Graphical Representations of Data in STEM Education

Investigation of Graphing and Graph Comprehension

Dissertation zum Erwerb des Doctor of Philosophy (Ph.D.)
am Munich Center of the Learning Sciences
der Ludwig-Maximilians-Universität
München

Vorgelegt von

Verena Ruf
aus Lauingen (Donau)

München, den 8. November 2023

1st supervisor / Erstgutachter: Prof. Dr. Jochen Kuhn, LMU München

2nd supervisor / Zweitgutachter: PD Dr. Sarah Malone, Universität des Saarlandes

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Frank Fischer, LMU München

Prof. Dr. Martin Fischer, LMU Klinikum

Tag der mündlichen Prüfung: 26.03.2024

Acknowledgements

There are a lot of people I would like to thank for their support. First, my sincere gratitude to my supervisors. Thank you for being my sounding board during the entire process and helping me plan during the final stages. To my first supervisor Jochen Kuhn: thank you for your help in structuring my thesis, your inspiration for finding new fascinating areas of research, and imagining new ways of teaching. Another thank you belongs to my second supervisor Sarah Malone: thank you for putting things into perspective and for your insightful questions during our discussions that always made me think one step further. I learned a lot from both of you. Another thank you to the third member of my supervisory committee, Frank Fischer, for your comments regarding the articles and for your immediate commitment when I was looking for a third supervisor on such short notice. To Stefan Küchemann: thank you for helping me every step of the way, always being open to answering any questions, and being my daily supervisor. Furthermore, thank you to Martin Fischer, for your support during the writing process of the articles and for being the final member of my examination committee.

I would also like to thank every member of my group – both in Munich and in Kaiserslautern – for all their help. A special thank you to Eva Rexigel for your support during work and your friendship outside of it. I very much appreciate you listening to all my concerns and always offering a helpful solution. Thank you to Steffen Steinert for helping me with coding-related questions and suggesting new directions. I would also like to thank my colleagues in Munich: Tatjana Lamparter and Yavuz Dinc, the former, particularly for our writing retreats and the latter for not being annoyed with me for talking to my computer and bringing gingerbread cookies. Thank you to all the girls from the DTP: I haven’t known you long but you’ve been an inspiration.

Besides my colleagues, I would like to thank all my family and friends. Pia Adam, for all her advice and encouragement. A special thank you to my parents: for being there during the COVID-19 pandemic and for their continuous support. Thank you to my sisters, Miriam and Theresa for always being there to talk. And thank you to my grandparents, especially my grandmother, Sophie Stadter, for showing me that anything is possible and always having my back.

Executive Summary

Learning materials usually consist of various types of representations. For example, graphical representations, such as illustrations or graphs, are often used in instructions in combination with text. Graphical representations of data are a subgroup of graphical representation that is common not only in education but also in news media. These types of representations depict data and can be informative to learners when presenting them solely or in addition to text. Dealing with such information is a key skill of the 21st century and has been frequently researched. Skills dealing with graphs can be summarised under the term *graphing competence,* describing the creation (graphing) and the comprehension of graphs. However, graphing competence is not an easy skill for students to learn and students' difficulties are frequently reported. This thesis presents research that aims to contribute to previous findings regarding graphing competence, thereby enhancing the use of graphs as an educational tool. Both aspects of graphing competence – graphing and graph comprehension – are addressed in this thesis. The first research direction concerns graphing; how graphing is investigated, what benefits it has, and the types of difficulties students have during graphing. The second research direction addresses the second aspect of graphing competence: graph comprehension. Graph comprehension skills change with varying levels of expertise. Expertise differences can be analysed using eye movements as indicators of cognitive processing.

Therefore, this thesis analyses eye movements during learning and problem-solving with graphs, specifically paying attention to the differences between the visual processing of experts and non-experts. Furthermore, differences in graph comprehension between various study disciplines are examined. Based on current empirical research, physics students can be considered experts compared to students of other disciplines, because they seem to solve graph comprehension tasks better, independently of the task context. Building on previous research, the visual behaviour of physics and non-physics students is studied. Extending previous research, machine-learning methods are used to predict correct and incorrect solvers based on their eye movements.

The three research directions are addressed in the three studies presented in this thesis. The first study describes a systematic literature review of the empirical research on graphing in K-12 science, technology, engineering, and maths (STEM) education. The second study reviews the literature comparing experts' and non-experts' visual processing

during learning and problem-solving with graphs. The third study investigates the differences in learning gain and visual behaviour between physics and non-physics students solving graph comprehension tasks.

The first study narratively summarises how graphing is implemented in studies researching graphing in K-12 education. Furthermore, information on the added value of graphing and students' difficulties during graphing are considered. Fourty-four studies investigated this topic published from 1979 until March 2022, when the search was conducted. Many studies instructed the graphing of line graphs over more than one lesson. The synthesis of the study results indicates that different types of graphing instruction have a positive effect not only on graphing skills but also on graph comprehension. However, the review findings indicate that students have difficulties both with the graphing conventions as well as with the theoretical implications of the data depicted in the graph. As theoretical difficulties are also common in graph comprehension, this indicates that both types of difficulties influence graphing skills. Furthermore, the two aspects of graphing competence – graphing and graph comprehension – might affect each other.

The second study presents a literature review of studies comparing the visual processing of experts and non-experts during learning and problem-solving with graphs. Thirty-two studies published between 2003 and 2022 were analysed regarding the eye-tracking metrics used to investigate visual behaviour and the reported differences between experts and non-experts. Most studies used more than one eye-tracking metric. The findings indicate that experts pay more attention to relevant areas of the graph than non-experts. This is in line with the information-reduction hypothesis, suggesting that experts can ignore irrelevant information on a perceptual level. Definitions of expertise vary, implying that an overarching definition of expertise is missing. However, over the course of this review, four possibly relevant factors for expertise in graph comprehension were identified: (1) graphical literacy, (2) domain knowledge, (3) prior mathematical knowledge, and (4) task knowledge.

The third study empirically investigates differences in learning gain and visual behaviour of physics and medical or veterinary students. Twelve physics and twelve non-physics students, respectively, voluntarily solved 24 graph comprehension tasks in the contexts of math, physics, and medicine at the beginning and the end of their first semester. There were no statistically significant differences in learning gain between groups. This might indicate similar transfer skills between these study disciplines as both participant groups took STEM courses. Correct and incorrect solvers could be predicted via machine learning based on their eye movements. Therefore, machine learning optimised for small datasets can be a valuable tool for assessing expertise by analysing eye movements.

