia research in recent years. Because eye-tracking can reveal insights
into ongoing cognitive processes and visual attention during learning,
it offers unique contributions to better understand multimedia learn-
ing (Hyönä, 2010; Mayer, 2010; van Gog & Scheiter, 2010). Using
eye-tracking as a research tool requires the interpretation of the rela-
tion between eye movements and cognitive processing. The eye-mind
assumption claims a direct relationship between fixations and informa-
tion being processed: The longer the fixation duration, the more
extensive the ongoing cognitive processing (Just & Carpenter, 1976;
1980; Rayner, 1998). Furthermore, there is also evidence that a shift
of attention precedes a saccade to the corresponding position and
that transitions are associated with integration processes (Deubel &
Schneider, 1996; Hoffman & Subramanian, 1995; Kragten, Admiraal,
& Rijaardsdam, 2015; Mason, Pluchino, Tornatora, & Ariasi, 2013;
Rayner, 1998; Rayner, McConkie, & Ehrlich, 1978; Schonweke,
Berthold, & Renkl, 2009). Although results of several studies support
the hypothesis of a linkage between eye fixation behavior and cogni-
tive processing (e.g., Hannus & Hyönä, 1999; Hegarty & Just, 1993;
Hegarty, Mayer, & Green, 1992; Kragten et al., 2015; She & Chen,
2009; Underwood, Humphrey, & Foulsham, 2008), eye-tracking data
should always be interpreted with caution: When a learner’s eyes
move from a picture to a text, for example, the transition could indi-
cate an attempt to make a connection between the picture and corre-
sponding words, but we do not know if the integration process was
actually successful.

However, in recent years, eye-tracking methodology has been
used in multimedia research more frequently to examine cognitive
processing. Johnson and Mayer (2012), for example, conducted an
eye-movement analysis of the spatial contiguity effect in multimedia
learning. They analyzed transitions between text and diagram (which
they refer to as the cognitive process of integrating), and total fixation
time on diagram and text (which they refer to as the cognitive process
of selecting) as a more direct measure of online cognitive processing
in order to clarify theoretical assumptions of the CTML. Ponce and
Mayer (2014) used eye-tracking methodology to examine how graphic
organizers affect cognitive processing during learning by measuring
fixation durations (indicating attempts to select elements) and transi-
tions (indicating attempts to organize and integrate the graphic orga-
nizer with the text).

Mason et al. (2013) conducted an eye-tracking study to examine
how students learn from a science text with concrete or abstract illus-
trations. Eye-tracking methodology was used to trace text and picture
processing. Among other variables, they used fixation duration to
measure processing during reading and the frequency of transitions
from verbal to graphic representations as indicators of integrative
effort. Based on the eye-movement analysis, they concluded that students who learned with text and abstract illustrations processed the text more efficiently (indicated by fewer fixations on the text, though both groups had almost the same learning gains) and made a greater effort of integrating verbal and pictorial information (indicated by a higher proportion of transitions between the abstract illustration and the text). Schnotz et al. (2014) examined the strategies students used for integrating text and picture information. Based on the analysis of participants’ number of fixations, fixation durations, and transitions between texts, pictures, and comprehension items, they concluded that texts are more likely to be used for coherence-oriented general processing, whereas pictures are more likely to be used as scaffolds for initial mental model construction and afterwards as easily accessible visual representations on demand for specific mental model updates.

Kragten et al. (2015) examined students’ learning activities while studying biological process diagrams. As an indication for learning activities, they collected verbal data and eye-tracking data, namely, fixation durations and transitions between areas of interest (AOIs) in the process diagrams. Both fixation duration and transitions were significant predictors of comprehension scores.

All listed exemplary studies used a computer monitor to present the learning material and a stationary monitor-bound eye tracker to record participants’ eye movements. In the present experiments, we extend eye-tracking methodology to the study of generative drawing on paper while reading a scientific text on paper using mobile eye-tracking glasses (see Figure 1).

5 OVERVIEW OF THE EXPERIMENTS

Asking learners to create drawings depicting the content of a science text they are reading has been shown to be a promising learning strategy in promoting better learning outcomes (Fiorella & Mayer, 2015; van Meter & Garner, 2005), but there is a need to more directly investigate the cognitive processes underlying learning. According to van Meter and Firetto’s (2013) CMDC, two major processes primed by generative drawing are (a) focusing more on the main aspects of the text and (b) mentally integrating related portions of the text rather than processing the text in linear order.

In the present studies, we use eye-tracking methodology to examine these learning processes, in addition to examining students’ learning outcomes using posttest measures. In particular, we are interested in whether the eye-movement patterns underlying generative drawing can be distinguished from eye-movement patterns of students using other generative instructional or learning strategies. Finally, we examine whether there is further support for the prognostic drawing principle—the idea that the quality of learner-generated drawings is related to their performance on tests of learning outcomes.

For the two experiments, we adapted learning materials about an influenza infection and the immune system’s response as well as pretests and posttests from Schmeck et al. (2014). Schmeck et al. (2014) have already shown that a drawing group performed significantly better on learning outcome measures than a reading only control group (Experiments 1 and 2) and a group provided with instructor-generated pictures (Experiment 2), so we decided to forego a reading only group in the present experiments.

The primary goal of Experiment 1 is to compare the learning processes and outcomes when students draw their own illustrations as they read a scientific passage versus when students read a scientific passage that contains author-provided illustrations. The primary goal of Experiment 2 is to compare the learning processes and outcome when students draw their own illustrations as they read a scientific passage versus when students write a summary as they read a scientific passage. The main research questions in both experiments are the following:

1. Do the groups differ in learning outcomes (as measured by posttests)?
2. For the drawing group in each experiment, does the accuracy of the drawing correlate with posttest scores?
3. Do the groups differ in learning processes (as measured by eye-tracking metrics as indicators of cognitive processing)?
4. Are the learning processes and learning outcomes of the drawing group equivalent across the two experiments?

According to the CMDC, we predict that the drawing group will exhibit more generative processing during learning in both experiments, but we are not able to make firm predictions concerning learning outcome measures. On the basis of previous findings, we expect a significant correlation between the accuracy of drawing and posttest scores for the second research question, which would indicate a

FIGURE 1 Example of a learner sitting in front of the drafting table and engaging in generative drawing while wearing eye-tracking glasses
replication of the prognostic drawing principle. We expect that the drawing group's performance on learning outcome measures and the eye-movement patterns will be consistent across the two experiments.

