Fostering integration of informational texts and virtual labs during inquiry-based learning

Joep van der Graaf\textsuperscript{a,b,*}, Eliane Segers\textsuperscript{a,b}, Ton de Jong\textsuperscript{a}

\textsuperscript{a} Department of Instructional Technology, University of Twente, the Netherlands
\textsuperscript{b} Behavioural Science Institute, Radboud University, the Netherlands

ARTICLE INFO

Keywords:
Inquiry-based learning
Eye-tracking
Integration
Signaling
Informational text

ABSTRACT

Inquiry-based learning allows students to learn about scientific phenomena in an exploratory way. Inquiry-based learning can take place in online environments in which students read informational texts and experiment in virtual labs. We conducted two studies using eye-tracking to examine the integration of these two sources of information for students from vocational education (78 and 71 participants, respectively, mean age of 13 years and 7 months). In Study 1, we examined whether the amount of time spent on reading text and on integrating the text content with information from a virtual lab (as measured via gaze switches between the text and the lab) affected the quality of the inquiry-based learning process in the lab (i.e., correctly designed experiments and testable hypotheses created) and the learning gain (increase in domain knowledge from pretest to posttest). Results showed, on average, a gain in domain knowledge. Pretest scores were related to posttest scores, and this relation was mediated by the score for correctly designed experiments in the lab. There was no relation between informational text reading time and inquiry process quality or learning gain, but more frequent integration was associated with a higher score for experimentation in the virtual lab, and more frequent integration attenuated the relation between pretest score and designing correct experiments. Integration could thus compensate for the negative effects of lower prior knowledge. In Study 2, we examined whether integration was stimulated by highlighting correspondences between the informational text and the virtual lab (i.e., signaling). Integration was higher than in Study 1, but this did not further improve the quality of the inquiry process or the learning gain. A general conclusion is that integration fosters inquiry-based learning, but that stimulating additional integration may not result in further improvement.

1. Introduction

Practices within STEM (Science, Technology, Education, and Mathematics) classrooms are changing. There is an increasing focus on engaged forms of learning and acquiring deep knowledge (see De Jong, 2019), often achieved via increased use of student-centered methods and with the help of technology (OECD, 2014, 2016). The goal of such STEM teaching is to prepare students for jobs in the 21st century, in which the ability to understand and use scientific methods and concepts to gain conceptual knowledge is increasingly important (Fischer et al., 2014). Inquiry-based learning is often used in this context (Lazonder & Harmsen, 2016), and can be defined as learning by engaging in the authentic scientific research activities of hypothesis generation, setting up and conducting experiments, and drawing conclusions (Chinn & Malhotra, 2002). Overall, learning by inquiry is most effective when students are well guided. This can be done by different actors (e.g., teachers, online learning environments, or peers) and in different ways (e.g., by combining inquiry with direct instruction, as in the form of informational text, or by providing students with scaffolds or apps; see De Jong, 2019). When proper guidance is provided, inquiry-based learning results in deeper knowledge acquisition than direct instruction alone (Aditomo & Klieme, 2020; Lazonder & Harmsen, 2016).

When inquiry-based learning takes place online, the virtual lab is often accompanied by informational text. Reading (and understanding) this text is important, because if students lack the necessary prior domain knowledge, the inquiry process suffers (Eberbach & Crowley, 2009). Indeed, for inquiry-based learning to be effective in improving domain knowledge, students need to know how to conduct scientific
investigations, what the relevant concepts are within the domain, and how to use these concepts in their investigations (Klahr, Zimmerman, & Jirout, 2011). Informational texts that provide instruction and support in digital learning environments help students to acquire such knowledge (Graesser, McNamara, & VanLehn, 2005). However, little is known about how the use of informational text in this way is related to carrying out inquiry processes, and how both are related to prior knowledge and learning from inquiry. In the present study, therefore, we examined the association of reading informational text and integration of the informational text with a virtual lab with inquiry-based learning in a pair of studies using eye-tracking.

1.1. Functions of multiple representations

The text and the virtual lab together form what is called a multiple representations situation. Multiple representations seem to play a large role in the experimentation phase of inquiry-based learning, as compared to hypothesis generation and drawing conclusions (Chinn & Malhotra, 2002), because the representation that is pivotal in this phase is the virtual lab in which experiments can be conducted. Three main functions of multiple representations, complementing, constraining, and constructing (Ainsworth, 2006), explain how multiple representations might affect inquiry-based learning. Among these functions, ‘constructing’ is most closely related to integration (Mayer, 2014), which might be especially relevant for inquiry-based learning in a digital learning environment with informational text and a virtual lab. For example, studies in multimedia learning contexts have used signaling to promote integration and learning (Richter, Scheiter, & Eitel, 2016).

1.1.1. Complementing

Multiple representations can complement each other in the processes they support or the information they provide. While both informational texts and virtual labs provide information about the domain, the processes they support differ. Reading about the domain happens with the informational texts; conducting investigations happens with the virtual lab. This means that students can learn via reading the informational text and/or conducting experiments in the lab. In inquiry-based learning, it is common that these two processes complement each other. For example, theory is provided in informational texts in the orientation phase before experimentation starts (Pedaste et al., 2015), and information that is generated during experimentation in a virtual lab may not be present in the text. Indeed, additional actions by learners in a virtual lab are necessary in order to generate additional knowledge, as shown by the finding that the possession of good experimental skills is associated with more adequate and consistent domain knowledge after an inquiry-based lesson series (Edelsbrunner, Schalk, Schumacher, & Stern, 2015).

1.1.2. Constraining

It may be complicated for a learner to relate knowledge that is inferred from the text to the virtual lab, as the latter has inherent limitations (constraints) on what can be tested and how (see Samarapungavan, 2018). Virtual representations tend to be more specific than textual representations. Pictorial representations, as an adjunct to the text, have been found to have beneficial effects for learning about science; one of their useful functions is to provide a structural framework for the text (Carney & Levin, 2002). While such studies have used a static pictorial representation, this organizational function might be provided by a virtual lab, as well.

1.1.3. Constructing

Finally, construction includes the integration of information from multiple representations (Ainsworth, 2006). When informational text and a virtual lab are provided, the student must combine and integrate the information from these two forms of media for meaningful learning (Mayer, 2014). Such integration is important, as shown by a study by Mason, Tornatora, and Pluchino (2013), who found that children engaging in more frequent integration (operationalized in that study as switching between text and illustration) when studying an illustrated science text had a higher gain in text-based factual knowledge.

1.2. Integration

Different types of information can be integrated during inquiry-based learning. Informational texts can include information, for example, about the specific domain (i.e., domain knowledge), about how knowledge is created (i.e., epistemic knowledge), or about how to execute specific steps (i.e., procedural knowledge). In line with Eberbach and Crowley’s study (2009), previous studies have shown that domain knowledge helps in inquiry-based learning; for example, it helps with understanding what are the relevant mechanisms or processes that explain the observed phenomena, how variables are operationalized, how data can be generated, what counts as evidence, and so forth (see Samarapungavan, 2018, for a review). To foster inquiry-based learning, Klahr et al. (2011) suggested three strategies. The first is to provide information about the domain. Indeed, domain knowledge in the text can be directly linked to representations in the virtual lab (e.g., via the organizational role of the pictorial representation; see Carney & Levin, 2002), which could help with understanding the concepts that are under investigation. The second is to provide information on how to conduct experiments procedurally. This may help with knowing what to do in the lab, which helps with designing and conducting experiments (e.g., Millar, Lubben, Gott, & Duggan, 1995). Finally, information about how to use domain knowledge in experiments can foster the translation of research questions into an experimental design, and also the translation of experimental results into conceptual frameworks. In other words, knowing how to use domain knowledge in experiments can foster coordination of theory and evidence (Kuhn, 2004).

To summarize, offering both textual and pictorial representations has several beneficial functions, among which integration (part of the constructing function; Ainsworth, 2006) seems to be the most relevant for online inquiry-based learning, as it links understanding of the domain to understanding (and use) of the lab. While integration has been addressed in science learning, that research has mostly considered static pictures (Mayer, 2014). Simulations have also been addressed; for example, transitions between a graph of the results and an animated presentation of the topic under investigation (gas molecules) or sliders to manipulate variables were related to improved comprehension (O’Keefe, Letourneau, Homer, Schwartz, & Plass, 2014). The role of informational texts was not addressed in that study. However, other studies (Van der Meij & de Jong, 2006, 2011) have found beneficial effects of the integration of informational texts and simulations, although they did not use eye-tracking. Integrated representations have been shown to result in larger learning gains compared to separate, non-linked representations (Van der Meij & de Jong, 2006). Directing learners’ focus toward relating and translating between representations was related to better learning compared to general prompts (Van der Meij & de Jong, 2011).