The research presented in this thesis supports the relevance of instructing graphing competence. Both aspects of graphing competence, graphing and graph comprehension, should be considered during teaching. In particular, graphing instruction could be beneficial for students because it does not only seem to facilitate graphing skills but also graph comprehension. Furthermore, graphing instruction seems relatively easy to implement as the findings indicated that it was advantageous in various forms. However, students had difficulties during graphing. Student difficulties based on graphing conventions or based on theoretical aspects, such as with interpretation, were reported in many studies, indicating that both types of student difficulties should be considered during instruction. Furthermore, future research should consider the visual behaviour of K-12 students and experts during graphing because eye movements can indicate expertise in processing graphs.

During learning and problem-solving with graphs, a comparison of the visual processing of experts and non-experts supports the information-reduction hypothesis. This indicates that experts can ignore irrelevant information on a perceptual level and process information more efficiently than non-experts. Showing students experts' strategies might, therefore, be beneficial for them by guiding their focus to relevant information. Future research should consider levels of expertise based on measurable factors due to the diverse possibilities in which graph comprehension might be facilitated. For example, STEM instruction could promote the transfer of problem-solving skills, such as graph comprehension, to other domains. In summary, the results of this thesis highlight influencing factors for graphing competence, both graphing and graph comprehension, not only in K-12 but also in higher education.

Deutsche Zusammenfassung

Lernmaterial enthält normalerweise verschiedene Arten von Repräsentationen. Beispielsweise werden in Lernmaterialien neben Text oft graphische Repräsentationen, wie Abbildungen oder Graphen, verwendet. Eine Untergruppe graphischer Repräsentation sind graphische Repräsentationen von Daten; sie sind nicht nur in der Bildung sondern auch in den Nachrichten verbreitet. Repräsentationen von Daten können für Lernende sowohl alleine, als auch in Kombination mit Text, informativ sein. Der Umgang mit solchen Repräsentationen ist eine Schlüsselkompetenz des 21. Jahrhunderts und wurde oft untersucht. Die Fähigkeit, mit Graphen umzugehen, kann unter dem Begriff *Graphing-Kompetenz* zusammengefasst werden. Dieser beschreibt nicht nur die Fähigkeit Graphen zu erstellen (Graphing), sondern auch die Fähigkeit Graphen zu verstehen. Graphing-Kompetenz ist allerdings keine leicht erlernbare Fähigkeit und über Schwierigkeiten damit wird häufig berichtet. Diese Dissertation präsentiert Forschung, die bisherige Erkenntnisse zu Graphing-Kompetenz erweitern und die Verwendung von Graphen als pädagogisches Hilfsmittel in der Lehre verbessern möchte. Beide Aspekte von Graphing-Kompetenz – Graphing und Graphenverständnis – werden in dieser Dissertation adressiert. Die erste Forschungsrichtung ist die Untersuchung von Graphing, wie Graphing untersucht wurde, welche Vorteile es hat, und die Arten von Schwierigkeiten von SchülerInnen während des Erstellens von Graphen. Die zweite Forschungsrichtung adressiert den zweiten Aspekt von Graphing-Kompetenz: das Graphenverständnis. Die Fähigkeit Graphen zu verstehen ändert sich je nach Expertise. Visuelles Verhalten kann anhand von Augenbewegungen untersucht werden und unterschiedliche Level von Expertise können mittels kognitiver Prozesse analysiert werden.

Deswegen befasst sich diese Dissertation mit der Analyse von Augenbewegungen während des Lernens und Problemlösens mit Graphen, mit einem Fokus auf den Unterschieden in visuellem Verhalten von ExpertInnen und Nicht-ExpertInnen. Außerdem werden Unterschiede im Graphenverständnis zwischen verschiedenen Studienfächern untersucht. Physikstudierende werden im Vergleich zu Studierenden anderer Fachrichtungen in aktuellen Studien als ExpertInnen betrachtet, weil sie Aufgaben zum Graphenverständnis besser lösen können, unabhängig vom Kontext der Aufgabe. Basierend auf bisheriger Forschung wird das visuelle Verhalten von Physik- und Nicht-Physik-Studierenden untersucht. Die bisherige Forschung wird durch eine

Analyse mittels maschinellem Lernen erweitert, durch die korrekte und inkorrekte LöserInnen anhand ihrer Augenbewegungen prädiziert werden.

Die drei Forschungsrichtungen werden in den drei Studien dieser Dissertation adressiert. Die erste Studie beschreibt ein systematisches Literatur-Review über die Forschung zu Graphing in der schulischen Mathematik, Informatik, Naturwissenschaften und Technik (MINT) Bildung. Die zweite Studie ist eine Übersicht über Literatur, die das visuelle Verhalten von ExpertInnen und Nicht-ExpertInnen während des Lernens und Problemlösens mit Graphen vergleicht. Die dritte Studie untersucht Unterschiede im Lernfortschritt und visuellen Verhalten zwischen Physik und Studierenden anderer Fachrichtungen beim Lösen von Aufgaben zum Graphenverständnis.

Die erste Studie ist ein systematisches Review über die Implementation von Graphing in der schulischen Bildung in Studien zu diesem Thema. Außerdem werden Informationen über den Wert von Graphing und Schwierigkeiten von SchülerInnen dabei berücksichtigt. Vierundvierzig Studien wurden zwischen 1979 und der Literatursuche im März 2022 zu diesem Thema veröffentlicht. Bei vielen dieser Studien wurde das Erstellen von Liniengraphen während einer Instruktion über mehrere Unterrichtsstunden hinweg untersucht. Die Synthese der Ergebnisse deutet darauf hin, dass verschiedene Arten von Instruktionen von Graphing nicht nur einen positiven Effekt auf die Fähigkeit, Graphen zu erstellen, sondern auch auf das Graphenverständnis haben. Allerdings zeigen die Ergebnisse des Reviews, dass SchülerInnen sowohl Schwierigkeiten mit den gebräuchlichen Konventionen für das Graphing als auch mit der theoretischen Bedeutung der im Graphen gezeigten Daten haben. Da Schwierigkeiten mit der theoretischen Bedeutung auch beim Graphenverständnis vorkommen, könnten beide Arten von Schwierigkeiten einen Einfluss auf die Graphing-Fähigkeiten haben. Außerdem könnten die beiden Aspekte von Graphing-Kompetenz – Graphing und Graphenverständnis – einander beeinflussen.