6 | EXPERIMENT 1

The main goal of Experiment 1 was to use eye-tracking methodology to shed more light on the cognitive processing of students who learn by drawing as they read a text about influenza in comparison with those who read the same text with author-generated pictures. Students were instructed either to draw pictures for each paragraph of a text or to read the text along with author-generated pictures for each paragraph. First, if hands-on drawing activity is especially effective, we would expect students in the drawing group to perform better than students in the picture group on learning outcome posttests measuring retention, transfer, and drawing. If students naturally make connections between words and graphics, then there should be no difference or the picture group should perform better because they have better illustrations. Second, the quality of students' drawings should also predict learning outcomes. Third, according to the theoretical assumptions of the CMDC, we expected students in the drawing group to reread parts of the text several times, to focus more on the important content of the text, and to switch from their drawing to corresponding text passages more often than students in the picture group.

6.1 | Method

6.1.1 | Participants and design

The participants were 62 eighth and ninth graders in German higher track secondary schools. The data for seven participants were excluded from analyses due to eye-tracking system calibration problems or due to measurement errors, and three were removed because they did not follow the instructions. For the analyses that follow, there were 52 participants (19 in eighth grade and 33 in ninth grade) with an average age of 14.08 years (SD = 0.79). The proportion of females was 44.2%. There were two groups based on a between-subjects design, with 26 students randomly assigned to the drawing group (who were instructed to draw pictures on paper that reflect the main elements and relations described in the text) and 26 students randomly assigned to the picture group (who received author-generated pictures on paper along with the text and thus served as the control group).

6.1.2 | Materials and apparatus

The materials were administered in paper-and-pencil form and consisted of a demographics questionnaire, a content-related knowledge pretest, a spatial and verbal ability test, two learning booklets, a usability questionnaire, and three posttests. The demographics questionnaire solicited basic background information concerning the students' age, sex, and latest grades in science subjects (e.g., biology). The content-related knowledge pretest (based on Schmeck et al., 2014) consisted of 19 multiple-choice items and was intended to assess students' prior knowledge of information covered in the text (Cronbach's $\alpha = .76$). An example item is: “What are the components of an influenza-virus? (a) capsule, membrane, and antibody, (b) capsule, membrane, and glycoproteins, (c) antibody, membrane, and glycoproteins, or (d) nucleus, capsule, and membrane” [b] is the correct answer]. The content-related knowledge pretest consisted of the same items as the retention posttest.

To measure students' spatial ability, the 10-item Paper Folding Test developed by Ekstrom, French, and Harman (1976) was used (Cronbach's $\alpha = .75$). Students' verbal ability (Cronbach's $\alpha = .62$) was measured with 30 multiple-choice items from the German Cognitive Ability Test (KFT; Heller & Perleth, 2000). Each item presents words that have something in common (stem). Of five alternatives, students have to select the one word that fits to the stem. An example item is: “Mouse, wolf, bear: (a) rose, (b) lion, (c) running, (d) hungry, (e) brewing” [b] is the correct answer].

There were two versions of the learning booklet—one for each group. They consisted of a motivation questionnaire (which was identical for both groups) and the drawing version or the picture version of the learning material. After the students read the instructions for their particular learning material, but before they engaged in their tasks, students had to fill in a questionnaire addressing their current motivation for doing the learning task (Cronbach's $\alpha = .84$). The questionnaire consisted of nine items from the Challenge and Interest subscale of the Questionnaire on Current Motivation (FAM; Rheinberg, Vollmeyer, & Burns, 2001). The learning material for both groups consisted of a science text that explained the causal steps of an influenza infection and the immune system's response (adapted from Schmeck et al., 2014), which is part of the German biology curriculum for ninth graders in higher track secondary schools. The text consisted of 816 words (in German) and was divided into seven paragraphs, which were presented on seven different pages. The second paragraph, translated into English, is shown in Figure 2. For both groups, each paragraph was presented on the left side of a 29.7 × 42.0 cm (DIN A3) sheet of paper. The right side of the paper contained a drawing prompt for the drawing version or contained an author-generated picture for the picture version. The drawing prompt, adapted from Schmeck et al. (2014), included a legend showing all eight elements that were relevant to be included in various drawings (e.g., a template for an influenza virus) and a partly predrawn background on which the drawing had to be generated. This level of baseline instructional support was used based on research by Schwamborn et al. (2010), who proposed using a scaffolded drawing prompt to counter overstraining students with the mechanics of drawing itself (see also Leutner, Leopold, & Sumfleth, 2009). Providing a drawing prompt that includes a legend showing all relevant elements for drawing and a predrawn background is intended to leave enough cognitive capacity for learners to make sense of the text and benefit from the drawing strategy. The author-generated pictures consisted of the same elements as used in the legend for the drawing condition, and each element was labeled.
All elements that were shown in the legend or were used for the author-generated pictures as well as their relations were described in detail in the text.

The usability questionnaire developed by the first author assessed whether the eye-tracking glasses affected participants’ performance (Cronbach’s $\alpha = .75$). Students were asked to rate on a scale from (1) does not apply to (7) does apply whether the eye-tracking glasses made them move unnaturally, made them feel unwell, and disturbed or restricted them during learning with the text and graphics.

The three posttests intended to assess learning outcomes were a retention posttest, a transfer posttest, and a drawing posttest (based on Schmeck et al., 2014). The retention posttest (Cronbach’s $\alpha = .85$) consisted of 19 multiple-choice items (the same items as in the content-related knowledge pretest) and measured students’ retention of the factual and conceptual information covered in the text. In addition, the transfer posttest consisted of nine multiple-choice items (Cronbach’s $\alpha = .73$) addressing the ability to apply the learned information to new situations. An example item is: "T-helper cells recognize not only viruses, but also agents that are extraneous to the body. Which medication would you administer to a patient, who has received a new kidney? (a) a medicine that suppresses the immune response to the body, (b) a medicine that activates the immune response of the body, (c) a medicine that contains antigens, or (d) a medicine that contains blood of the kidney donor? [(a) is the correct answer]. The drawing posttest consisted of four items asking students to reproduce the main ideas in the text by drawing sketches depicting key concepts of the text and their spatial relations. An item example for the drawing posttest is: "Sketch the binding between influenza viruses and appropriate antibodies." The four drawing posttest items are condensed versions of the seven drawings that the students in the drawing group were asked to generate during the learning phase.

The apparatus was an SensoMotoric Instruments (SMI) mobile eye-tracking glasses system (SMI ETG 1.5) with a sampling rate of 30 Hz, which was used to record the participants’ eye movements during learning. One advantage of having participants wear mobile eye-tracking glasses (in contrast to sitting at a screen monitored by a stationary eye tracker) is that students could be in their natural sitting position and move their heads freely around while learning. SMI’s BeGaze software (version 3.7) was used for fixation and saccade detection. We also used a drafting table in order to help students handle the big paper format. Another advantage of the drafting table is that the angle between the learning material and participants’ eyes could be optimized in order to reduce measurement errors (Holmqvist et al., 2011).