1.3. Signaling

Integration can be stimulated by means of signaling: the highlighting of corresponding information in text and pictures (Van Gog, 2014). Signaling has been shown to be effective in fostering comprehension in various domains (Richter et al., 2016). In virtual labs, the information is not static, as is often the case with the illustrations used in signaling studies. Variables can be manipulated in a virtual lab, which causes the settings of these variables and the outcomes of the experiments to vary. Therefore, the implementation of signaling is different when texts are combined with a virtual lab instead of an illustration. Relevant variables can be highlighted to promote using
them, but highlighting of relevant variable settings and experimental outcomes is difficult to do, because the settings and outcomes that are relevant depend on the research question. In addition, identifying potential explanatory variables is essential for learning about the domain, and might be the most difficult step in inquiry-based learning (Siegler & Chen, 1998). Whether signaling also improves the integration of informational texts and a virtual lab during inquiry-based learning remains unclear. Signaling in dynamic displays, such as a virtual lab, is an effective support procedure in learning from multiple representations. However, the effectiveness of signaling for learning also seems to depend on certain prerequisites (such as prior knowledge), and the learning process itself can mediate the effects of these prerequisites on learning outcomes (Renkl & Scheiter, 2017). This notion is in line with the previously mentioned role of domain knowledge in experimentation and the role of experimentation in learning gain.

1.4. The present study

Even though online inquiry-based learning in virtual labs has been extensively studied (e.g., Brinsson, 2015; De Jong, Linn, & Zacharias, 2013), the focus has been on providing scaffolds. The effect of reading informational text and using it in inquiry-based learning has remained a black box up till now, although informational text can be seen as a crucial part of the learning environment. Therefore, in the present study, we examined the reading informational text and integration of the text with a virtual lab during inquiry-based learning in a pair of studies using eye-tracking. Reading time was included as a measure because text can only be used when it is read. In Study 1, we related reading and the quality of the inquiry process to learning gains on a domain knowledge test. Those findings were then incorporated in an adapted version of the learning environment that used signaling to stimulate integration, which was tested in Study 2. The same informational texts were used in both studies. The texts provided three related types of information that are helpful for inquiry-based learning (see Klahr et al., 2011): domain knowledge, such as what is electric current; how to conduct experiments procedurally in the specific virtual lab presented, such as how to place a light bulb in the electrical circuit; and how to use domain knowledge in the experiments, for example, that an ammeter can be used to measure amperage. Note that information about how to conduct experiments did not include epistemological content about the nature of knowledge creation. When assessing reading behavior and inquiry-based learning in relation to learning gains, it is important to control for cognitive and motivational factors. We did so in both studies. Reading performance has been shown to be affected by word decoding skills and vocabulary (White, Graves, & Slater, 1990), as well as by reading motivation (Schiefele & Schaffner, 2016). The quality of the inquiry process has been associated with scores on scientific reasoning measures, such as assessments of the understanding of experimental strategies (Stender, Schichow, Zimmerman, & Härtig, 2018). Furthermore, knowledge of how hypotheses and evidence are related to theories, what is called coordination of theory and evidence, is an important factor in inquiry-based learning (Kuhn, 2004). STEM motivation also plays a role in science learning. Students with higher general interest in science scored higher on the PISA science assessment (OECD, 2007).

2. Study 1

The aim of Study 1 was to investigate the role of reading informational text in an online inquiry-based lesson that was centered around a virtual lab. The following research questions were addressed:

1. What is the association of reading time with acquiring domain knowledge from learning in an inquiry-based learning space, while controlling for cognitive and motivational factors?
2. What are the mediating effects of the quality of the inquiry process (hypothesis generation and experimentation) on the relation between domain knowledge before and after learning in an inquiry-based learning space, and how does integration moderate the relations between prior knowledge and the quality of the inquiry process?

First, we expected reading time to be positively related to learning gains, based on the beneficial effects of informational text in self-regulated learning (Graesser et al., 2005). Second, it was hypothesized that the number of testable hypotheses and score for correctly designed experiments would be positively related to learning gains, as they are part of the inquiry process (Klahr, 2000). Third, integration of the informational text and the virtual lab was expected to be positively related to the score for correctly designed experiments, because integration helps with applying knowledge from the text when using the virtual lab. Correctly designed experiments, in turn, would positively influence the acquisition of domain knowledge, as they provide complementary knowledge to the informational text (Ainsworth, 2006). Different relations were expected for hypothesis generation, because a virtual lab is not needed to create testable hypotheses. The informational text provided all information needed to create hypotheses. Furthermore, guidance was implemented for this step by constraining what hypotheses could be created. Integration might help with hypothesis generation at a more advanced stage of learning, when students have more experience with the topic under investigation and with the virtual lab. Therefore, no association of integration with hypothesis generation in this case was hypothesized. In contrast, the informational text can inform hypotheses. Thus, reading time was hypothesized to be related to hypothesis generation.

2.1. Method

2.1.1. Participants

Seventy-eight students, with a mean age of 13 years and 7 months ($SD = 5$ months), participated in Study 1. There were 46 girls and 32 boys. They were from three eighth-grade classes at a secondary school for pre-vocational education in the Netherlands. The participants’ level of education matched ISCED (International Standard Classification of Education) Level 2, which is slightly below average, but after completing their pre-vocational education, they can continue with education that matches ISCED Level 3. ISCED levels 0, 1, and 2 are attained by 26% of the country’s population, while 74% attain a higher level of education (Eurostat, 2019). Parents and/or caretakers of the students were approached via the school. They received information about the study and were given the opportunity to ask questions of either the researchers directly or the contact person at the school. They could also indicate if they did not want their child to participate in the study. Similarly, the students themselves could refuse to participate and none of them did. The research project was approved by the ethical committee of the faculty.

2.1.2. Materials

2.1.2.1. Inquiry-based learning space about electrical circuits. An Inquiry-based Learning Space (ILS) was designed for this study. An ILS is an online learning environment revolving around inquiry-based learning. We created an ILS that aimed to teach students about electrical circuits, using the Go-Lab ecosystem (www.golabz.eu). With Go-Lab, the designer of an ILS can select the texts to use and can choose from a large range of virtual labs. In addition, there are additional tools in Go-Lab that can be added to support the inquiry process, such as a tool to record experimental outcomes. Informational text and the lab were presented beside each other; see Fig. 1. The electrical circuits lab allowed students to create a circuit by dragging and dropping components and changing them, as well as changing the power levels, to investigate the principles of electrical circuits (Sikken, 2017). The informational text consisted of explanations about
inquiry-based learning, electrical circuits (Kappers & Schatorjé, 2014), and inquiry-based learning with electrical circuits, as well as how to navigate through the ILS, and was divided into 16 lesson phases, as separate pages, which the students could navigate through. The informational text changed over the phases, but the lab remained the same. All students’ actions related to the informational text and in the lab were stored. The ILS was self-paced, which meant that students could proceed to the next phase whenever they were ready.

The informational text included instructions for the students about how to create hypotheses, conduct experiments, and draw conclusions about electrical circuits. In the first phase, the learning goal was presented, along with instructions on how to navigate through the ILS. Electrical circuits were introduced in the second phase. In the third phase, the electrical circuit lab was introduced. Thus, the first three phases allowed students to gather information about the topic and the learning activity. In the fourth phase, electric current and tension and how to measure them were explained; this information could subsequently be used to generate hypotheses. Phases 5 and 6 helped students to explore the variables and the virtual lab. In the fifth phase, students were instructed to create an electrical circuit in the lab, using an example. The example served as a basis for the first two experiments. In the sixth phase, how the power level from the power supply could be changed was explained. Next, students created hypotheses about current (phase 7) and tension (phase 8; called 7b in Fig. 1). For this phase, a Hypothesis Scratchpad was presented to aid the construction of hypotheses; see Fig. 2. Students could drag the terms to the hypothesis field with their mouse and rearrange them, with the goal of creating testable hypotheses. In phase 9, students were instructed to design an experiment; this was the experimentation phase of inquiry-based learning. Phase 10 was for writing down the results, that is, drawing conclusions. In phase 11, an explanation was given of parallel and serial circuits, which indicated the start of another inquiry cycle, beginning with information about the topic. Next, hypotheses were created about current and tension (phase 12) and about whether where measurement happens in the circuit affects current and tension (phase 13) in the two types of circuits. Phase 14 was used for designing experiments, that is, experimentation. Results could be written down in phase 15, drawing conclusions, and finally, in phase 16, students were thanked for their participation and asked to write down any additional questions they had concerning the topic or inquiry-based learning.

2.1.2.2. Domain knowledge (pretest and posttest). A multiple-choice test was designed to assess domain knowledge. The test consisted of 28 questions about electrical circuits, with four answer options each. There were 12 questions that addressed reproduction of knowledge, such as ‘What is an electrical circuit?’ The remaining 16 questions addressed application, with questions such as: ‘If the power from the power source increases, what will happen to the amperage and voltage?’