Die zweite Studie präsentiert eine Literaturrecherche von Studien, die das visuelle Verhalten von ExpertInnen und Nicht-ExpertInnen während des Lernens und Problemlösens mit Graphen vergleichen. Zweiundreißig Studien wurden zwischen 2003 und 2022 publiziert und in dieser Arbeit anhand ihrer Eye-Tracking Metriken untersucht und Unterschiede im visuellen Verhalten von ExpertInnen und Nicht-ExpertInnen wurden verglichen. Die meisten Studien haben mehrere Eye-Tracking Metriken verwendet. Die Ergebnisse zeigen, dass ExpertInnen länger auf relevante Bereiche von Graphen fokussieren als Nicht-ExpertInnen. Dies unterstützt die

Informations-Reduktions-Hypothese, was darauf hinweist, dass ExpertInnen unwichtige Informationen auf einer wahrnehmungsbezogenen Ebene ignorieren können. Expertise wurde in den Studien auf unterschiedliche Arten ermittelt, was darauf hindeutet, dass es keine übergreifende Definition von Expertise gibt. Allerdings wurden im Laufe des Reviews vier Faktoren als mögliche Indikatoren von Graphenverständnis identifiziert: (1) Lese- und Schreibkompetenz für Graphen, (2) Domänenwissen, (3) mathematisches Vorwissen und (3) Wissen über die Aufgabe.

Die dritte Studie untersucht empirisch Unterschiede im Lernzuwachs und im visuellen Verhalten von Physik und (Tier-)Medizin Studierenden. Je zwölf Physik und zwölf Nicht-Physik Studierende haben freiwillig insgesamt 24 Aufgaben zum Graphenverständnis in Mathe, Physik und Medizin am Anfang und am Ende ihres ersten Semesters beantwortet. Es wurden keine statistisch signifikanten Unterschiede im Lernzuwachs zwischen den Gruppen gefunden. Dies könnte daran liegen, dass Studierende aller Disziplinen MINT-Kurse belegt haben und die Probandengruppen daher ähnliche Transfer-Fähigkeiten zwischen den Kontexten entwickeln konnten. Korrekte und inkorrekte Löser konnten anhand ihres visuellen Verhaltens mit maschinellem Lernen vorhergesagt werden. Dies zeigt, dass maschinelles Lernen mittels eines für kleine Datensätze optimierter Algorithmus ein gutes Werkzeug zur Auswertung von Expertise mittels einer Analyse von Augenbewegungen sein kann.

Die in dieser Dissertation präsentierte Forschung betont die Relevanz, Lernenden Graphing-Kompetenz zu vermitteln. Beide Aspekte von Graphing-Kompetenz, Graphing und Graphenverständnis, sollten dabei während des Unterrichts berücksichtigt werden. Besonders eine Instruktion von Graphing kann dabei für SchülerInnen hilfreich sein, weil dadurch nicht nur die Fähigkeit Graphen zu erstellen gefördert wird, sondern auch das Graphenverständnis. Außerdem scheinen verschiedene Arten von Graphing-Instruktionen lernförderlich zu sein, was darauf hindeutet, dass eine solche Anleitung einfach zu implementieren ist. Allerdings hatten SchülerInnen Schwierigkeiten beim Erstellen von Graphen. Ihre Schwierigkeiten basierten entweder auf den der Graphenerstellung zugrundeliegenden Konventionen oder waren in theoretischen Aspekten begründet, beispielsweise Schwierigkeiten bei der Interpretation. Diese Schwierigkeiten wurden in vielen Studien berichtet, was darauf hindeutet, dass die Schwierigkeiten während der Instruktion berücksichtigt werden sollten. Zukünftige Forschung sollte außerdem das visuelle Verhalten von SchülerInnen und ExpertInnen während des Erstellens von

Graphen analysieren, da Augenbewegungen ein Indikator für Expertise während der Verarbeitung von Graphen sein können.

Ein Vergleich des visuellen Verhaltens von ExpertInnen und Nicht-ExpertInnen während des Lernens und Problemlösens mit Graphen hat die Informations-Reduktions-Hypothese gestützt. Dies deutet darauf hin, dass ExpertInnen unwichtige Informationen wahrnehmungsbezogen ignorieren können und Informationen dadurch effizienter als Nicht-ExpertInnen verarbeiten. Es könnte daher hilfreich für Lernende sein, wenn sie Strategien von ExpertInnen sehen, um auf relevante Informationen zu achten. Zukünftige Forschung sollte verschiedene Level von Expertise anhand messbarer Faktoren berücksichtigen, da es verschiedene Möglichkeiten zur Förderung von Graphenverständnis gibt. Beispielsweise könnte MINT-Instruktion womöglich hilfreich sein, um Übertragungvon Problemlöse-Fähigkeiten, wie beispielsweise Graphenverständnis, zwischen Domänen zu lernen. Zusammenfassend heben die Ergebnisse dieser Dissertation Einflussfaktoren für Graphing-Kompetenz, für Graphing und für Graphenverständnis, nicht nur für die schulischen Bildung, sondern auch für die höhere Bildung, hervor.

x

Table of Contents

1. General Introduction

1.1. Aims and Outline of the Thesis

Graphical representations, such as images or pictures, are common in many media. For example, graphical representations of numerical data (graphs) are used in countless aspects of everyday life, such as newspapers. They are crucial tools for conveying information (Mahmoud & Zoghaib, 2023). Especially during the COVID-19 pandemic, graphs were widespread, for example, to visualise the number of COVID-19 cases (Engledowl & Weiland, 2021). Representations of data are also used in scientific publications (Midway, 2020). They can, for example, to show the relation of variables (Lachmayer et al., 2007). Graphs are often easier to interpret than other types of data representations, such as verbal descriptions, because they can show relations more explicitly (Larkin & Simon, 1987). Therefore, graphs are well-suited for educational purposes (Shah & Hoeffner, 2002). However, not all types of graphs are equally appropriate for conveying information (Mahmoud & Zoghaib, 2023). Consequently, the ability to judge the quality of a graph depending on its context is a valuable one (Rubel et al., 2021). Furthermore, using information is a key skill of the 21st century (Program for International Student Assessment, 2022). Using information includes correctly employing mathematical representations, such as graphs, extracting information from them, and interpreting the results correctly (Program for International Student Assessment, 2022). Therefore, an essential aspect of education is teaching students how to use graphs (Glazer, 2011) and graphs are accepted tools in education (e.g., Shah & Hoeffner, 2002), especially in science, technology, engineering, and mathematics (STEM) subjects (Fyfe et al., 2014).

This thesis describes research on how graphs are used in educational STEM practice. The main foci are (a) analysing graph creation as an educational tool, (b) investigating expertise differences in visual processing during learning and problem-solving with graphs, and (c) exploring differences between study disciplines during problem-solving with graphs.