6.1.3 | Procedure

The study took place in two separate sessions. In the first session, students were tested together in their classrooms. First, the experimenter handed out the demographics questionnaire and the content-related knowledge pretest and collected them when the students were finished. Second, the experimenter provided instructions for the Paper Folding Test and the verbal ability test and handed each student both tests. The students had a 3-min time limit for the Paper Folding Test and a 10-min time limit for the verbal ability test.

After about a week, individual learning sessions were held in an unoccupied classroom during school hours. The students were individually taken out of class and randomly assigned to one of the two learning conditions. First, the experimenter explained the eye-tracking
equipment, asked the participant to put on the eye-tracking glasses, and proceeded with a 3-point eye-tracking calibration for the participant. Second, the student received written and spoken instructions on how to proceed based on his or her treatment group. Third, the student was given the learning booklet based on the condition he or she was assigned to. Each student was provided with a pen, a pencil, an eraser, and several highlighters with the spoken instruction to use whatever they needed. After these instructions, the student was asked to complete the motivation questionnaire. Next, the participant started learning with the material corresponding to the given treatment while the eye movements were recorded. The reason why participants had to wear the eye-tracking glasses before they started learning with the text was, on the one hand, to give students time to get used to the eye-tracking glasses and, on the other hand, to allow the experimenter to supervise the eye-tracking accuracy and to recalibrate if necessary.

Both groups were instructed to carefully read the text paragraph by paragraph in order to comprehend the biological topic. Students in the picture group were instructed to read the text and refer to the pictures on the right side of the booklet if they needed support. Students in the drawing group were instructed to read the text and to make a clear and simple drawing for each of the seven paragraphs representing its main ideas. They were also advised to use the pictorial templates given in the legend and to draw on the predrawn backgrounds. Both groups were told that after learning, they would be asked questions about what was learned. Learning time was computed by awarding one point for each correct answer, yielding a total possible of 9 points. Similar to the drawing accuracy score during learning, the score for the drawing posttest was calculated by counting the total number of correct main ideas in each student’s answer across the four drawing items based on an adapted and revised coding scheme from Schmeck et al. (2014). Students could earn a maximum of 20 points. Again, the quality was scored by the first author and a biology teacher with an intraclass correlation of ICC unjust = .99.

A maximum of 19 points was possible for the content-related knowledge pretest and for the retention posttest, 1 point for each correctly answered item. The transfer posttest contained nine multiple-choice items, and students scored 1 point for each correct answer, yielding a total possible of 9 points. Similar to the drawing accuracy score during learning, the score for the drawing posttest was calculated by counting the total number of correct main ideas in each student’s answer across the four drawing items based on an adapted and revised coding scheme from Schmeck et al. (2014). Students could earn a maximum of 20 points. Again, the quality was scored by the first author and a biology teacher with an intraclass correlation of ICC unjust = .97.

The Paper Folding Test was scored by adding up the number of correctly answered items and subtracting 1 point for each incorrectly answered item, with no points for unanswered items. A maximum of 10 points was possible. The verbal ability test score for each student was computed by awarding one point for each correct answer and by adding up the points to obtain the total verbal ability score (out of a possible total of 30 points). The median was calculated for the motivation and the eye-tracking usability questionnaire based on the seven-level Likert scale used in each questionnaire.

6.1.4 Scoring of the paper-based materials

The drawing accuracy score (concerning drawing during learning in the drawing condition) was computed by using a revised coding scheme adapted from Schmeck et al. (2014), which was based on expert drawings and a checklist specifying important relational features of the drawings. Students received points when they identified critical elements representing the main ideas of each text paragraph and constructed pictures by drawing each of those elements correctly and spatially correct on the predrawn background. One important aspect in the second paragraph (as shown in Figure 2) is, for example, that there are several new influenza viruses located outside the somatic cell after the replication process. Students received 2 points if their drawing included at least two correctly sketched influenza viruses (including capsule, membrane, and glycoproteins) outside the somatic cell. Points were subtracted if the influenza viruses were indeed recognizable, but missed parts, and for drawing fewer than two viruses. Students had to draw seven pictures in total, one picture for each text paragraph. Maximum drawing accuracy was scored with 32.5 points. The first author and a biology teacher scored the quality for each of the seven drawings for each student with an intraclass correlation of ICC unjust = .99.

To analyze eye-movement data, we established AOs. An AOI is a defined region on a stimulus, in which eye movements (such as fixations and saccades) are analyzed separately (Holmqvist et al., 2011). Using BeGaze, version 3.7, we created two main AOs for each of the seven pages: One covered the author-generated picture (for the picture group) or the drawing prompt (for the drawing group) on the right side of the page (workspace AOI) and the other one covered the whole text paragraph on the left side of the page (text AOI), as shown in Figure 3. In addition, within each text AOI, we defined further AOs covering the most important content of each paragraph (important-text AOI). The determination of the most important content of each paragraph was done in collaboration with a biology teacher. The size and coverage of all AOs were identical for both experimental conditions.

There are many possible eye-tracking measures (Holmqvist et al., 2011), but we focus on four indices as indicators of cognitive processing, as shown in Table 1:

1. Number of fixations per word inside the text AOs (which we name rereadings) is computed as the total number of fixations in the text AOs divided by the total number of words in the text AOs and provides a measure of intentional focus on words. The
purpose of the measure is to quantify how much of the text has been read and how carefully. If the entire text is read once, a value about 0.8 is expected, because due to chunking, not all words are fixated. If participants read the text or parts of the text several times, the value is expected to be higher (Holmqvist et al., 2011; Poole, Ball, & Phillips, 2005).

2. Proportion of relevant fixation time (which we name focused fixation time) is the sum of all fixation durations on important-text AOIs divided by the sum of all fixation durations on text AOIs and provides a measure of selective reading attention.

3. The number of times eye fixations move from the workspace AOIs to the text AOIs per minute (which we name transition rate) provides a measure of the attempts to integrate visualizations—either self-generated or instructor provided—and words (e.g., Schmidt-Weigand, Kohnert, & Glowalla, 2010).

4. Proportion of relevant transitions (which we name meaningful transitions) is computed as the total number of transitions from the workspace AOIs to the important-text AOIs divided by the total number of transitions from the workspace AOIs to the text AOIs and provides a measure of the proportion of meaningful attempts to integrate visualizations with corresponding text passages (e.g., Johnson & Mayer, 2012).

These indices were chosen because they have been used in previous studies to tap the learning processes that we propose occur in generative learning.
6.2 | Results and discussion

6.2.1 | Are the groups equivalent on basic characteristics?

The first step was to determine whether the groups differed on basic characteristics. A chi-square analysis indicated that the drawing and picture groups did not differ significantly in the proportion of boys and girls, $\chi^2(1) = 0.08, p = .780$. Independent t tests show that there were no significant group differences in spatial ability, $t(50) = -1.09, p = .281$, verbal ability, $t(50) = -1.49, p = .142$, and age, $t(50) = 0.35, p = .729$. The drawing group ($M = 40.89\%$ correct, $SD = 17.44$) and the picture group ($M = 48.12\%$ correct, $SD = 23.43$) also did not differ in the content-related knowledge pretest, $t(50) = -1.27, p = .209$. As the data for verbal ability and age are not normally distributed, non-parametric tests were used additionally, which also confirmed that there were no significant differences between the groups on the median rating scores at the .05 level.