Reliability was analyzed for the pretest assessment. After removal of three items (two reproduction of knowledge items and one application of knowledge item), the test was sufficiently reliable (Kline, 1993), Cronbach’s $\alpha = 0.60$, Guttman’s $\lambda_2 = 0.62$. Reliability for the posttest was acceptable as well, Cronbach’s $\alpha = 0.60$, Guttman’s $\lambda_2 = 0.64$. The 25 questions addressed reproduction of knowledge (10 questions) and insight (15 questions), and were used to test domain knowledge at pretest and posttest in both studies. While two components might be assumed in this questionnaire, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.43, showing that the data were not suitable
for a separate analysis of components (Field, Miles, & Field, 2013).

2.1.2.3. Inquiry process quality measures. We assessed the number of testable hypotheses and quality of experimentation in the ILS. Testable hypotheses consist of a grammatically correct statement that can be tested with an experiment, regardless of whether it is in line with scientific theory. An example of a hypothesis that would be scored as testable is: “If the power from the power supply increases, then the electric current increases”. This resulted in a maximum score of six, one point per hypothesis. Reliability of the hypothesis score was good (Kline, 1993), Cronbach’s α = 0.75, Guttman’s λ₂ = 0.77.

To determine the score for correct experiments, students’ electrical circuits were evaluated, based on the assignment. The assignment was to test a specific hypothesis. If the circuit was designed in such a way that the hypothesis could be tested, it was counted as a correctly designed experiment. The first circuit students designed was correct if they created a copy of the example presented in the ILS. The parallel circuit was correct when it included a power supply and at least two branches with lightbulbs. The serial circuit was correct when it included a power supply and at least two lightbulbs connected in series. In addition, an additional point could be scored for each circuit when meters were appropriately placed. An example of an experiment that correctly addressed the research question about the effect of power from the power supply on electric current included: a power supply, a light bulb, an ammeter, and all these components connected via wires. This resulted in a total of six possible points for correctly designed experiments. The correct experiment score was reliable (Kline, 1993), Cronbach’s α = 0.62, Guttman’s λ₂ = 0.64.

2.1.2.4. Cognitive and motivational factors

2.1.2.4.1. Word decoding. Word decoding was assessed using a paper-and-pencil lexical decision task (Van Bon, 2006). The task consisted of a card with three columns with 120 bi-syllabic words, of which 90 were (high-frequency) real words and 30 were pseudowords. Students were asked to mark pseudowords by striking through them. Students were given one minute for this task. The last item read was underlined by the student. The reliability of the test has been shown to be good (Kline, 1993), Cronbach’s α = 0.77-0.84 (Van Bon, 2006). One point was given for each correctly marked pseudoword and each correctly unmarked real word, resulting in a possible total of 120 points.

2.1.2.4.2. Vocabulary. The Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) in Dutch (Schlichting, 2005) was used to assess vocabulary. The task for the student was to indicate which out of four pictures matched the spoken word. The assessment was changed compared to the original (Dunn & Dunn, 1997; Schlichting, 2005) in three ways. First, the test was administrated in the classroom, on computers. Second not all items were used, but every third item from set 10 onwards was selected, which resulted in 32 items. Third, we did not use a discontinuation rule. The experimenter read the words out loud, so all students in the classroom could hear them. Students were given enough time to select their answer. The test proved to be unreliable (Kline, 1993), Cronbach’s α = 0.31, Guttman’s λ₂ = 0.37. A likely cause is that a selection of items was used. In addition, most words in the test appeared to be archaic. Therefore, this task was not used in the analyses.

2.1.2.4.3. Reading motivation. Reading motivation was assessed using 21 items that were selected and translated (DUO Onderwijsonderzoek [DUO Educational research], 2017) from the Reading Motivation Questionnaire (RMQ; Schiefele & Schaffner, 2016). We used this selection of items to assess general reading motivation. All items were statements about reading and students answered whether they agreed on a 5-point Likert scale. A higher score on an item indicated greater agreement and a higher score in total indicated higher reading motivation. This measure showed excellent reliability (Kline, 1993), α = 0.86, Guttman’s λ₂ = 0.87. The minimum score was 21 and the maximum score was 105.

Based on the three second-order reading motivation factors addressed in the original RMQ, namely intrinsic, extrinsic, and regulatory motivation (Schiefele & Schaffner, 2016), a confirmatory factor analysis (CFA) was conducted. While factor loadings were all significant, supporting the componental structure, the model did not have a good fit, χ²(168, N = 138) = 416.75, p < .001, CFI = 0.833, RMSEA = 0.104, SRMR = 0.084. Model fit might have been better if all 34 original items were used instead of the 21 selected items. Given that results were the same when using components instead of the total score and we were interested in reading motivation in general, we reported analyses using the total score for reading motivation.

2.1.2.4.4. Scientific reasoning: coordination of theory and evidence. To assess scientific reasoning, a test of coordination of theory and evidence was used, which included questions about generating hypotheses and evaluating evidence. This is one of the components of the Scientific Reasoning Inventory, which was found to be valid and reliable with grade 4 students (Van de Sande, Verhoeven, Kleemans, & Segers, 2019). The set of nine multiple-choice items was not reliable (Kline, 1993) in the current sample, Cronbach’s α = 0.28, Guttman’s λ₂ = 0.38, which might be because students in the present sample were two years older and, as a result, there were ceiling effects on six items (over 80% correct). Reliability could not be improved, and therefore, this task was not used in subsequent analyses.

2.1.2.4.5. STEM attitude. Attitude towards STEM (Science, Technology, Engineering, and Mathematics) was assessed using a questionnaire with 20 items (Denessen, Vos, Hasselman, & Louws, 2015). The items were statements about doing STEM in school, preference for a job in STEM, whether STEM is easy, and whether STEM is useful. Students were asked to what extent they agreed with the statements. A Likert scale with four options was used: strongly agree, agree, disagree, or strongly disagree. Strongly disagree was assigned a score of 1, disagree a score of 2, agree a score of 3, and strongly agree a score of 4. All statements were phrased in such a manner that a higher score indicated a more positive STEM attitude. The test was found to have good reliability (Kline, 1993), Cronbach’s α = 0.78, Guttman’s λ₂ = 0.82. The minimum score was 20 and the maximum score was 80.

2.1.3. Eye-tracking

During the online inquiry-based lesson, eye movements were recorded using Tobii Pro Glasses 2®. The wearable eye-tracker was placed in front of the eyes like regular glasses. If students already wore regular glasses, additional glasses for the wearable eye-tracker were used to match the strength of the regular glasses. The eye-tracking system used the dark pupil technique, with infrared illumination to track the pupil center corneal reflection, at a sampling rate of 50 Hertz. Two cameras were aimed at each eye and a fifth camera recorded the scene (i.e., what the participants looked at). To classify eye movement, the Tobii I-VT Gaze Filter was used.

2.1.4. Apparatus

The online inquiry-based lesson was presented on a LG Flatron E2210® monitor (36.89 × 50.64 cm LED TN computer screen, refresh rate 60 Hertz, and resolution 1680x1050). The screen, as well as a mouse and keyboard, were connected to a laptop with Windows 10® operating system and default screen and color settings. The online lesson was available via the Go-Lab server and all process data were stored securely on the server for later analyses.

2.1.5. Procedure

All tests were administered digitally, except for the word decoding test. Digital scores were recorded online in a secured database. Administration of the cognitive and motivational measures and the domain knowledge pretest was done in the classroom; it lasted for
approximately 60 min, including a small break. One to two weeks later, the ILS about electrical circuits was completed, which lasted for 26 min on average. Directly thereafter, students were given the domain knowledge posttest. Working in the ILS and on the posttest happened in a quiet room inside the school, with two randomly selected students at a time. These students were seated at different desks on opposite sides of the room. Eye-tracking glasses recorded where the students looked in the ILS by measuring their pupils and by recording what the students had in front of them, so that the eye movements could be mapped onto it. The experimenter was in the room and made sure that testing went well. The only instruction students received was that provided in the ILS. When they had a question related to the ILS and its assignments, they were told to use the information that was available in the ILS. Electrical circuits had not been taught yet at school, but they were in the curriculum for later that school year.

### 2.1.6. Data analysis

Before statistical analyses could be performed, eye-tracking data was pre-processed. Two Areas of Interest (AOIs) were created: one for the text and one for the lab. The AOIs were placed on each side of the screen; see Fig. 1. The left side of the screen contained the informational text and the right side of the screen contained the virtual lab. The AOIs were used to determine what eye movements took place within these areas. Fixation time per AOI and switches between the AOIs were established. Fixation time was the total time of all fixations within the AOI. The measure of reading used was the relative amount of reading time. Reading time was adjusted in relation to time in the lab by dividing reading time by lab time. The text AOI was fixated upon for 14 min 38 s, on average (SD = 4 min and 8 s), and the lab AOI for 5 min 44 s on average (SD = 2 min and 45 s).

Switches were used to assess integration. A switch required moving from one AOI to the other; there had to be a fixation on one AOI and then subsequently on the other AOI. To extract switches, first only fixations were extracted from the eye-tracking data. Second, a cut-off of 500 ms was used as a maximum between two subsequent fixations. There could be a delay because there were missing data in-between the fixations; for example, sometimes the eyes could not be found.