Students can use graphs in various ways during learning and problem-solving (Leinhardt et al., 1990). Graph comprehension skills are necessary to successfully use graphs (Shah, 1997). Many studies analyse how to facilitate students' graph comprehension skills (e.g., Strobel et al., 2019) and although the terms used by the authors can vary – such as graph interpretation (e.g., Smit et al., 2016), graph understanding (e.g., Klein et al., 2020), or graph reading (Ludewig et al., 2020) – the described skills are all essentially similar. Graph creation is another aspect of graph interpretation (Glazer, 2011). There is much previous research on graph comprehension skills, including reviews (Shah & Hoeffner, 2002), case studies (Roth & Bowen, 2001), and comparisons between countries (Galesic & Garcia-Retamero, 2011). Although there are reviews about drawing in education (Cromley et al., 2020; Van Meter & Garner, 2005; Y. Zhang et al., 2021), a review about how the creation of graphs is implemented in research has not been published so far. Such a review could provide much insight into possible overarching benefits for students and difficulties they might have (aim a).

Students' difficulties can also be identified by looking at their visual processing indicated by eye movements. There have been differences in the visual processing of experts and non-experts when looking at graphs. For example, science and non-science undergraduate students, science graduate students, and science faculty reported similar strategies to solve graph-based tasks (Harsh et al., 2019). However, only faculty and graduate students seemed to implement their plan whereas the undergraduate students' procedures varied (Harsh et al., 2019). In physics, the answer correctness of the participants could be predicted based on their eye movements (Küchemann et al., 2020, 2021). Paying attention to the task-irrelevant parts of the graph might be due to misconceptions (Klein, Küchemann, et al., 2019; Wang et al., 2022). Misconceptions are wrong beliefs due to prior knowledge or intuition (Leinhardt et al., 1990). Misconceptions about graphs are often based on a misunderstanding of previous instruction (Leinhardt et al., 1990) and using incorrect strategies can lead to misinterpretations of the depicted data (Clement, 1989). Student understanding can improve if students overcome misconceptions and learn the right concepts (Marisda et al., 2020), although this may be difficult (Lem et al., 2013). Differences in visual processing between experts and non-experts when learning and problem-solving with graphs could help identify problems. A review of this topic considering various types of graphs could fill this gap (aim b).

For line graphs, previous research has shown that physics students outperform psychology (Susac et al., 2018) and economics (Brückner et al., 2020; Klein, Küchemann, et al., 2019) students in problem-solving tasks independently of the task context. The main difference between physics and the other subjects is that physics students have physics courses. Therefore, expertise differences between physics students and other students also taking STEM courses should be smaller than between students of other subjects. A comparison of physics and medical students could provide valuable insights (aim c).

The thesis is structured in three parts. The first part describes the theories of learning with graphical representations (see section 1.2). Theories about learning with multimedia, such as the cognitive theory of multimedia learning (Mayer, 2014a), are introduced. These theories apply to theories of how learners can generate graphical representations themselves, which is, for example, useful to externalise information (Schmidgall et al., 2019). This practice is common in many subjects, such as using graphs in STEM problem-solving (Zacks & Tversky, 1999). Theories of multimedia learning relate to cognitive processes (see section 1.3) which can be analysed via visual processing as indicated by eye movements (Alemdag & Cagiltay, 2018). For example, previous research has found that experts' and non-experts' visual processing differs in visualisation comprehension (e.g., Gegenfurtner et al., 2011). Based on previous research, the research aims of the articles included in this thesis as well as a short overview of every study are presented (see section 1.4).

The second part of this thesis describes the two literature reviews and the empirical study conducted to address the research aims. First, a systematic literature review about generating graphical representations of numerical data in STEM education is presented (see section 2). The second article is a literature review about differences in the visual processing of experts and non-experts when learning and problem-solving with graphs (see section 3). Third, an article describing an empirical study comparing physics and medical students' visual behaviour during problem-solving with graphs is included (see section 4).

The last section of this thesis summarises the results of the three studies (see section 5). The findings are assimilated with the theory as well as the educational practice. The thesis ends with a conclusion after considering the limitations and future research.

1.2. Graphical Representations in Education

There is various research about the theoretical background for learning and problem-solving with graphical representations. This section first presents general theories for learning with more than one representations, such as graphs and text, for example, the cognitive theory of multimedia learning (CTML) (Mayer, 2014a). Learners cannot only learn by learning with provided graphical representations but also by generating them (aim a). This aspect is elaborated in the following before describing learning with graphs, specifically generating (aim a) and comprehending (aims b and c) them, as defined by the term *graphing competence*.

1.2.1. Theories of Learning with Multimedia

Presenting multiple types of representations in combination is considered multimedia learning. During instruction, graphical representations are usually presented in combination with other types of representations, such as text (Mayer, 2014c). In STEM education, graphs are typically presented with text or equations (e.g., Díaz-Levicoy et al., 2018). Learning with multimedia describes the construction of a mental representation based on learning material consisting of words, such as text, and pictures, such as illustrations (Mayer, 2014b). The goal of learning is to build a mental model of the represented information (Mayer, 2021). The multimedia assumption expects that learning with multimedia, such as words and pictures, is more effective than learning only from text (Butcher, 2014). There are several theories about learning and problem-solving with multiple representations (Ayres, 2015), that also apply to learning in STEM education.

The cognitive theory of multimedia learning (CTML) (Mayer, 2014a, 2021) is based on three assumptions: that the auditorial or verbal and the visual or pictorial information-processing channels process information separately (Baddeley, 2012; Camp et al., 2021; Paivio, 1969), that each channel has a limited capacity (Camp et al., 2021; Chandler & Sweller, 1991), and that processing information, as is necessary for learning, is an active process (Wittrock, 1974). Learners need to select relevant information, organise the selected information, and integrate it with each other and with their prior knowledge in order to construct a mental model (Mayer, 2014a, 2021). Therefore, the goal of learning about topics in a particular domain is to transfer information from limited-capacity working memory to unlimited long-term memory. Cognitive load theory can be employed to facilitate this process by improving the instructional design (Sweller, 2020). According to the original theory (Sweller et al., 1998), cognitive load can be split into three separate categories (Orru & Longo, 2019; Paas et al., 2003; Sweller et al., 1998): Extraneous cognitive load is based on the way information is represented and is not related to the learning aim, intrinsic cognitive load is related to task difficulty and the amount of relevant information that has to be processed, and germane cognitive load is needed to understand the learning material and learn effectively. In a review, Kalyuga (2011) argued that germane and intrinsic cognitive load are related. Consequently, in current interpretations of cognitive load theory, germane cognitive load is not viewed as a separate type of cognitive load but as a working memory resource allocating capacity from extraneous to intrinsic processing (Sweller et al., 2019). Current reviews continue to differentiate between the three types of cognitive load (Mutlu-Bayraktar et al., 2019; Orru

& Longo, 2019; Skulmowski & Xu, 2022). Effective learning should use mostly intrinsic and germane cognitive load and little extraneous cognitive load (Orru & Longo, 2019), although a small amount of extraneous cognitive load related to creating an appealing learning environment might be beneficial (Skulmowski & Xu, 2022). Cognitive load, therefore, plays an important role in learning environments containing multiple representations, such as text and pictures (Mutlu-Bayraktar et al., 2019).