We conclude that the groups do not differ on basic characteristics.

6.2.2 | Learning outcomes: Do students learn better from a science text when they generate drawings than when they are given author-generated pictures?

Our basic research question concerns whether students learn better from a science text when they are asked to create drawings as they read as compared with being given a text with author-generated illustrations. If the drawing group engages in deeper learning, we predicted that the students in the drawing group will perform better on the posttests. The top three rows of Table 2 show the adjusted means and standard deviations of the drawing group and the picture group on the retention posttest, transfer posttest, and drawing posttest, respectively. An analysis of covariance (ANCOVA) with group as a between-subjects factor, retention posttest score as the dependent measure, and pretest score as the covariate shows that the covariate was significantly related to the retention posttest score, $F(1,49) = 20.06, p < .001, \eta^2_p = .291, r = .54$, and that the drawing group scored significantly higher than the picture group on the retention posttest, yielding a medium effect size, $F(1,49) = 4.94, p = .031, \eta^2_p = .092, r = .30$. An ANCOVA with group as the between-subjects factor, transfer posttest score as the dependent measure, and pretest score as the covariate shows that the covariate was significantly related to the transfer posttest score, $F(1,49) = 18.21, p < .001, \eta^2_p = .271, r = .52$, and that the drawing group scored significantly higher than the picture group on the transfer posttest, yielding a large effect size, $F(1,49) = 17.95, p < .001, \eta^2_p = .268, r = .52$. None of the analyses violated the assumption of homogeneity of regression slopes.

It should be noted, however, that students in the drawing group invested more time on the learning task compared with students in the picture group, $t(50) = 7.925, p < .001, r = .75$ (for $M$ and $SD$ see the last row of Table 2). Furthermore, across both groups time on task predicted, when controlling for pretest scores, the retention posttest ($\beta = .349, p = .006$) and the drawing posttest ($\beta = .481, p < .001$), but not the transfer posttest ($\beta = .095, p = .458$), as learning outcome measures. Thus, the question arises whether the effect of drawing instruction on learning outcome measures might be mediated by time on task. Following Baron and Kenny (1986), ANCOVAs with group as the between-subjects factor, learning outcome measures as dependent variables, and time-on-task scores as well as pretest scores as covariates indicated that the drawing effect was no longer significant for the retention posttest, $F(1, 49) < 1, \eta^2_p < .01, r = .03$, and the drawing posttest, $F(1, 48) = 3.803, p = .057, \eta^2_p = .07, r = .27$, when controlling for time on task. This indicates that the drawing effect on the retention and the drawing posttest was mediated by time on task: Drawing instruction increased the time invested for learning, and this increased time increased students' scores on the retention and the drawing posttests. Note that such a mediation analysis was not performed for the transfer posttest because there

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>Drawing</th>
<th>Mdn</th>
<th>M</th>
<th>SD</th>
<th>Picture</th>
<th>Mdn</th>
<th>M</th>
<th>SD</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention test* (% correct)</td>
<td>Drawing</td>
<td>79.26</td>
<td>19.26</td>
<td>67.30</td>
<td>19.26</td>
<td>.30*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer test* (% correct)</td>
<td>Drawing</td>
<td>67.79</td>
<td>22.70</td>
<td>69.39</td>
<td>22.70</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing test* (% correct)</td>
<td>Drawing</td>
<td>75.46</td>
<td>19.26</td>
<td>52.61</td>
<td>19.26</td>
<td>.52***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Learning processes |       |         |     |   |    |         |     |   |    |   |
| Rereadingsb | Drawing | 2.93 | 3.12 | 0.95 | 1.91 | 2.12 | 1.32 | .60*** |
| Focused fixation time (%) | Drawing | 59.69 | 3.75 | 54.53 | 3.99 | .56*** |
| Transition rateb | Drawing | 2.46 | 2.72 | 1.18 | 1.23 | 1.36 | 0.79 | .59*** |
| Meaningful transitions (%) | Drawing | 59.33 | 7.59 | 45.61 | 22.4 | .39** |

| Learning time |       |         |     |   |    |         |     |   |    |   |
| Time on learning task (min) | Drawing | 22.18 | 6.13 | 9.65 | 5.23 | .75*** |

*aAdjusted means and standard deviations.

*bA Mann-Whitney test was used for the analysis. *$p \leq .05$. **$p \leq .01$. ***$p \leq .001$. 

\(\eta^2_p = .208, r = .46\), and that the drawing group scored significantly higher than the picture group on the drawing posttest, yielding a large effect size, $F(1,49) = 17.95, p < .001, \eta^2_p = .268, r = .52$. None of the analyses violated the assumption of homogeneity of regression slopes.
was neither a correlation with time on task nor a drawing effect on the transfer test.

To sum up, results provide some evidence concerning the cognitive processing of students who engage in generative drawing. In the present study, both groups received treatments intended to improve learning, but the drawing group outperformed the picture group on two of three measures of learning outcomes, suggesting that asking students to generate drawings is more effective in improving learning outcomes than providing already drawn illustrations. However, the drawing group also needed more time to learn, and the drawing effect turned out to be mediated by the time students invested in their learning task. Thus, the question arises as to what kind of cognitive processing was initiated by the instruction to generate drawings.

### 6.2.3 | Learning outcomes: Does the quality of learner-generated drawings predict learning outcomes?

A further research question concerns whether the quality of drawings produced by students in the drawing group is related to performance in tests of learning outcomes. The mean proportion correct on drawing accuracy during learning was 67.93% (SD = 21.00). A correlation analysis revealed that the drawing accuracy score of learner-generated drawings correlated significantly with the retention posttest score, \( r = .82, p < .001 \), with the transfer posttest score, \( r = .58, p = .002 \), and with the drawing posttest score, \( r = .93, p < .001 \). These findings provide further evidence for the prognostic drawing principle (Leutner & Schmeck, 2014; Schwamborn et al., 2010), which states that the quality of students' drawings during learning from science texts is predictive of learning outcomes.

### 6.2.4 | Learning processes: Do students exhibit different eye-movement patterns as indicators of cognitive processing during learning when they generate drawings than when they are given author-generated pictures?

Having established that the drawing group performs better on learning outcome posttests than the picture group, the major new focus of this experiment, compared with previous experiments, was to examine how drawing affects cognitive processing during learning, indicators of which we examine through eye-tracking measures.