Inspection of the data revealed no outlier for posttest domain knowledge, one for STEM attitude, and four for relative reading time. Outliers were identified based on the outlier inter-quartile range rule with a multiplier of 2.2, as suggested by Hoaglin and Iglewicz (1987). Outliers were removed before conducting the analyses. Skewness and kurtosis were acceptable, with corresponding z-values below 1.96 (Field et al., 2013), which was confirmed by visual inspection.

Mediation, moderation, and moderated mediation analyses were performed using the Process plugin (Hayes, 2018) in SPSS. It applies ordinary least squares regression with bootstrapping of mediation effects, because mediation effects are rarely normally distributed (Preacher & Hayes, 2008). The number of bootstrap samples was set at 5000. Significance of mediation effects was evaluated using a 95% CI.

### 2.2. Results

Before answering the research questions, it was necessary to determine whether learning occurred. This was the case; on average, participants got higher scores on the domain knowledge test at posttest than at pretest, as revealed by a paired-samples t-test, t(76) = 5.91, p < .001, d = 0.59, see Table 1. The effect size indicated a medium-sized effect (Cohen, 1992).

The first research question about the association of reading with learning gains was assessed using hierarchical regressions. Posttest domain knowledge was the dependent variable. In the first step, pretest domain knowledge was added to the regression model. In the second step, the control variables were tested, one by one, and none of the control variables was found to explain additional variance in posttest domain knowledge. Next, the relative reading time was added to the model with pretest domain knowledge. Reading time had no effect. Only pretest domain knowledge was associated with posttest domain knowledge, which was in line with the correlations, see Table 1.

To answer the second research question, the role of integration was studied in relation to inquiry-based learning and knowledge acquisition. First, a mediation model was built, with domain knowledge at pretest as the independent variable, domain knowledge at posttest as the dependent variable, and the inquiry process quality variables (testable hypotheses and correct experiments) as mediators. The results showed that pretest domain knowledge predicted testable hypotheses, B = 0.1231, t(75) = 2.18, p = .033; correct experiments, B = 0.1889, t (75) = 4.87, p < .001; and posttest domain knowledge, B = 0.4584, t (73) = 4.98, p < .001. The total effect of pretest domain knowledge on posttest domain knowledge (i.e., without the mediators) was B = 0.6181, t(75) = 7.37, p < .001. Regarding the mediators, correct experiments significantly predicted the posttest score, B = 0.7768, t (73) = 3.30, p = .002, and was a significant mediator of the relation between pretest and posttest domain knowledge, B = 0.1467, CI [0.0472, 0.2851]. This means that higher pretest scores were related to higher correct experiment scores, and higher correct experiment scores were related to higher posttest scores. Testable hypotheses were not related to any of the other variables. The complete model was significant, F(73, 3) = 24.34, p < .001, and explained 50% of the variance in posttest domain knowledge.

Second, in a moderated mediation model, the effect of integration was added; see Fig. 3 and Table 2. Integration moderated the relation between pretest domain knowledge and correct experiments, B = - 0.0009, t(69) = 3.13, p = .003. It appeared that the relation between pretest scores and correct experiments was absent for students with a high level of integration (84th percentile), but the relation was significant for students with an average (50th percentile) or low (16th percentile) level of integration; see Table 3 and Fig. 4.

In line with its moderation of the relation between pretest domain knowledge and correct experiments, integration also had a significant moderating effect on the mediation by correct experiments of the relation between pretest and posttest domain knowledge, namely, a moderated mediation effect, B = -0.0021, CI [-0.004, -0.0001]. In other words, the mediation of the relation between pretest and posttest knowledge by the score for correct experiments was significant in the below-average integration group, B = 0.2061, CI [0.0611, 0.3945], and the average integration group, B = 0.1072, CI [0.0278, 0.2248], but it was not significant for the above-average integration group, B = 0.0097, CI [-0.0731, 0.1362]. Integration was also directly related to the score for correct experiments, B = 0.0394, t(69) = 3.45, p = .001. The coefficients for the final model with moderated mediation can be found in Table 2.

### 2.3. Discussion

The aim of Study 1 was to investigate the association of reading time and integration with the quality of the inquiry-based learning process and learning gains. On average, there was a significant learning gain. Relative time spent reading (as compared to time spent in the lab) was not related to learning gain or inquiry process quality. Regarding the inquiry process, the score for correct experiments mediated the relation between pretest and posttest domain knowledge, but no effect of the number of testable hypotheses on learning gains was found. Finally, integration was related to the score for correct experiments and moderated the relation between pretest domain knowledge and the score for correct experiments. This showed that for participants with high integration, there was no relation between pretest domain knowledge and score for correct experiments, while there was a positive relation between pretest domain knowledge and score for correct experiments for participants with medium and low levels of integration.

In contrast to our expectations, reading time was not related to inquiry process quality or knowledge acquisition. It might be that most...
students did read the text, as suggested by the total fixation time on the text, which was 15 min, on average. In addition, the online lesson was self-paced, which allowed students to spend more time on reading, if needed.

3. Study 2

The results of Study 1 indicated that integration positively affects performance in an online inquiry-based lesson. It can thus be hypothesized that support of integration could lead to higher learning gains. One effective approach to supporting integration is the use of signals. Signaling has been shown to promote learning gains in various domains (Richter et al., 2016). Signals can highlight corresponding features of the informational text and a pictorial representation (i.e., the virtual lab in the present study). In this way, signals catch the learners’ attention, as revealed by fixations on the signaled features, and this looking behavior is related to better performance on a knowledge test after learning in a digital environment with signals (Scheiter & Eitel, 2015). Therefore, color signals might lead to more integration as revealed by fixations, resulting in larger learning gains. Signaling should lead to longer processing time, as it stimulates integration, which takes time (Schneider, Bege, Nebel, & Rey, 2018). Therefore, signals may also lead to more time spent in phases where the signals appear. To test these hypotheses, students in Study 2 used the same materials as in Study 1 with one difference, namely, the addition of color signals. To sum up, the following research questions were addressed in Study 2:

1. How do performance scores (including learning gains) compare between Study 1 (control) and Study 2 (signaling)?
2. To what extent does signaling have the expected effects: more integration and more time spent in phases with signals?
3. How do the relations between pretest domain knowledge, correctly designed experiments and testable hypotheses, correct experiments, and integration, as well as Pearson’s Correlations.

Hypotheses were that (1) students in Study 2 would show a larger learning gain than students in Study 1, while no difference would exist on the pretest in both studies; (2) signaling would be associated with more integration and time spent in phases with signals compared to the same phases with no signaling in the control condition, and (3) condition would moderate the effects of integration, namely its direct effect on score for correct experiments, resulting in a two-way interaction, and its moderating effect on the relation between prior knowledge and score for correct experiments, resulting in a three-way interaction.

3.1. Methods

3.1.1. Participants

Seventy-one students, mean age of 13 years and 7 months (SD = 6 months), participated in Study 2. There were 30 girls and 41 boys. The participants were recruited in the same way as in Study 1. Study 2 was conducted one year after Study 1. The students were from the same school, followed the same educational program, had the same teachers, and were tested in the same setting. Hence, the participants in Study 2 were comparable to those in Study 1.

3.1.2. Materials

3.1.2.1. Inquiry-based learning space with color signals. As in Study 1, the Inquiry-based Learning Space (ILS) aimed to teach students about electrical circuits via inquiry-based learning. In Study 2, color signals for five components that could be used in the lab were added to the informational text and online lab, namely, the power supply and light...
bulb in phase 3, the voltmeter and ammeter in phase 5, and the power controller in phase 6. These words were underlined in the text and students were instructed to click on them. After they clicked on the word, both the word and the corresponding component in the lab blinked for four seconds, after which the background of the word and the component remained colored; see Fig. 5. The components had different colors. We chose to implement the signals in this way, because color signaling has a beneficial effect on learning and blinking helps to draw the attention, while flashing signals that disappear have a negative effect on learning (Schneider et al., 2018).

3.1.2.2. Inquiry measures. As in Study 1, correctly designed experiments and testable hypotheses were assessed during inquiry-based learning. In addition, time spent per phase was assessed. Time spent per phase was defined as the time from clicking to open a phase until clicking to open on another phase. When a student revisited a phase, that time was added to the time spent in that phase. The time spent in the final phase could not be determined exactly, because there was no phase after it. In the final phase, students were thanked and could write down additional questions. Two scores for time spent were added as a covariate. The analyses showed that reading motivation had no association with the other measures, and therefore the analyses was adjusted. To correct for multiple testing, the linear step-up procedure was used (Benjamini & Hochberg, 1995) with the ‘mutoss’ package (Team, 2017) in R (R Core Team, 2019).