Similarly to the CTML, the integrated model of text and picture comprehension (ITPC) assumes different ways of processing text and pictures (Schnotz, 2014). First, learners analyse textual and pictorial features and then process them on a deeper cognitive level (Schnotz, 2014). The resulting propositional representation of the text and the mental model of the pictures are integrated during information processing (Schnotz, 2014). In this context, graphs would be considered pictorial information. The different functions of representations in this process constitute the main difference between the CTML and ITPC (Ayres, 2015).

Principles of multimedia learning have been formulated based on the CTML and the effects of cognitive load, which can be used for designing learning material (Mayer, 2014c). One example is the signalling (or cueing) principle (Mayer, 2014c). This principle uses signals (cues) to draw learners' attention to relevant information or highlight the organisation of important information (van Gog, 2014, 2021) and can facilitate learning (Alpizar et al., 2020) by supporting learners' cognitive processes, especially selecting information (van Gog, 2021). Various multimedia learning principles have been adopted in multiple STEM contexts (Herrlinger et al., 2017; Klein, Viiri, et al., 2019; Rodemer et al., 2021; Ruf et al., 2022). These principles can be transferred to other educational situations, for example, to computer-based testing environments (Dirkx et al., 2021). It is therefore important to consider multimedia principles not only during instruction but also in the context of a testing environment (Lindner et al., 2021) because cognitive processes during problem-solving happening in testing are similar to those found during multimedia learning (Lindner et al., 2017).

Besides the principles of multimedia learning, the design, function and task framework (DeFT) can be used to construct supportive learning material containing various representations (Ainsworth, 2006). This framework is based on the functions multiple representations fulfil during learning as well as relevant tasks and design considerations (Ainsworth, 1999). Representations can have three main functions during learning: (1) they can complement each other, for example, by providing different information, e.g., in real and virtual experiments (Flegr et al., 2023); (2) they can constrain each other, such as one familiar representation constraining an unfamiliar representation, which, e.g., is the case in concreteness fading where learners first see concrete pictures of a situation that become more abstract during the learning process (Kokkonen & Schalk, 2021); and (3) representations can construct deeper understanding, e.g., during constructive learning activities (Chi & Wylie, 2014). This is the case when students generate graphical representations, such as graphs.

1.2.2. Generating Graphical Representations

Generating information during the learning process can be more useful for learners than merely reading the material (Bertsch et al., 2007). This is called the generation effect (Slamucka & Fevreiski, 1983). The generation effect is related to the generative theory (Wittrock, 1974), which theoretically grounds the effectiveness of active processing in the CTML (Mayer, 2014a). Creating graphical representations, such as diagrams or graphs, is termed, for example, *drawing*, *sketching*, or *graphing*. Drawing can facilitate generative learning (Fiorella & Mayer, 2016). Thus, actively presenting information can promote learning more effectively than passively committing information to memory (Chi & Wylie, 2014; Fiorella & Zhang, 2018). This is in line with the interactive, constructive, active, passive (ICAP) framework positing that active learning is better than passive learning, which in turn is superseded by constructive followed by interactive learning (Chi & Wylie, 2014). With each level, learning is assumed to be more effective because learner engagement increases.

Furthermore, creating a new representation from a provided representation can potentially fulfil each of the functions described in the DeFT framework (Ainsworth, 2006): The representations can complement each other as the generated representation is based on the provided representation, they can constrain each other, for example, the generated representation can clarify a process described in a text, and generation can create deeper understanding because learners have to actively engage with the material. Generating representations based on provided representations involves more than one type of representation (e.g., the graph and text) and is consequently part of learning with multimedia. Generation is also connected to the cognitive processes of selecting, organising, and integrating information (Fiorella & Mayer, 2016; Mayer, 2014a; Van Meter & Garner, 2005) because the information presented in additionally generated representations has to be selected from the provided representation and the selected information has to be organised into a fitting format for the generated representation. Generating representations aids in prior knowledge activation (Wetzels et al., 2010), because learners have to integrate their prior knowledge with the provided information during the construction process.

The generative sense-making framework (Fiorella, 2023) focuses on how internal and external representations interact to facilitate sense-making processes. It follows similar assumptions as the ICAP framework (Chi & Wylie, 2014) and those based on cognitive processes (Fiorella & Mayer, 2016; Van Meter & Garner, 2005). This framework describes how sense-making and learner characteristics as well as the generated representations influence each other. The generated representations shape the instruction, which in turn can affect sense-making. Learning outcomes depend on the success of sense-making processes. Generative learning activities consist of explaining, visualising, and enacting (Fiorella & Mayer, 2016, 2021). In the generative sense-making framework, each activity has a function, for example, visualising can help organise knowledge (Fiorella, 2023). The generative sense-making framework, therefore, specifies constructive activities in the ICAP framework (Chi & Wylie, 2014).

Constructing representations, such as graphs, can have many advantages (Ainsworth et al., 2011; Ainsworth & Scheiter, 2021). By generating representations learners can represent visuo-spatial information (Scheiter et al., 2017) as well as make inferences visible (Larkin & Simon, 1987). Generating representations also improves self-regulation (Kollmer et al., 2020). For example, drawing can encourage learners to provide detailed explanations (Fiorella & Kuhlmann, 2020). Furthermore, learners use the generated representations as visualisations of information and to externalise information (Schmidgall et al., 2019).

Generation can be implemented in teaching in various ways, for example, by generating a representation before comparing it to a provided representation (Q. Zhang & Fiorella, 2021). Instructional support for the generation of representations can further facilitate learning (Cromley et al., 2020; Fiorella & Zhang, 2018; Van Meter et al., 2006; Wu & Rau, 2019). However, the generation of new representations takes up cognitive resources (Schwamborn et al., 2011) and has to be implemented carefully (Fiorella & Zhang, 2018). For example, drawing seems to be especially helpful for older students in secondary or higher education (Brod, 2021; Y. Zhang et al., 2021) and for undergraduate students with low prior knowledge (Lin et al., 2017). Older learners seem to have the necessary cognitive skills to deal with generation tasks (Brod, 2021) but should not have so much prior knowledge that generating representations would be redundant (Lin et al., 2017).