The bottom portion of Table 2 shows the medians, means, and standard deviations of the drawing group and the picture group on eye-tracking measures of rereading, focused fixation time, transition rate, and meaningful transitions. Our major prediction was that the drawing group will score significantly higher than the picture group on each of these indices of cognitive processing during learning. As the data for rereadings and transition rate are not normally distributed, nonparametric tests were used; for focused fixation time and meaningful transitions, independent \( t \) tests were used.

A Mann–Whitney test on rereadings shows that students who engaged in generative drawing read parts of the text significantly more often than students provided with author-generated pictures, \( U = 101.50, z = -4.33, p < .001, r = .60 \), indicating that the drawing group has read parts of the text more carefully than the picture group. An independent \( t \) test on focused fixation time revealed that the drawing group scored significantly higher than the picture group, \( t(50) = 4.81, p < .001, r = .56 \), indicating that the drawing group was more focused on key elements while reading than the picture group.

Furthermore, a Mann–Whitney test on transition rate revealed that the drawing group exhibited significantly more workspace-to-text transitions per minute than the picture group, \( U = 104.00, z = -4.28, p < .001, r = .59 \), indicating more attempts to engage in integrating visualizations and corresponding text passages. In addition, an independent \( t \) test on meaningful transitions shows that the drawing group had a higher proportion of meaningful transitions than the picture group, \( t(50) = 2.96, p = .006, r = .39 \), indicating a higher probability of success in integrating visualizations and corresponding text passages. These differences in indicators of cognitive processing during learning represent the primary new findings in this experiment.

Experiment 1 was designed to provide information on cognitive processing during generative drawing by analyzing eye-movement patterns as indicators of cognitive processing. In particular, the eye-movement analysis shows that when learners are instructed to engage in generative drawing, (a) they had more fixations per word inside the text AOIs, indicating more attentional focus on the text; (b) they had a higher proportion of fixation time on the most important content of the text while reading, indicating more success in selecting relevant information; (c) they had more transitions per minute, indicating more attempts in making connections between visualizations and text; and (d) they had a higher proportion of meaningful transitions, indicating a higher probability of success in making meaningful connections between visualizations and corresponding text passages than learners who received a different instructional strategy (such as providing graphics). However, differences in eye-tracking measures did not translate into differences in transfer test performance, suggesting that deep learning may be possible even without corresponding increases in eye-movement indices.

### 7 | EXPERIMENT 2

Experiment 2 addressed several issues. First, we were interested in whether the eye-tracking patterns of the drawing group can be distinguished from eye-tracking patterns of another generative learning strategy such as summarizing. Like drawing, summarizing is intended to foster appropriate cognitive processing during learning, which should be beneficial for learning outcomes; but unlike drawing, summarizing relies mainly on verbal processing whereas drawing includes visual processing (Fiorella & Mayer, 2015; van Meter & Fretto, 2013). Thus, although both learning strategies are intended to promote better learning outcomes, they may accomplish that goal by fostering somewhat different cognitive processes during learning. Following Leopold and Leutner (2012), we predicted that drawing would be more helpful than summarizing to help students to create...
a deep level understanding of the text’s content. Thus, we expected students in the drawing group to focus more on the important content of the text while reading and to have a higher proportion of meaningful transitions than the summarizing group. Second, we wanted to examine whether the performance on learning outcome measures and the eye-tracking patterns obtained in Experiment 1 can be replicated for the drawing group in Experiment 2, as similar to Experiment 1, the quality of students’ drawings should also predict learning outcomes.

7.1 | Method

7.1.1 | Participants and design

The participants were 43 ninth graders in German higher track secondary schools. The data for seven participants were excluded from analyses due to eye-tracking system calibration problems or due to measurement errors, and one was removed because the student did not follow the instructions. For the analyses that follow, there were 35 participants with an average age of 14.09 years (SD = 0.51). The proportion of females was 71.4%. There were two groups based on a between-subjects design, with 17 students randomly assigned to the drawing group (who were instructed to draw pictures on paper that reflect the main elements and relations described in the text) and 18 students randomly assigned to the summarizing group (who were instructed to write summaries on paper that reflect the main ideas of the text and thus served as the control group).

7.1.2 | Materials and apparatus

The materials were identical to those used in Experiment 1, except for the necessary adjustments to the summarizing group. The pretest consisted of a demographics questionnaire, a content-related knowledge pretest (Cronbach’s α = .29), a spatial ability test (Cronbach’s α = .55), and a verbal ability test (Cronbach’s α = .53).

Again, there were two versions of the learning booklet, depending on the task. They consisted of a motivation questionnaire (identical for both groups, Cronbach’s α = .61) and the drawing version or the summarizing version of the learning material. To ensure the comparability of the two groups, the summarizing version also included baseline instructional support (as suggested by Schmeck et al., 2014). The right side of the paper contained a legend showing all the key words that were used in the drawing prompt for the drawing group and therefore gave a hint on what is important in the text. Furthermore, there were lines for the summaries (as shown in Figure 4) beneath the legend. The drawing prompt for the drawing group was identical to that in Experiment 1.

As in Experiment 1, a usability questionnaire was intended to assess whether the eye-tracking glasses affected participants’ performance (Cronbach’s α = .83) and three posttests intended to assess retention (Cronbach’s α = .76), learner’s ability to transfer what was presented in the text to new situations (Cronbach’s α = .56), and student’s retention of conceptual information presented in the science text by means of drawing.

The same eye-tracking glasses and software were used as in Experiment 1.

FIGURE 4 Second text paragraph of the immunology lesson for both groups: Example for text, two-part drawing prompt (left), and two-part summarizing prompt (right). Translated from the German original
7.1.3 | Procedure

The procedure was the same as in Experiment 1, except that the summarizing group was asked to produce a written summary of each paragraph. As in Experiment 1, the students were randomly assigned to one of the two learning conditions. Both groups were instructed to produce drawings or summaries, respectively, that are short, simple, and clear, reflect only the main content of each paragraph, and use the elements given in the legends.

7.1.4 | Scoring of the paper-based materials

All paper-based instruments were scored with the same procedures used in Experiment 1. The first author and a student assistant (teacher trainee) scored the quality for each of the seven drawings for each student with an intraclass correlation of ICC_{unjust} = .97 and each of the four drawing items of the drawing test for each student with an intraclass correlation of ICC_{unjust} = .91.

Similar to the drawing accuracy score, the summary accuracy score was computed by using a revised coding scheme adapted from Schütz (2015). The summary accuracy score scheme is designed to be comparable with the drawing accuracy score. Students received points when they identified critical elements representing the main ideas of each text paragraph and wrote them down in their summaries. One important aspect in the lesson is, for example, that there are several new influenza viruses located outside of the somatic cell after the replication process. Students received a maximum of 2 points for this aspect if they mentioned in their summaries that there were several new influenza viruses created (1 point) and that they were located outside the somatic cell (1 point). Again, the first author and a student assistant (teacher trainee) scored the quality for each of the seven summaries for each student with an intraclass correlation of ICC_{unjust} = .97. Maximum summary accuracy score was 32 points. Both accuracy scores are specified in percentage to be comparable.