3.2. Results

3.2.1. Comparison of Study 1 (control) and Study 2 (signaling)

First, performance in the two studies was compared; see Table 4. Study 1 is the control condition and Study 2 is the experimental condition with signaling. The participants in the two conditions did not differ on pretest or posttest domain knowledge, inquiry process quality measures, relative reading time, integration, or cognitive and motivational factors, except for reading motivation and integration. Participants in the control condition (Study 1) scored higher on reading motivation than participants in the signaling condition (Study 2). Therefore, whenever conditions were compared in subsequent analyses, we checked whether reading motivation impacted the results, when added as a covariate. The analyses showed that reading motivation had no association with the other measures, and therefore the analyses without reading motivation are presented. Regarding integration, the results showed that participants in the signaling condition on average

Table 2

Results of Moderated Mediation Analysis on Posttest Domain Knowledge by Pretest Domain Knowledge, and Testable Hypotheses and Correct Experiments as mediators and Integration as Moderator.

| Dependent Variable               | Independent Variables | B     | SE   | t     | p      | R²   |
|----------------------------------|-----------------------|-------|------|-------|--------|------|
| Testable hypotheses              | Pretest¹              | 0.1932| 0.1908| 1.0124| .3149  |      |
|                                  | Integration           | 0.0044| 0.0175| 0.2497| .8035  |      |
|                                  | Interaction*²         | −0.0003| 0.0014| 0.2198| .8267  | 0.0952|
| Correct experiments              | Pretest¹              | 0.5332| 0.1243| 4.2893| .0001  |      |
|                                  | Integration*²         | 0.0594| 0.0114| 3.4538| .0009  |      |
|                                  | Interaction*³         | −0.0029| 0.0009| 3.1258| .0026  |      |
| Posttest domain knowledge        | Pretest¹              | 0.4333| 0.0943| 4.5942| <.0001 |      |
|                                  | Testable hypotheses    | 0.1627| 0.1716| 0.9479| .3465  |      |
|                                  | Correct experiments*²  | 0.7128| 0.2426| 2.9382| .0045  |      |

Note. ¹Pretest = Pretest domain knowledge. ²Interaction between pretest domain knowledge and correct experiments. *p < .05.

Table 3

Conditional Effects of Pretest Domain knowledge on Correct Experiments for Three Groups of Integration.

| Dependent variable | Independent variable | Moderator | Moderator group | B     | SE   | t     | P     |
|--------------------|----------------------|-----------|-----------------|-------|------|-------|-------|
| Experiment¹        | Pretest²             | Integration | Low*            | 0.2892| 0.0554| 5.221 | <.0001|
|                    |                      |           | Medium*         | 0.1505| 0.0382| 3.9390| .0002 |
|                    |                      |           | High            | 0.0136| 0.0611| 0.2229| .8243 |

Note. ¹Experiment = Correct experiments. ²Pretest = Pretest Domain Knowledge. *p < .05.
had more switches between text and lab. The data that were used for
subsequent analyses combined data from Studies 1 and 2 in order to be
able to investigate the possible moderating effect of condition. Corre-
lations in Study 2 revealed a similar pattern as the correlations of Study
1; see Table 5. In the combined data, STEM attitude was significantly
correlated with pretest and posttest domain knowledge, while this was
not the case in Study 1. We therefore also ran all analyses with STEM
attitude as a covariate, but as this did not affect the results, we report
the analyses without this covariate.

3.2.2. Domain knowledge learning gains

We next compared the learning gains over the two conditions (hy-
pothesis one). To investigate the gain in domain knowledge, a mixed
ANOVA with condition (control/signaling) as between-subjects factor,
and time (pretest/posttest) as within-subjects factor was conducted.
The main effect of time was significant, \(F(1, 144) = 56.45, p < .001\),
partial \(\eta^2 = 0.28\), indicating that, on average, posttest scores were
higher than pretest scores. The effect of condition and the interaction
between time and condition were not significant, \(F(1, 144) < 0.50\) in
all cases. This indicated that scores of the participants in the two con-
ditions on the domain knowledge test did not differ at pretest and
posttest, and neither did their learning gain.

3.2.3. The effect of signaling

In order to address the second hypothesis about the effect of
signaling on integration, a validity check was conducted and time spent
per phase was investigated. The comparison of conditions showed that
students in the signaling condition did more integrating than students
in the control condition. As a validity check of the signal manipulation,
integration directly preceding the activation of the signals (by clicking
on them) was compared to integration directly after activation of the
signals. It was expected that there would be more integration after the
signal. In two phases (phases 3 and 5) there were two signals that were
activated directly after each other, as revealed by the time between
activation of the signals in phase 3 (\(M = 9.77\) s, \(SD = 18.85\)) and in
phase 5 (\(M = 5.89\) s, \(SD = 6.01\)). To control for differences in the total
amount of integration and the total time spent per phase, instances of
integration per second were calculated from the start of a phase (in
phase 3 or 5) to the first signal and following the second signal until the
end of the phase (in phase 3 or 5). In phase 6, there was only one signal
and the instances of integration per second before and after the signal
were compared. A paired samples \(t\)-test showed that there were more
instances of integration per second after the signals than before; see
Table 6. This indicated that the manipulation (signaling) had the ex-
pected effect.

Processing the signal might also lead to more time spent in the
phases with the signals. To test whether the time spent in the phases
differed between the conditions, a repeated measures ANOVA with
condition (control/signaling) as between-subjects factor, and phase
(1–15) as within-subjects factor was conducted. A Huyhn-Feldt

---

### Table 4

|                      | Control   |       |       | Signal    |       |       |
|----------------------|-----------|-------|-------|-----------|-------|-------|
|                      | \(M\)     | \(SD\) | \(N\) | \(M\)     | \(SD\) | \(N\) |
| Pretest              | 11.88     | 3.63  | 78    | 12.12     | 3.88  | 68    |
| t-test               | \(t(144) = 0.37\) | .746  | −0.23 | 0.06      |
| Posttest             | 13.86     | 3.43  | 78    | 13.66     | 3.99  | 71    |
| Word decoding        | 69.62     | 20.65 | 76    | 77.07     | 22.15 | 68    |
| t-test               | \(t(142) = 2.09\) | .115  | −7.46 | 0.35      |
| STEM attitude        | 48.71     | 6.77  | 77    | 47.21     | 8.79  | 68    |
| Reading motivation*  | 43.79     | 9.29  | 78    | 35.25     | 11.88 | 68    |
| t-test               | \(t(144) = 5.04\) | <.001 | 8.54  | 0.84      |
| Reading- vs lab-time | 2.74      | 0.93  | 70    | 2.62      | 1.12  | 65    |
| t-test               | \(t(133) = 0.64\) | .670  | 0.11  | 0.11      |
| Correct experiments  | 4.05      | 1.83  | 77    | 3.83      | 1.99  | 71    |
| t-test               | \(t(146) = 0.71\) | .670  | 0.22  | 0.12      |
| Integration*         | 3.90      | 1.41  | 78    | 3.49      | 1.64  | 69    |
| t-test               | \(t(145) = 1.61\) | .248  | 0.40  | 0.27      |
| Integration**        | 125.76    | 44.31 | 74    | 157.07    | 55.73 | 69    |
| t-test               | \(t(141) = 3.73\) | .001  | 26.63 | 0.56      |

* \(p < .05\).

---

Fig. 5. A screenshot of the task set-up. The left side consisted of informational text and the right side of the electrical circuits lab.
correction was used for the within-subject effect (phase) and the interaction (phase × condition), because the sphericity assumption was violated (Mauchly’s $W = 0.001$, $p < .001$). There was an interaction effect of phase and condition, $F(7.37, 1068.59) = 3.11$, $p = .002$, partial $\eta^2 = 0.02$, and a main effect of phase $F(7.37, 1068.59) = 67.19$, $p < .001$, partial $\eta^2 = 0.32$. The effect of condition was not significant, $F(1, 145) = 0.05$, $p = .831$, partial $\eta^2 < 0.01$. The interaction effect of phase and condition showed that the conditions differed in the pattern of time spent across the phases. As a follow-up, the effect of condition was investigated per phase using independent samples $t$-tests; see Table 7. Whenever equal variances could not be assumed, as revealed by a significant $F$-test comparing the variances of the two samples, the Welch approximation of the degrees of freedom was used in the $t$-test.

The phases with signals did show slowing down in the signaling condition. This effect was not significant for phase 5. There was a trend towards more time spent in phase 3 in the signaling condition compared to the control condition. In phase 6, the signaling condition spent more time. In two out of four phases involving hypotheses, phases 7 and 13, the signaling condition spent less time than the control condition. As an extra test, the number of testable hypotheses was compared between the signaling and control condition for the individual phases and they did not differ, $p$s > .05.

### 3.2.4. The effect of condition on the relations between domain knowledge, the inquiry process, and integration

The third research question concerned the relation of signaling with domain knowledge, correct experiments, testable hypotheses, and integration. We hypothesized that condition would moderate all effects of integration. To investigate whether the relations were different between Study 1 (control condition) and Study 2 (signaling condition), a moderated mediation model was constructed, just as in Study 1, but condition was added as a moderator of integration. Pretest domain knowledge was the independent variable, correct experiments and testable hypotheses were mediators, integration was a moderator of the relations between pretest domain knowledge and each of the mediators, condition (control/signaling) was a moderator of the same relations as integration and also a moderator of the relations of integration with testable hypotheses and correct experiments, and posttest domain knowledge was the dependent variable; see Fig. 6.