In secondary education, drawing has been researched in multiple ways: Drawing tools have been developed to compare the representational competence of high school students in chemistry and biology education (Chang, 2018) and the sketching of functions in mathematics by pre-university students has been analysed to gain insights into what mathematical connections students make during these tasks (García-García & Dolores-Flores, 2021). Generation activities are also used at universities, for example, sketching tasks as part of a quantum education curriculum (Kohnle et al., 2020) and drawing of best-fit lines has been analysed during physics lab activities (Nixon et al., 2016).

Technological tools can support generating graphical representations (Cromley et al., 2020; Donnelly-Hermosillo et al., 2020). For example, graphing calculators have been common tools in the last decades (Kastberg & Leatham, 2005; Penglase & Arnold, 1996). Technological tools also include digital drawing tools, such as GraphSmarts, in comparisons of paper-and-pencil with technology-based drawing (Gardner et al., 2021) or providing drawing prompts, for example, in interactive chemistry tutorials (Wu & Rau, 2018). Computer software, such as Excel (Åberg-Bengtsson, 2006), has also been used to research graph generation.

1.2.3. Graphing Competence

Generating graphs is a part of graphing competence (Glazer, 2011). Students learn and generate representations in many subjects. In STEM disciplines, graphical representations often depict numerical data (graphs). *Graphing* is, therefore, defined as generating convention-based graphs. Learning with graphs is theoretically based on learning with multimedia (see section 1.2.1) and generating graphs is also grounded in theories about generating graphical representation (see section 1.2.3). Apart from graphing, graphing competence includes the ability to analyse graphs (Glazer, 2011).

The ability to analyse graphs is often referred to as either "graph comprehension" (Curcio, 1987; Kanzaki & Miwa, 2011; Zacks & Tversky, 1999) or "graph interpretation" (Boels et al., 2019; Gültepe, 2016; Lachmayer et al., 2007; Nixon et al., 2016; Roth & Bowen, 2001). Using information – a key skill of the 21st century (Program for International Student Assessment, 2022) – is an integral part of graph comprehension. Various factors are important for graph comprehension (Friel et al., 2001): the tasks in which graphs are used as well as their purpose, the discipline/ context of the graphs, and the characteristics of the learner. This is an important topic in education and has been analysed in various ways in multiple reviews (Boels et al., 2019; Glazer, 2011; Leinhardt et al., 1990; Shah & Hoeffner, 2002). In their review, Leinhardt et al. (1990) focus on functions in the context of mathematics education with a special view on students' misconceptions and difficulties with graphs. Misconceptions are defined as a "reasonably well-formulated system of ideas" (Leinhardt et al., 1990, p. 5) consisting of explicit pieces of knowledge. Misconceptions can be related to difficulties but do not necessarily cause them. Leinhard et al. (1990) distinguish three types of tasks which can include the construction of a graph: prediction tasks concerning the data pattern, translation tasks between types of representations, and scaling tasks involving the scales and units of the depicted data. Besides constructions, all types of tasks include interpretation. Another review addresses the instructional implications of students' graph comprehension (Shah & Hoeffner, 2002). They analyse three factors that can influence interpretations: visual characteristics of the graphs, students' prior knowledge about graphs, and "expectations about the content of the data in a graph" (Shah & Hoeffner, 2002, p. 47). They recommend four aspects for teaching graphical literacy: (1) Teaching graphical literacy in a specific context, (2) using translation tasks, (3) focusing on linking the visual features with the meaning in the context, and (4) viewing graph comprehension not as simple fact retrieval but as an evaluation activity. Other reviews focus on challenges (Glazer, 2011) or misconceptions (Clement, 1985) with graph comprehension, sometimes examining only specific types of graphs, such as histograms (Boels et al., 2019). Most of these reviews refer in various ways to generating graphs as well as analysing them (Boels et al., 2019; Clement, 1985; Leinhardt et al., 1990). However, their focus is not exclusively on graphing although researchers recommend that the creation of graphs "should be explicitly taught given its importance and its complexity" (Glazer, 2011, p. 183).

Therefore, constructing graphs is a common topic in research and many studies specifically analyse students' difficulties with graphing. Typical errors include confusing the slope and the height of a graph as well as interpreting the graph like a picture (Clement, 1985). Undergraduate students taking an introductory physics course seemed to have both of these difficulties in the context of kinematics (McDermott et al., 1987). This indicates that students have trouble connecting the graph to the underlying concept (McDermott et al., 1987). A study with 32 undergraduate students taking an introductory physics lab course, also found that students had trouble connecting the data to the physics concept and often used “rote procedures” (Nixon et al., 2016, p. 11) instead of strategies involving a deeper understanding during lab activities including the construction of best-fit lines. Other errors that students have made in the context of graph creation are related to scaling (von Kotzebue et al., 2015) or finding the best type of graph for given data (Ozmen et al., 2020). These difficulties are comparable between contexts. For example, Dewi et al. (2018) and Gultepe and Kilic (2015) reported corresponding graphing difficulties in physics and chemistry, respectively. They are also similar among learners of various ages: Scaling difficulties were found in a study with 437 university science students solving problems in a biology context (von Kotzebue et al., 2015) as well as in a study with 40 elementary school students in the context of math (Åberg-Bengtsson, 2006). These examples indicate that students' difficulties with generating graphs can be persistent across contexts and different learning levels, although students improve with higher grades (Wavering, 1985).

This demonstrates the prevalence of students’ difficulties during the generation of graphical representation. One possible reason could be that students have difficulties processing such tasks. The CTML posits three relevant cognitive processes performed during learning with multiple representations: selection, organisation, and integration (Mayer, 2014a; see also section 1.2.1). These cognitive processes can be analysed by looking at learners’ eye movements (Alemdag & Cagiltay, 2018).

1.3. Cognitive Processes and Eye Movements During Learning with Graphical Representations

Learners’ eye movements can indicate their cognitive processes and can be employed to analyse multimedia learning processes (see section 1.3.1). Using eye tracking as an investigative technique for recording eye movements, previous research has found differences between the eye movements of experts and non-experts when learning or problem-solving with graphs (see section 1.3.2). Eye movements have also been used to predict performance, for example in graph comprehension tasks (see section 1.3.3).

1.3.1. Eye Movements and Cognitive Processes

Eye movements are useful process measures because they can indicate attention (Just & Carpenter, 1980). This is called the eye-mind hypothesis. Although this hypothesis is based on reading research, its assumptions also hold in other circumstances (Schindler & Lilienthal, 2019), and eye tracking is often used as a method to analyse visual attention during learning processes (Alemdag & Cagiltay, 2018; Hahn & Klein, 2022; Lai et al., 2013; Strohmaier et al., 2020). Eye movements are either recorded via stationary eye trackers, which are used during computer-based studies (Strohmaier et al., 2020), or via mobile eye-tracking glasses, which are used, for example, during experimenting (Kumari et al., 2021). Mobile eye-tracking is also relevant in other learning environments, for example, in learning applications in augmented reality (Fleischer et al., 2023).