7.1.5 | Measurement of eye-movement behavior

Identical AOIs were used as in Experiment 1 (as shown in Figure 5), created with BeGaze, version 3.7. Again, the main eye-tracking metrics were rereadings, focused fixation time, transition rate, and meaningful transitions.

7.2 | Results and discussion

7.2.1 | Are the groups equivalent on basic characteristics?

Before testing the hypotheses, we analyzed whether groups differed on basic characteristics. A chi-square analysis indicated that the drawing and summarizing groups did not differ significantly in the proportion of boys and girls, \( \chi^2(1) = 0.01, p = .915 \). Independent t tests show that there were no significant group differences in spatial ability, \( t(33) = -0.57, p = .572 \), verbal ability, \( t(33) = -0.43, p = .671 \), and age, \( t(33) = -0.30, p = .765 \). The drawing group (\( M = 34.67\% \) correct, \( SD = 12.35 \)) and the summarizing group (\( M = 35.67\% \) correct, \( SD = 11.78 \)) also did not differ in the content-related knowledge pretest, \( t(33) = -0.25, p = .808 \), but because of the low reliability in this small sample (Cronbach’s \( \alpha = .29 \)), the content-related knowledge...
pretest will not be considered in the following analyses. As the data for spatial ability and age are not normally distributed, nonparametric tests were used additionally, which also confirmed that there were no significant differences between the groups at the .05 level. Similarly, as the motivation questionnaire and usability questionnaire were measured with an ordinarily scaled Likert scale, nonparametric tests were used to investigate group differences, which also confirmed that there were no significant differences between the groups on the median rating scales at the .05 level. We conclude that the groups do not differ on basic characteristics.

7.2.2 Learning outcomes: Does the performance of the drawing group in Experiment 1 correspond to the performance of the drawing group in Experiment 2?

A main goal in Experiment 2 was to determine whether the drawing group in Experiment 2 produced a drawing accuracy and learning outcome scores as in Experiment 1. Independent t tests show that there were no significant differences between the two drawing groups on the drawing accuracy, $t(41) = -0.51, p = .605$, the retention posttest $t(41) = 0.96, p = .342$, the transfer posttest, $t(41) = 1.07, p = .289$, and the drawing posttest, $t(41) = -0.72, p = .474$. We conclude that the drawing accuracy and learning outcomes of the drawing group in Experiment 2 were equivalent to the drawing group in Experiment 1.

7.2.3 Learning outcomes: Do students learn better from a science text when they generate drawings than when they generate summaries?

A secondary goal was to determine whether the drawing group and the summarizing group in Experiment 2 differed in learning outcome performance. The top three rows of Table 3 show the means and standard deviations of the drawing group and the summarizing group on the retention posttest, transfer posttest, and drawing posttest, respectively. Independent t tests show that there was no significant effect of group on the retention posttest, $t(33) = -1.03, p = .314$, and the transfer posttest, $t(33) = -0.46, p = .647$. However, as expected, the drawing group scored significantly higher than the summarizing group on the drawing test, yielding a large effect size, $t(33) = 4.63, p < .001$, $r = .63$. Overall, we conclude that the groups achieved equivalent learning outcomes in terms of verbal content from the lessons, that is, they achieved equivalent levels of learning of the text material.

In contrast to Experiment 1, where students in the drawing group invested more time on the learning task than students in the picture group, students in the drawing group in Experiment 2 invested less time on the learning task compared with students in the summary group, $t(33) = 6.016, p < .001, r = .72$ (for M and SD, see the last row of Table 3). Furthermore, across both groups, time on task negatively predicted the drawing posttest ($\beta = -.434, p = .009$), but not at all the retention posttest ($\beta = .138, p = .429$) and transfer posttest ($\beta = .106, p = .433$) as learning outcome measures. Thus, only for the drawing posttest, the question arises whether the negative effect of drawing instruction on these learning outcome measures might have been mediated by time on task. Following Baron and Kenny (1986), an ANCOVA with group as the between-subjects factor, drawing posttest scores as dependent variables, and time-on-task scores as covariates indicated that the drawing effect was still significant, $F(1, 32) = 10.897, p = .002, \eta^2 = .25$, $r = .50$, when controlling for time on task. This indicates that the drawing effect on the drawing posttest was not mediated by time on task.

7.2.4 Learning outcomes: Do the quality of learner-generated drawings and summaries predict learning outcomes?

As in Experiment 1, we tested whether the quality of students’ drawings predicts learning outcomes and therefore provides further support for the prognostic drawing principle. The mean proportion correct on drawing accuracy during learning was 68.24% (SD = 15.91). A correlation analysis revealed that the drawing accuracy score of learner-generated drawings correlated significantly with retention posttest score, $r = .87, p < .001$, and drawing test score, $r = .89, p < .001$. There was no significant correlation, however, with the transfer test score, $r = .29, p = .266$. The lack of significance for transfer may reflect the small sample size of 17 participants in the drawing group. These findings are partly consistent with the results of Experiment 1 and provide further evidence for the prognostic drawing principle (Leutner & Schmeck, 2014; Schwamborn et al., 2010), which states that the quality of students’ drawings during learning from science texts is predictive of learning outcomes.

The summary accuracy score did not correlate significantly with retention posttest score, $r = -.14, p = .576$, transfer test score, $r = -.09, p = .731$, or drawing test score, $r = -.26, p = .290$, after all. The data provide no evidence for the summary accuracy score to be a predictor for learning outcomes.

### Table 3: Means and standard deviations of two groups on three learning outcome measures and four learning process measures for Experiment 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Drawing</th>
<th>Summarizing</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention test (% correct)</td>
<td>71.83</td>
<td>78.07</td>
<td>.19</td>
</tr>
<tr>
<td>Transfer test (% correct)</td>
<td>57.52</td>
<td>61.11</td>
<td>.08</td>
</tr>
<tr>
<td>Drawing test (% correct)</td>
<td>78.09</td>
<td>50.72</td>
<td>.63***</td>
</tr>
<tr>
<td>Learning processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rereadings</td>
<td>3.11</td>
<td>5.12</td>
<td>.57***</td>
</tr>
<tr>
<td>Focused fixation time (%)</td>
<td>61.95</td>
<td>58.41</td>
<td>.40*</td>
</tr>
<tr>
<td>Transition rate</td>
<td>2.89</td>
<td>5.65</td>
<td>.76*.69***</td>
</tr>
<tr>
<td>Meaningful transitions (%)</td>
<td>65.06</td>
<td>58.48</td>
<td>.38*</td>
</tr>
<tr>
<td>Learning time</td>
<td>21.78</td>
<td>9.79</td>
<td>.72***</td>
</tr>
<tr>
<td>Time on learning task (min)</td>
<td>4.11</td>
<td>37.22</td>
<td></td>
</tr>
</tbody>
</table>

*p ≤ .05, **p ≤ .01, ***p ≤ .001.
Learning processes: Does the performance of the drawing group in Experiment 1 correspond to the performance of the drawing group in Experiment 2?