Pretest domain knowledge was related to correct experiments, $B = 0.8265, t(131) = 2.88, p = .005$, and posttest domain knowledge, $B = 0.4825, t(135) = 7.39, p < .001$. The score for correct experiments was related to posttest domain knowledge, $B = 0.8283, t(135) = 5.13, p < .001$. Integration was related to correct experiments, $B = 0.0704, t(131) = 2.70, p = .008$. Furthermore, there was a three-way interaction between pretest domain knowledge, integration, and condition on correct experiments, $B = 0.0028, t(131) = 2.30, p = .023$. Integration moderated the relation between pretest domain knowledge and correct experiments in the control condition, $B = -0.0029, t(131, 1) = 8.50, p = .004$, but not in the signaling condition, $B = -0.0001, t(130, 1) = 0.02, p = .891$. This effect can be explained by the fact that pretest domain knowledge and correct experiments were not related for participants with high integration in the control condition, whereas they were related in the signaling condition; see Table 8.

In addition, two two-way interactions were found; namely, pretest domain knowledge interacted with integration on correct experiments, $B = -0.0058, t(131) = 2.70, p = .008$, and integration interacted with condition on correct experiments, $B = -0.0310, t(131) = 2.00, p = .048$. The first interaction showed that pretest domain knowledge and correct experiments were related for low and medium integration participants, but not for the high integration group. However, interpretation of this result should consider the three-way interaction, because that interaction showed that for the high integration group, there was a significant relation between pretest domain knowledge and correct experiments in the signaling condition and not in the control condition. The second interaction showed that integration and correct experiments were related in the control condition, $B = 0.0394, t(69) = 3.45, p = .001$, and not in the signaling condition, $B = 0.0085, t(62) = 0.83, p = .410$.

Finally, the possible moderating effect of condition on the moderated mediation effect detected in Study 1 was tested. In Study 1, integration was found to moderate the mediating effect of correct experiments in the relation between pretest and posttest domain knowledge. The present analysis showed that this moderated mediation effect was moderated by condition, $B = 0.0024, CI [0.0002, 0.0048]$, which indicated that the mediating effect of correct experiments was moderated by integration in the control condition, $B = -0.0024, CI [-0.0046, -0.0005]$, but not in the signaling condition, $B = -0.0001, CI [-0.0013, 0.0010]$.

### Table 6
Comparison of the Instances of Integration per Second Before and After the Signals for Phase 3 and 5.

| Integration per second | Before the signal(s) | After the signal(s) |
|-----------------------|---------------------|---------------------|
|                       | $M$    | $SD$ | $n$ | $M$    | $SD$ | $n$ | $t$  | Adj. $p$ | $M1-M2$ | $d$  |
| Phase 3$^*$ | 0.11  | 0.11  | 60  | 0.33  | 0.16  | 62  | 8.89 | <.001  | −0.22  | 1.64 |
| Phase 5$^*$ | 0.08  | 0.11  | 62  | 0.32  | 0.12  | 63  | 11.15 | <.001  | −0.24  | 1.99 |
| Phase 6$^*$ | 0.06  | 0.13  | 63  | 0.13  | 0.12  | 63  | 2.97  | .004  | −0.07  | 0.55 |

Note. $^*$ $p < .05$.
3.3. Discussion

The aim of Study 2 was to investigate inquiry-based learning and domain knowledge when signaling was incorporated in the inquiry-based lesson. To evaluate the effect of signaling, the results were compared to Study 1 without signaling. Condition (control/signaling) did not affect performance, because performance was the same in the control and signaling conditions for pretest and posttest domain knowledge, testable hypotheses, and correct experiments. Learning gain also did not differ between the signaling and control conditions; students in both conditions had higher posttest than pretest domain knowledge scores. In addition, signaling showed the expected effects on integration and time spent per phase. Students in the signaling condition showed more integration in all phases with signals and more time spent in some phases with signals compared to students in the control condition. Interestingly, in the signaling condition, there was speeding up in some phases with hypotheses, while number of testable hypotheses did not drop compared to the control condition. Finally, condition did affect the inquiry process by moderating the effects of integration. Integration was not related to the inquiry process in the signaling condition, while it was in the control condition.

4. General discussion

The general aim of the studies was to investigate processing of informational text and a virtual lab in inquiry-based learning and potential effects on learning about the domain. Study 1 showed that integration of text and the virtual lab fostered inquiry-based learning, in the sense that participants who integrated more also got higher scores for correct experiment design, and those who got higher scores for correct experiments had higher learning gains. It was therefore expected that stimulation of integration through signaling would lead to improvements in experiment design and gains in posttest domain knowledge. Table 7

| Phase | Control | Signal |
|-------|---------|--------|
|       | M | SD | n | M | SD | n | t-test | Adj. p | M1-M2 | d |
| 1 | 71.51 | 73.51 | 77 | 70.82 | 79.96 | 71 | t(146) = 0.05 | .957 | 0.69 | 0.01 |
| 2 | 74.76 | 43.75 | 78 | 63.20 | 26.03 | 71 | t(127.35) = 1.98 | .149 | 11.56 | 0.32 |
| 3* | 124.54 | 58.35 | 78 | 148.46 | 59.31 | 71 | t(147) = 2.48 | .054 | −23.92 | 0.41 |
| 4 | 50.05 | 26.31 | 78 | 51.29 | 22.92 | 71 | t(147) = 0.30 | .849 | −1.24 | 0.05 |
| 5 | 92.81 | 66.78 | 78 | 105.30 | 52.92 | 71 | t(144.32) = 1.27 | .496 | −12.49 | 0.21 |
| 6* | 39.50 | 22.69 | 78 | 50.79 | 28.86 | 71 | t(132.75) = 2.64 | .047 | −11.29 | 0.44 |
| 7* | 95.50 | 62.49 | 78 | 70.51 | 46.14 | 71 | t(141.15) = 2.79 | .045 | 24.99 | 0.45 |
| 8 | 48.23 | 40.21 | 78 | 50.69 | 34.19 | 71 | t(141) = 0.40 | .849 | −2.46 | 0.07 |
| 9 | 122.07 | 68.35 | 78 | 131.32 | 95.80 | 71 | t(125.42) = 0.67 | .838 | −9.25 | 0.11 |
| 10 | 108.49 | 64.67 | 78 | 125.78 | 104.24 | 71 | t(114.83) = 1.20 | .496 | −17.29 | 0.20 |
| 11 | 81.59 | 61.75 | 78 | 77.33 | 73.87 | 71 | t(147) = 0.38 | .849 | 4.26 | 0.06 |
| 12 | 103.04 | 54.97 | 78 | 101.33 | 61.82 | 71 | t(147) = 0.18 | .859 | 1.70 | 0.03 |
| 13* | 129.59 | 77.57 | 78 | 79.70 | 38.77 | 71 | t(115.50) = 5.03 | <.001 | 49.88 | 0.80 |
| 14 | 200.89 | 121.89 | 77 | 222.78 | 159.69 | 71 | t(130.68) = 0.93 | .662 | −21.89 | 0.15 |
| 15 | 148.25 | 79.15 | 78 | 151.82 | 86.09 | 71 | t(147) = 0.26 | .916 | −3.57 | 0.04 |
| Total | 1485.69 | 421.19 | 76 | 1501.13 | 454.29 | 71 | t(145) = 0.21 | .916 | −15.44 | 0.04 |

Note. *p < .05. † p < .10.

Table 8

| Condition | Integration | B | SE | t | p |
|-----------|-------------|---|----|---|---|
| Control   | Low*        | 0.2700 | 0.0548 | 4.93 | <.001 |
|           | Medium*     | 0.1269 | 0.0426 | 2.98 | .003 |
|           | High        | −0.0203 | 0.0750 | 0.27 | .787 |
| Signal    | Low*        | 0.2310 | 0.0529 | 4.36 | <.001 |
|           | Medium*     | 0.2263 | 0.0396 | 5.71 | <.001 |
|           | High*       | 0.2214 | 0.0530 | 4.18 | <.001 |

Note. *p < .05.
knowledge. This was tested in Study 2. Although no improvement in score for correct experiments or domain knowledge was found, signaling did lead to more integration and more efficient hypothesis generation.

4.1. Learning gains

Students learned about the domain in both studies, which supports the claim that inquiry-based learning is effective in promoting knowledge acquisition. This is in line with previous research on (online) inquiry-based learning (De Jong et al., 2013; Lazonder & Harmsen, 2016). The effect sizes indicated a medium-sized effect (Cohen, 1988, 1992): a $d$ of 0.59 in Study 1 and a partial $\eta^2$ of 0.28 for Studies 1 and 2 combined. Compared to other effect sizes within educational research, the present effect sizes for learning gain are above what is usually found (Hattie, 2009). Furthermore, this effect was found after a single inquiry-based learning session, which suggests that effective learning materials were used in the present studies.