Various eye-tracking metrics can be used to analyse eye movements. Interesting areas of the stimulus are called areas of interest (AOIs) and they are the basis for calculating various eye-tracking metrics, such as fixations, for example, as sums or averages (Holmqvist & Andersson, 2017). Among the most common metrics are longer stops at a location – called fixations – and small movements between fixations – called saccades (Holmqvist & Andersson, 2017; Salvucci & Goldberg, 2000). Fixation durations or fixation counts can indicate how much attention is spent on various areas. For example, Malone et al. (2020) compared single representations with heterogeneous and homogeneous multiple representations consisting of text, an equation, and a graphical representation to determine which type of representations were most beneficial for problem-solving. This was indicated by the fixations on the representations, which suggested that the graphical representation was the most helpful one in that specific task. It should be noted that fixation durations and fixation counts can be correlated (Atkins & McNeal, 2018). There are also eye-tracking metrics that are more dynamic and include information about how the focus of attention changes over time, such as saccades and transitions. For example, when comparing students learning with animation or with interactive feedback, the animated group was more likely to fixate after a short saccade (Hoyer & Girwidz, 2020). In contrast, the interactive feedback group was more likely to fixate after a long saccade indicating an influence of the group on visual behaviour (Hoyer & Girwidz, 2020). Combined with an increased performance, these results indicate that longer saccades are related to deeper processing. Other, more static, metrics are roughly based on fixations, such as dwell time, which includes the duration of the saccades and can include multiple fixations (Holmqvist & Andersson, 2017). Dwell time is used similarly to fixations, for example, to analyse the effectiveness of animating contextual elements in learning games, with longer dwell times indicating attention (Javora et al., 2021). Transitions describe gaze switches between certain AOIs (Holmqvist & Andersson, 2017). They can indicate whether students connect pieces of information, for example, by comparing the portion of transitions between different combinations of representations, such as text and graph (Bayri & Kurnaz, 2015). A large number of transitions suggests that information is being integrated with each other and this can be used, for example, for developing adaptive learning systems (Kennel, 2022).

Information processing is a common topic in educational eye-tracking studies (Lai et al., 2013). A literature review of 57 studies analysing animations in multimedia learning found that most studies investigated cognitive processes (Coskun & Cagiltay, 2022). In the context of multimedia learning (see section 1.2.1), eye movements can be indicators of cognitive processes, such as selection, organisation, and integration (Alemdag & Cagiltay, 2018). Researchers assume that the percentages of, for example, fixation durations or counts, can illustrate selection, whereas average and total fixation duration can signify organisation (Alemdag & Cagiltay, 2018; Coskun & Cagiltay, 2022). Percentages of fixation duration contain the distribution of attention over AOIs, indicating relative (proportional) visual attention (Ruf et al., 2022; van Meeuwen et al., 2014). This could imply that one area is perceived as more important than another to the viewer. For example, middle-school students who solved a physics task correctly paid more relative attention to relevant areas than those who solved the task incorrectly (Wang et al., 2022). The number of transitions can indicate integration processes (Alemdag & Cagiltay, 2018; Coskun & Cagiltay, 2022).

Selection processes are important to find relevant information (Mayer, 2014a). They are used, for example, in eye-movement modelling examples by showing students the eye movements of an expert and drawing their attention to relevant information (Tunga & Cagiltay, 2023). In multiple-choice questions, this includes selecting the correct answer, which students tend to fixate longer (Tsai et al., 2012). The selection process can, for example, be facilitated by using the singalling principle (van Gog, 2021).

The most commonly examined processes are organisational processes (Coskun & Cagiltay, 2022). Organisation is needed to coherently structure the selected information (Mayer, 2014a). Organisation can be fostered via various means, such as providing information about the structure of a video (Cojean & Jamet, 2022). When using the signalling principle a combination of text-based cues and reflection prompts as well as visual cues facilitated students' understanding (Zheng et al., 2023). This combination of text-based cues and reflection prompts seemed to foster reorganisation and integration of information and was especially helpful in a transfer test.

Integration processes are needed to integrate the organised information with prior knowledge and – in the case of multiple representations – integrate different representations, such as verbal and graphical ones, with each other (Mayer, 2014a). Similar to organisation, integration processes have also been investigated in the context of instructional videos, for example, the integration between textual (speech) and pictorial information (Schüler & Merkt, 2021). Results indicate that inconsistent information between speech and pictures influences the participants' visual behaviour but not the learning outcome. Few participants noticed the discrepancies and they generally recalled more pictorial than speech information, which could explain the equal learning outcomes. Integration between text and other related representations is also important outside of instructional videos. For example, in a study investigating fourth-graders' processing of a scientific text, those who made more integrative transitions between the text and the pictorial representation learned better than those who did not (Mason et al., 2013). There are ways to promote integration: For example, cues have proven effective in facilitating integration processes between the vector field representation and the equation and text also presented (Klein, Viiri, et al., 2019).

Cognitive processes can change with increasing expertise, for example, experts encode information in long-term working memory differently than non-experts (Ericsson & Kintsch, 1995). As cognitive processes can be distinguished via eye movements, it is possible to analyse expertise differences this way (Gegenfurtner et al., 2011).

1.3.2. Expertise Differences in Eye Movements

Expertise is a common topic in STEM education research, for example, the differences in visual processing between experts and non-experts are often investigated (Hahn & Klein, 2022). Performance differences between experts and non-experts are probably due to differences in the cognitive processes they execute when dealing with information (Ericsson & Kintsch, 1995; Guida et al., 2012). Differences in visual processing are discernible in the eye movements of experts and non-experts (Brams et al., 2019; Gegenfurtner et al., 2011). Several theories can be used to interpret such differences based on various eye-tracking metrics.