A main goal of Experiment 2 was to determine whether the eye-tracking patterns of the drawing group in Experiment 1 could be replicated and compared against another learning generative strategy, that is, summarizing. The bottom four rows of Table 3, showing scores on the four eye-tracking measures for the drawing group in Experiment 2, are partly comparable with the bottom four rows of Table 2 for the drawing group in Experiment 1. The drawing group in Experiment 1 and the drawing group in Experiment 2 yielded eye-tracking scores that did not differ significantly on rereadings, \( t(41) = 0.04, p = .967 \), focused fixation time, \( t(41) = -1.80, p = .079 \), and transition rate, \( t(41) = -0.47, p = .638 \). The groups only differ significantly in the proportion of meaningful transitions, \( t(41) = -2.13, p = .039 \), \( r = .32 \). The higher proportion of meaningful transitions in Experiment 2 could be due to the fact that in Experiment 2 only ninth grade students participated, whereas in Experiment 1, the drawing group consisted of both eighth and ninth grade students. Ninth graders are cognitively further developed and may have more sophisticated cognitive schemas than eighth graders.

We conclude that the eye-movement patterns as indicators of cognitive processes of the drawing group in Experiment 2 are almost equivalent to those of the drawing group in Experiment 1, indicating a replication of the pattern of eye-tracking measures.

Learning processes: Do students exhibit different eye-movement patterns as indicators of cognitive processing during learning when they generate drawings than when they generate summaries?

A major research question for Experiment 2 concerns whether the drawing group and summarizing group differ in their indicators of cognitive processing during learning, as reflected in the four eye-tracking measures. The bottom portion of Table 3 shows the means and standard deviations of the drawing group and summarizing group on each of the four eye-tracking indices. On the one hand, an independent \( t \) test on rereadings revealed that the drawing group read parts of the text significantly less often than the summarizing group, \( t(33) = -5.12, p < .001 \), \( r = .57 \), indicating that the summarizing group has read parts of the text more carefully than the drawing group. On the other hand, an independent \( t \) test on focused fixation time revealed that the drawing group scored significantly higher than the summarizing group, \( t(33) = 2.48, p = .018 \), \( r = .40 \), indicating that the drawing group was more focused on key elements while reading than the summarizing group. Concerning the attempts to engage in the process of integrating, an independent \( t \) test on transition rate shows that the drawing group exhibited significantly less workspace-to-text transitions per minute than the summarizing group, \( t(33) = -5.54, p < .001 \), \( r = .69 \). However, the drawing group scored significantly higher than the summarizing group on meaningful transitions, \( t(33) = 2.35, p = .025 \), \( r = .38 \), indicating that students who drew engaged more deeply in making meaningful connections between relevant text passages and respective visualization aspects than students who summarized and thus were supposed to make these connections between relevant text and the notes that they extracted from this text.

Overall, the summarizing group displayed more attentional focus on the text (as indicated by rereadings) and more attempts in making connections between the generated summaries and the text (as indicated by transition rate), whereas the drawing group displayed more strategically focused processing of the text by focusing more attention on relevant text passages (as indicated by focused fixation time) and connections between relevant text passages and generated drawings (as indicated by meaningful transitions). Thereby, as expected, the summarizing group also engaged in self-monitoring and self-regulation processes but had lower proportions of focused fixation time and meaningful transitions than the drawing group. The higher frequency of transitions and rereadings in the summarizing group could be due to the difficult technical terms (such as “glycoproteins” and “pseudopodia”) that students in the summarizing group had to copy from the text to their summaries. In contrast to the summarizing group, however, the eye-movement patterns of the drawing group provide further evidence for the assumed self-regulation cycle, which forces learners’ attention more on relevant words and connections (van Meter & Firetto, 2013).

We conclude that the eye-movement patterns of the students who engaged in generative drawing can be distinguished from the eye-movement patterns of students who engaged in summarizing. However, additional research is needed to determine why differences in learning process measures were not reflected in differences in learning outcome measures.

8 | GENERAL DISCUSSION

8.1 | Empirical contributions

The present set of experiments enables us to better understand how generative drawing works as a learning strategy, especially in contrast to other learning strategies. The primary empirical contribution of Experiment 1 is that students who are asked to draw illustrations as they read a scientific text display different learning processes (as measured by four eye-tracking metrics) and, to some extent, different learning outcomes (as measured by two of three posttests) than students who read the same text with author-generated pictures. Concerning learning processes, the drawing group had more fixations per words inside the text AOIs, a higher proportion of focused fixation time in relevant text, more workspace-to-text transitions per minute, and a higher proportion of meaningful transitions from visualizations to relevant text than the picture group. Concerning learning outcomes, the drawing group outperformed the picture group on the retention and drawing tests, but not the transfer test, which is intended to measure deep learning. In a recent review, Fiorella and Zhang (2018) reported a similar pattern for learning outcome measures, in which the drawing group scored higher than the picture group on retention
tests ($d = 0.45$ based on six comparisons) but not on transfer tests ($d = 0.04$ based on 11 comparisons). The new contribution of Experiment 1 is that the drawing group’s change in cognitive processing during learning as compared with the picture group (i.e., indicated by eye-tracking measures) did not translate into increased deep understanding of the material as compared with the picture group (i.e., indicated by transfer test performance).

The primary empirical contribution of Experiment 2 is that students who are asked to draw as they read display different learning processes (as measured by four eye-tracking metrics) than students who are asked to write summaries but display equivalent learning outcomes as students who are asked to summarize as they read (as measured by retention and transfer posttests). Concerning learning process, the drawing group had fewer fixations per words inside the text AOIs and fewer workspace-to-text transitions per minute, but a higher proportion of focused fixation time in relevant text and a higher proportion of meaningful transitions from the generated content to corresponding text passages than the summarizing group. Concerning learning outcome, the groups did not differ significantly on the key posttest measures of retention and transfer, with the drawing group scoring lower on both measures. In a recent review, Fiorella and Zhang (2018) reported similar learning outcome results in which drawing produced slightly lower scores on retention ($d = −0.09$ based on two comparisons) and transfer ($d = −0.05$ based on three comparisons) as compared with other generative strategies such as summarizing. In short, the drawing group’s superiority in cognitive processing (as measured by eye movements) as compared with the summarizing group did not translate into improvements in learning outcome on retention or transfer tests.