In both Study 1 and Study 2, learning was found to be related to experimentation. When experiments were designed correctly, the outcome of the experiments offered an opportunity to acquire domain knowledge; these opportunities would obviously be missed when experiments were designed incorrectly. Others have also found a relation of experimentation quality with knowledge acquisition in inquiry-based learning (Edlesbrunner, Schalk, Schumacher, & Stern, 2018). We found no effect of hypothesis generation on acquisition of domain knowledge, which might be explained by the learning trajectory in the inquiry process, in which learning about the domain usually takes place at the end, that is, when translating evidence into theory, while at the start of inquiry, prior knowledge (theory) is used to create testable hypotheses (Kuhn, 2004). Indeed, there was a positive relation between pretest scores and number of testable hypotheses. The results also showed a relation between pretest domain knowledge and experimentation quality in a regression with an effect size ($R^2$) of 0.33, indicating a large effect (Cohen, 1992). An increase of one point for pretest domain knowledge was associated with an increase of half a point for correct experiment design. Knowledge about the domain, in this case electrical circuits, helped with setting up informative experiments. Prior knowledge could be used in understanding and recognizing what the variables and their settings were, which in turn would affect the quality of experimentation. This finding is in line with the notion that for most experiments, context matters, such that students use their pre-existing knowledge to set up an experiment (Samarapungavan, 2018).

The final model for Study 1 explained 47% of the variance in posttest domain knowledge; in the final model with Studies 1 and 2 combined, 57% of the variance in posttest domain knowledge was explained. Much of this was due to the use of pretest domain knowledge as a predictor. However, insofar as experiment quality explained additional variance, the results support the relevance of experimentation in explaining learning gains. Regarding the explained variance in posttest domain knowledge, the $R^2$ indicated a large effect (Cohen, 1992). Each increase of one point in the score for experiment design was associated with an increase of 0.71 points in posttest domain knowledge. Note that part of this effect could be because correct experiments mediated the effect of pretest domain knowledge on posttest domain knowledge. While taking this mediation effect into account, pretest domain knowledge still had an effect on posttest domain knowledge (a direct effect; Hayes, 2018). This effect showed that an increase of one point in pretest domain knowledge was associated with an increase of 0.43 points in posttest domain knowledge.

4.2. Reading and integration

In contrast to our hypothesis, time spent reading informational text was not related to the quality of the inquiry process or to learning gains. This can be explained by the finding that average reading time was rather high (about 15 min). The use of eye-trackers might have caused this effect (i.e., that students in general did read the texts). It has been shown that the knowledge that one’s eyes are being watched affects looking behavior to make it more socially desirable, compared to when one is unaware of one’s looking behavior being monitored (Risko & Kingstone, 2011). As the testing situation is also a social context (Hertwig & Ortmann, 2001), it might be assumed that the students in our experiment showed socially desirable behavior: paying attention to and putting effort into the inquiry-based lesson.

Our hypothesis was based on the assumption that if students started conducting experiments without reading the informational text, then experiment quality and posttest domain knowledge should be lower. Informational text can help with understanding the domain, inquiry-based learning, and their combination: inquiry-based learning within the domain (Klahr et al., 2011; Samarapungavan, 2018). The present results did show that informational text has a beneficial effect, because more integration of the informational text and the virtual lab was associated with a higher score for correct experiment design, which in turn was related to higher learning gains. Integration also moderated the relation of pretest domain knowledge and correct experiments in the control condition. This means that switching the gaze between informational text and the virtual lab helped with setting up experiments and learning about the domain, on average. Integration has previously been shown to be beneficial for learning. According to Mayer (2014), integration of corresponding information from different representations is required for meaningful learning. Our studies add that this is also the case when applied in an inquiry-based context and how: namely, via improved experimentation. The result showing integration to be a moderator also indicated that low pretest knowledge could be compensated for by more integration, as far as increasing the score for correct experiment design. This finding is in line with results from signaling studies that have shown larger beneficial effects of signaling in students with medium to low prior knowledge (Richter et al., 2016). Integration was not related to hypothesis generation. One explanation could be that the information that could help in generating hypotheses was provided only as text. While the virtual lab might have constrained the application of knowledge (Ainsworth, 2006), such as what meters could be used to measure what effects, the lab did not provide additional knowledge that could be integrated with the text to help generate hypotheses.

4.3. Signaling

When integration is stimulated by means of signaling, that is, using signals to explicitly highlight correspondences, transfer and comprehension are better than without the signals (Richter et al., 2016). In Study 2, we tested whether signaling would increase the score for correct experiments designed and posttest learning gains in domain knowledge compared to the control condition in Study 1. No difference in score for correct experiments or learning gains was found between the control and signaling conditions. Thus, signaling did not promote better experimentation or learning about the domain. In Study 1, the score for correct experiments depended on two sources of information, namely students’ pre-existing domain knowledge, because an association of pretest domain knowledge with correct experiments was found, and the available information from the online lesson, because an association of integration with correct experiments was found. The balance between relying on prior knowledge and on presented informational text about the domain appeared to shift to more dependence on informational text when integration was high, as revealed by the moderating effect of integration on the relation between pretest domain knowledge and correct experiments. There was no relation between pretest domain knowledge and correct experiments for high levels of integration. In addition, the moderating effect of integration suggested that the informational text about the domain could be disregarded when prior knowledge was high.
With this discussion of the findings from the control condition in mind, the results for the signaling condition can be explained. The balance between use of prior knowledge and informational text about the domain appeared to be different in the signaling condition, because here we found no effects of integration. The signals did foster integration, as integration scores were higher compared to the control condition. Therefore, more integration could have led to more use of the presented informational text when designing experiments in the virtual lab. However, other factors than those we investigated might be involved in additional improvement in experimentation. Examples of factors that relate to scientific reasoning (and experimentation) are general reading comprehension abilities (Van de Sande et al., 2019) and self-control during learning (Van der Graaf, Segers, & Verhoeven, 2018). The finding that signaling was related to more integration, but not to higher learning gain, was in line with a previous study on illustrated text (Scheiter, Schubert, & Schüler, 2018). The manipulation of the learning environment in the study by Scheiter et al. was related to more integration, but learning gain was not related to integration. The authors suggested that integration is only related to learning outcomes up to a certain level of integration and that excessive integration may reflect dysfunctional learning behavior.

Another explanation of the ineffectiveness of signals in promoting learning gains could be that the control condition already had some features that stimulated learning. The lesson was self-paced, which allows students to take the time they need to comprehend the content and finish the assignments. Self-paced lessons have been found to profit to a lesser extent from signaling than system-paced lessons (Richter et al., 2016). The control condition already included some signals of corresponding features, but only as text. Color signaling might not have additional benefits, because discursive (text) and visual signals appear equally effective in fostering knowledge acquisition (Richter et al., 2016). Therefore, signaling effects should be easier to detect when the control condition receives less support. A result from the present studies that supports the beneficial effects of signaling was the speeding-up of hypothesis generation.

### 4.4. Signaling and hypothesis generation

When checking the validity of the signal manipulation by analyzing the time spent per phase, including the phases with signals, an interesting result was found. While it was expected that phases with signals would show slowing down in the signaling condition, no hypothesis was formulated concerning the time spent in phases without signals. One phase with signals did not show an effect, one did show slowing down, and one showed a trend towards slowing down. Signals take time to process, because they stimulate integration (Schneider et al., 2018), which explains the slowing down in the phases with signals. Regarding phases with hypotheses, speeding up when generating a hypothesis in the signaling condition was found in two out of the four phases with hypotheses. This indicated that the process of generating a hypothesis was more efficient half of the time, while the quality of the hypotheses did not change (number of testable hypotheses was the same for the signaling and control conditions). Signaling stimulated integration, which in turn could have led to faster translation of mental theories into concrete investigations. Mental theories are assumed to play in role in generating hypotheses (Klahr & Dunbar, 1988; Kuhn, 2004). This process of going from mental theories to explicit hypotheses is fostered by verbal support (Van der Graaf, Gijsel, Van de Sande, & Segers, 2019) and therefore shows dependence on verbal abilities. The present finding fits with this notion, because integration of informational text (verbal) and the virtual lab was fostered in the signaling condition. Therefore, the students might have had better access to their mental models, insofar as these models were activated in order to assimilate the new knowledge from informational text. Similarly, new mental models could have been readily available to the students as a result of integration.

### 4.5. Limitations and suggestions

A first limitation is that the students were not tested longitudinally. It is unclear whether students improved in inquiry-based learning abilities or what levels of domain knowledge would be found in a delayed posttest. A second limitation is that integration was not compared to other approaches to learning. Therefore, future studies could compare the use of different learning strategies, such as summarization or enactment (Fiorella & Mayer, 2016). A related limitation was that the eye-tracking measures were relatively coarse. Fine-grained eye-tracking could reveal which sentences are read and what text is fixated upon when integrating the informational text with the virtual lab. While all of the text was designed to be useful for the inquiry-based lesson, some text was related to the topic and some text was related to inquiry-based learning. The expectation would be that the text that corresponds to elements in the virtual lab is used when integrating the two sources.