Furthermore, in both experiments, the accuracy of the drawings that students made correlated with posttest scores, consistent with prior results (Schwamborn et al., 2010). On the basis of these findings, we concur with the prognostic drawing principle: “The quality of learners’ drawings during learning predicts the quality of their learning outcomes” (Schwamborn et al., 2010, p. 878; see also Leutner & Schmeck, 2014). Finally, the drawing group’s drawing accuracy, performance on posttests, and, partly, eye-tracking measures do not differ significantly between Experiments 1 and 2, providing a useful replication of performance by the drawing group.

8.2 Theoretical contributions

The present set of experiments contributes to evaluating the theoretical assumptions of the Cognitive Model of Drawing Construction (van Meter & Firetto, 2013) by using mobile eye-tracking technology. The strength of eye-tracking methodology is its ability to measure indicators of cognitive processing during learning, in contrast to posttests that reflect learning outcomes. As van Meter and Firetto (2013) explain in their CMD model, “metacognitive control is responsible for guiding and affecting the operations a learner carries out” (p. 260). By creating a drawing, learners receive intuitive feedback on their comprehension of the text: If some portion of the drawing cannot be externalized, learners know that the learning material might not been understood well enough and that they need to seek necessary clarification in order to finish the drawing task. Our findings are consistent with these assumptions. In both experiments, learners with drawing instruction engaged in regulation processes to revise their mental model by switching back to the text and rereading parts of it in order to select the proper information and integrate it with their work-in-progress drawing. Importantly, the results of both experiments show that learners who engage in generative drawing during reading a scientific text are more likely to direct their attention towards key elements of the text and are more likely to make meaningful connections between their generated drawing and corresponding text passages when compared against a group that received a different instructional strategy (such as providing graphics in Experiment 1) or a group that was prompted to use a different learning strategy (such as summarizing in Experiment 2). Thus, this work helps move beyond demonstrating a generative drawing effect based on comparing a drawing group to a control group; instead, it examines the underlying cognitive processes involved in generative drawing as compared with other generative learning activities (e.g., summarizing) or incorporating generative instructional design features (e.g., adding illustrations).

8.3 Practical contributions

This study points to the potential of drawing as a generative learning strategy that primes deep cognitive processing (i.e., paying attention to the relevant material and making connections between visualization and corresponding text passages) during learning. It should be noted that students were provided with drawing tools during drawing, so the generative drawing principle may work best when students have support in drawing (Fiorella & Mayer, 2015). However, this study also shows that other generative study aids—such as summarizing—can be just as effective as drawing for promoting transfer test performance.

8.4 Limitations and future directions

We were not able to examine a large sample, largely due to the logistical requirements of eye-tracking methodology to test participants individually. Results should be interpreted in light of this limitation, with particular caution in interpreting null effects. Eye-tracking methodology provides valuable task-relevant information, but the data should be interpreted with caution. Although the eye-mind assumption claims a direct relationship between fixations and the way information is being processed, the eye-tracking data do not guarantee that students actually comprehend the relevant information while fixating it (Hyönä, 2010) or tell us what they are thinking while performing their tasks (Holsanova, Holmberg, & Holmqvist, 2008; Triesch, Ballard, Hayhoe, & Sullivan, 2003).

There was no classic (reading only) control group in either experiment, but rather the drawing group was compared with a group that received a different instructional strategy or was prompted
to use a different learning strategy. This is because our intention was to go beyond a further replication of the generative drawing effect (e.g., Leopold, Sunmfleth, & Leutner, 2013; Schmeck et al., 2014; Schwamborn et al., 2010, for an overview, see Fiorella & Mayer, 2015).

In both experiments, the transfer posttest failed to discriminate between the groups. Future studies on generative drawing, however, might investigate whether there are conditions under which generative drawing improves performance on transfer better than other generative learning strategies. Additionally, claims concerning the effectiveness of learning by drawing require a control group that does not draw. There is evidence that the benefits of generative drawing are limited to local knowledge, which is specific to the information presented in the instructional material and that is tied to the structure of the material (van Meter & Firetto, 2013; Waters, van Meter, Perrotti, Drogo, & Cyr, 2005, 2011). On the other hand, some studies reported positive effects of generative drawing on transfer performance (e.g., Gobert & Clement, 1999; Leopold & Leutner, 2012; Schwamborn et al., 2010; for an overview, see Fiorella & Mayer, 2015). Van Meter and Firetto (2013) predict in the CMDC, that when the drawing strategy is used, knowledge-based performance will be equally bound to the specific characteristics of the constructed mental model. They conclude that "improved performance can only be expected when the posttest assessments are well-matched to the characteristics of the specific knowledge representation that is constructed" (p. 263). However, particular attention should be paid to the transfer posttest to be used, because transfer posttest needs to be well-matched to the characteristics of the constructed mental model in order to be a sensitive measure.

In Experiment 2, Cronbach’s alpha was rather low for some of the instruments. Concerning the content-knowledge pretest, the low Cronbach’s alpha could indicate that the students had no prior knowledge at all on the immunology subject and only blind-guessed the answers. Therefore, we decided to exclude prior knowledge from further analyses. Furthermore, in Experiment 2, both verbal and spatial ability tests were only used to ensure that the groups were equivalent on basic characteristics. Thus, future research with more participants is needed to investigate the effects of generative drawing in comparison with summarizing on learning outcomes.

Finally, in Experiment 2, we did not find an advantage of the drawing instruction over the summarizing instruction in terms of learning outcome, as was expected and found by Leopold and Leutner (2012) at least for their transfer test ($d = 1.07, p = .002$). The main difference between the two studies is that in Leopold and Leutner (2012), as opposed to the present experiment, there were no scaffolded prompts, neither in the drawing condition (in form of a legend and a predrawn background) nor in the summarizing condition (in form of a legend). The eye-tracking data indicate that in our experiment, the summarizing group showed significantly fewer transitions to the toolbar and overall fixation durations on the toolbar than the drawing group. Thus, further research is needed to investigate the impact of the scaffolded prompts. In addition, summarizing procedures are taught in German schools and German textbooks, so our students might have been more familiar with the summarizing strategy than the drawing strategy.

Overall, this set of experiments provides evidence concerning the underlying cognitive processes during learning, which generally are assumed solely on a theoretical basis. However, understanding the connection between learning processes and learning outcomes remains a central challenge for the study of learning strategies.

ACKNOWLEDGMENT

This research was supported by the Interdisciplinary Centre of Educational Research, University of Duisburg-Essen.

CONFLICT OF INTEREST

Neither me nor any coauthor have a significant financial arrangement or affiliation with any product or services used or discussed in my paper, nor any potential bias against another product or service.

ORCID

Johannes Hellenbrand  https://orcid.org/0000-0002-8705-6626
Richard E. Mayer  https://orcid.org/0000-0003-4055-6938

REFERENCES


Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. Applied Cognitive Psychology, 23, 1227–1243. https://doi.org/10.1002/acp.1526