It should furthermore be noted that the operationalization of integration in the present study was coarse. In Study 2, the results showed no effects of integration; therefore, the way we measured integration might not reflect productive integration. To have more insight into the link between the eye movements and the learner’s (retrospective) intention, think-aloud protocols could possibly be included. There are several arguments to support our claim that assessing level of integration via number of switches between text and lab is relevant, as: (1) this is how integration has previously been operationalized (e.g., Mason et al., 2013; O’Keefe et al., 2014); (2) given its relations with the other variables in Study 1, it is reasonable to assume that the measure reflects a useful learning process; and (3) besides the required going back and forth from the text to the lab, which should result in a minimal amount of switching, additional going back and forth does seem to reflect a decision by the learner and thus, purposeful processing.

A final limitation is that the two studies were conducted in two cohorts, which were one year apart. Therefore, other factors might have affected the present results, especially the comparisons between Study 1 and 2. Potential differences between the cohorts were tested and one difference was found, for reading motivation, which appeared not to be related to the inquiry process or learning gain. Additional analyses also showed no effect of reading motivation. Furthermore, the participants were from the same school (with the teaching staff), with the same educational level and background. Condition did not affect any of the other relations in the mediation models besides the ones that were expected to change due to the addition of signaling.

### 4.6. Implications

To sum up, the results of our studies showed that learning took place and was related to the score for correct experiments. In addition, integration of the informational text and the virtual lab was beneficial for online inquiry-based learning. A theoretical implication regarding integration of text and pictures is that the beneficial effects can be extended to text and virtual labs in the domain of inquiry-based learning. Moreover, the present results provided a hint as to how integration was beneficial for learning. Inquiry-based learning has multiple phases (Klahr, 2000), and we found an association of integration with experimentation. Thus, integration of informational text and the virtual lab helps with applying the information in the virtual lab when conducting experiments.

Additionally, learning may be enhanced by providing instructional feedback on the experiments that the students have designed. If students used the feedback, then they would have the opportunity to improve the designs of their experiment in order to (better) address the research question at hand. As correctly designed experiments were related to greater learning gains in domain knowledge, it can be expected that learning gains would be larger if experiment designs improved. This might be especially beneficial for students who do not yet comprehend experimentation. One example is a study in kindergarten that
used feedback on the children’s actual experimental designs (Van der Graaf, Segers, & Verhoeven, 2015), which showed that by adding feedback and other supportive features, kindergartners could correctly design experiments with multiple variables.

A practical implication is that language and STEM education cannot be regarded as dissociated topics. The present studies suggested that students created mental representations of the inquiry process by integrating informational text and the virtual lab. Others have also found verbal factors to be related to scientific reasoning abilities and the development of these abilities (Van der Graaf et al., 2018). When teachers are trained to provide verbal support during inquiry-based learning, students’ inquiry-based learning and the outcomes improve (Van der Graaf et al., 2019). It is therefore relevant for teachers to elicit students’ conceptions and support their reasoning during inquiry-based learning. This support can be about the topic, inquiry-based learning, or a combination of the two (Samarapungavan, 2018).

4.7. Conclusion

Study 1 showed that inquiry-based learning led to learning gains and that experimentation and integration of the informational text and virtual lab affected learning. Thus, integration helps during inquiry-based learning, due to the effect of integration on the experimentation phase of inquiry-based learning. In Study 2, the effect of fostering integration via signaling was tested. Signaling did foster integration, but integration did not relate to experimentation or knowledge acquisition in Study 2. This explains why learning gains were not larger in Study 2 than in Study 1. Interestingly, signaling did make the process of hypothesis generation more efficient.

References

Aditomo, A., & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. International Journal of Science Education. https://doi.org/10.1080/09500693.2020.1716093.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with technology. Journal of the Royal Statistical Society, 57, 289–300. https://doi.org/10.1111/j.1467-9884.2005.tb02031.x.

Benjamin, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, 57, 289–300. https://doi.org/10.1111/j.1467-9884.2005.tb02031.x.

Brinon, J. R. (2015). Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research. Computers & Education, 87, 218–237. https://doi.org/10.1016/j.compedu.2015.07.001.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students’ learning from text. Educational Psychology Review, 14, 5–26. https://doi.org/10.1023/A:1013769026026.

Chinn, A. C., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education, 86, 175–218. https://doi.org/10.1002/sce.10001.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

De Jong, T. (2019). Moving towards engaged learning in STEM domains; there is no simple answer, but clearly a road ahead. Journal of Computer Assisted Learning, 35, 153–167. https://doi.org/10.1111/jcal.12137.

De Jong, T., Linn, M. C., & Zacharias, Z. C. (2013). Physical and virtual laboratories in science and engineering education. Science, 340, 305–308. https://doi.org/10.1126/science.1230579.

Denessen, E., Vos, V., Hasselman, F., & Louwes, M. (2015). The relationship between primary school teacher and student attitudes towards science and technology. Education Research International. https://doi.org/10.1155/2015/534696.

Dunn, L. M., & Dunn, L. M. (1997). PPVT-III: Peabody Picture Vocabulary Test (3rd ed.). Circle Pines, MN: American Guidance Service.

Duit, R., & Nuerdinwoerden [DUO Educational R&D] (2017). De lesmotivatie van Nederlandse kinderen en jongeren [Reading motivation of Dutch children and adolescents]. Amsterdam, the Netherlands: Stichting Lezen.

Eberbach, C., & Crowley, K. (2009). From everyday scientific observation: How children learn to observe the biologist’s world. Review of Educational Research, 79, 39–68. https://doi.org/10.3102/0034654308328589.

Edelsbrunner, P. A., Schall, L., Schumacher, R., & Stern, E. (2015). Pathways of conceptual change: Investigating the influence of experimentation skills on conceptual knowledge development in early science education. Pasadena, CA: Cognitive Science Society.
Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction, 36*, 11–26. https://doi.org/10.1016/j.learninstruc.2014.11.002.

Scheiter, K., Schubert, C., & Schüler, A. (2018). Self-regulated learning form illustrated text: Eye movement modelling to support use and regulation of cognitive processes during learning from multimedia. *British Journal of Educational Psychology, 88*, 90–94. https://doi.org/10.1111/bjep.12175.

Schiefele, U., & Schaffner, E. (2016). Factorial and construct validity of a new instrument for the assessment of reading motivation. *Reading Research Quarterly, 51*, 221–237. https://doi.org/10.1002/rrq.134.

Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam, the Netherlands: Harcourt Assessment BV.

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Education Research Review, 23*, 1–24. https://doi.org/10.1016/j.edurev.2017.11.001.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology, 36*, 273–310. https://doi.org/10.1006/cogp.1998.0686.

Sikken, J. (2017). The electrical circuits lab. Retrieved from: https://www.golabz.eu/lab/electrical-circuit-lab.

Stender, A., Schichow, M., Zimmerman, C., & Härtig, H. (2018). Making inquiry-based science learning visible: The influence of CVS and cognitive skills on content knowledge learning in guided inquiry. *International Journal of Science Education, 40*, 1812–1831. https://doi.org/10.1080/09500693.2018.1504346.

Van Bon, W. H. J. (2006). *Doorstreepleestoets: Een groepsgewijs af te nemen toets voor de technische leesvaardigheid* [Strike-through reading test: A group-administered test for word decoding]. Leiden, the Netherlands: PITS.

Van de Sande, E., Verhoeven, L., Kleemans, T., & Segers, E. (2019). The linguistic nature of children’s scientific reasoning. *Learning and Instruction, 62*, 20–26. https://doi.org/10.1016/j.learninstruc.2019.02.002.

Van der Graaf, J., Gijsel, M., Van de Sande, E., & Segers, E. (2019). A combined approach to strengthen children’s scientific thinking: direct instruction on scientific reasoning and training of teacher’s verbal support. *International Journal of Science Education, 41*, 1119–1138. https://doi.org/10.1080/09500693.2019.1594442.

Van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: dynamic assessment of the control of variables strategy. *Instructional Science, 43*, 381–400. https://doi.org/10.1007/s11251-015-9344-y.

Van der Graaf, J., Segers, E., & Verhoeven, L. (2018). Individual differences in the development of scientific thinking in kindergarten. *Learning and Instruction, 56*, 1–9. https://doi.org/10.1016/j.learninstruc.2018.03.005.

Van der Meij, J., & de Jong, T. (2006). Supporting students’ learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*, 199–212. https://doi.org/10.1016/j.learninstruc.2006.03.007.

Van der Meij, J., & de Jong, T. (2011). The effects of directive self-explanation prompts to support active processing of multiple representations in a simulation-based learning environment. *Journal of Computer Assisted Learning, 27*, 411–423. https://doi.org/10.1111/j.1365-2729.2011.00411.x.

Van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 263–278). (2nd ed.). New York, NY: Cambridge University Press.

White, T., Graves, M. F., & Slater, W. H. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. *Journal of Educational Psychology, 82*, 281–290. https://doi.org/10.1037/0022-0663.82.2.281.