The information-reduction hypothesis states that experts can ignore information that is irrelevant to the task at a perceptual level (Haider & Frensch, 1999). Therefore, experts can more efficiently assign attentional resources to the relevant parts of a stimulus (Brams et al., 2019; Gegenfurtner et al., 2011). This type of behaviour can be learned via repetition (Haider & Frensch, 1999). Based on this hypothesis, more and longer fixations on task-relevant areas can indicate expertise compared to more fixations on task-irrelevant areas indicating inexperience (Gegenfurtner et al., 2011). The information-reduction hypothesis can be used in learning environments, for example, eye-movement modelling examples helping non-experts recognise relevant areas more quickly (Xie et al., 2021). It is supported by studies across various domains (Brams et al., 2019). For example, in a study comparing advanced chemistry students with novice second-semester chemistry students, Topczewski et al. (2017) found that for nuclear magnetic resonance spectroscopy items novice students fixated more on the distractors compared to expert students. In a study comparing participants with high and low graph literacy, participants with high graph literacy paid more attention to relevant information necessary for correctly interpreting the data (Okan et al., 2016). These findings support the information-reduction hypothesis. However, research findings did not support the information-reduction hypothesis in the domain of medicine (Brams et al., 2019). In medicine, the holistic model of image perception seems to better explain differences in eye movements between experts and non-experts (Brams et al., 2019).

The holistic model of image perception proposes that experts can process images globally (Kundel et al., 2007). This may be due to parafoveal processing (Sheridan & Reingold, 2017). Global perception of the image influences future search processes, making experts more efficient (Gegenfurtner et al., 2011). Consequently, experts are expected to fixate on relevant AOIs more quickly than non-experts (Brams et al., 2019; Gegenfurtner et al., 2011). As mentioned above, the holistic model of image perception is prevalent in research about medical expertise (Brams et al., 2019). However, a concept of professional vision with experts being better at distributing their attention has also been found in experienced teachers when assessing classroom situations (Huang et al., 2023) and in pilots (Lounis et al., 2021).

Another assumption is that experts efficiently encode and store information in their long-term working memory (Ericsson & Kintsch, 1995). Novices start grouping information in working memory and, with practice, these so-called chunks move to long-term memory (Guida et al., 2012), where experts can access them via retrieval cues (Ericsson & Kintsch, 1995). Chunks can contain more familiar than unfamiliar or nonsensical information (Simon, 1974). Consequently, researchers assume that experts can concentrate more information in chunks than novices (Maries & Singh, 2023). Novices have less experience and are less knowledgeable about the topic and can, therefore, hold less information in working memory (Guida et al., 2012). As a result, novices process information less efficiently than experts (Ericsson & Kintsch, 1995; Guida et al., 2012). Evidence for chunking has, for example, been found in expert chess players (Chase & Simon, 1973). Foreign language learners can also use chunking strategies (Albelihi, 2022). According to the theory of long-term working memory, experts are assumed to have shorter fixation durations because they need less time to retrieve information (Gegenfurtner et al., 2011). Support for perceptual chunking was, for example, found in a study comparing novice, intermediate, and expert air traffic controllers as novices focused longer on irrelevant information than experts and intermediates (van Meeuwen et al., 2014).

Based on the CTML (see section 1.2.1) and these theories of expertise, Gegenfurtner et al. (Gegenfurtner et al., 2023). The authors propose a cognitive theory of visual expertise (CTVE) with three assumptions: First, that experts have a larger capacity for domain-specific information processing (e.g., Ericsson & Kintsch, 1995). Second, that visual processing of information changes with increased expertise (Haider & Frensch, 1999; Sheridan & Reingold, 2017) from a bottom-up to a top-down procedure. Third, that "experts interact with their environment when processing information of a visual scene" (Gegenfurtner et al., 2023, p. 153) and that they use meta-cognitive processes to evaluate information based on the current task. Apart from long-term working memory storing image chunks and prior knowledge, the CTVE proposes a visual register for (para-)foveal processing temporarily holding visual images. During information processing, experts use their meta-cognitive knowledge to monitor the cognitive processes processes proposed in the CTML: selecting, organising, and integrating information. These processes are refined and extended using the assumptions of the CTVE: Experts use para-foveal processing and are able to ignore irrelevant information. During this, they use their prior knowledge to determine which information is relevant. Experts also use their meta-cognitive knowledge to monitor their cognitive processes and apply domain knowledge to construct their visual field. The CTVE considers "educational usability" (Gegenfurtner et al., 2023, p. 150) and can, therefore, be applied to educational contexts.

Education researchers have compared differences in the visual behaviour of experts and non-experts in learning situations, for example, in the context of physics (Hahn & Klein, 2022). In this thesis, this is especially relevant in the context of graph comprehension. Previous research has found expertise-based differences in the visual processing of graphs. For example, differences in graph interpretation (Bowen et al.,

1999) and strategy during the process of solving graph-based tasks (Harsh et al., 2019) have been found between students and faculty. Although all participants reported using a comparable strategy, only experts implemented this strategy during their visual processing (Harsh et al., 2019). This is in line with the monitoring assumption proposed in the CTVE (Gegenfurtner et al., 2023) Undergraduate students seemed to learn their professor's interpretation instead of learning how to come to the right conclusion themselves (Bowen et al., 1999). Differences have also been found between science and engineering students considered experts and students studying other disciplines considered non-experts (Yen et al., 2012). When they first saw a graph problem, science students paid more attention to the question and less attention to the answers than non-science students. Presumably, the question area holds a lot of relevant information. Therefore, these results are in line with the information-reduction hypothesis (Gegenfurtner et al., 2011; Haider & Frensch, 1999). However, there were no statistically significant differences between science and non-science students in the viewing time on the graph (Yen et al., 2012). Another study compared the visual behaviour of professionals and students solving engineering problems (Ahmed et al., 2021). The results indicated that students fixated for a shorter time than professionals, although they seemed to make more fixations. The findings indicating longer fixations for experts are in line with the information-reduction hypothesis (Gegenfurtner et al., 2011; Haider & Frensch, 1999). However, experts making fewer fixations than non-experts indicates a slower visual search rate, which is in line with findings of previous reviews regarding decision tasks (Brams et al., 2019). This corresponds to the experts' ability to ignore irrelevant information (Brams et al., 2019) and is in agreement with the top-down procedure of experts as described in the CTVE (Gegenfurtner et al., 2023). A special focus of previous research has been on comparing physics students with students of different subjects during problem-solving with graphs: all studies found that physics students performed better than non-physics students independently of the task context (Brückner et al., 2020; Klein, Küchemann, et al., 2019; Susac et al., 2018). For example, physics students outperformed economics students in a post-replication study in both physics and finance tasks (Brückner et al., 2020). This might be due to the focus on graphs and their formulas in multiple topics of physics education (Pospiech et al., 2019). However, irrespective of the participants' subject, correct solvers focused more closely on relevant areas than incorrect solvers (Klein, Küchemann, et al., 2019), which is in line with the information-reduction hypothesis (Gegenfurtner et al., 2011; Haider & Frensch, 1999).

Solving domain-specific tasks correctly, therefore, seems to be an indicator of experience in a specific context as demonstrated by the differences in visual processing. This indicates that successful learners can use visual chunking strategies although more practice might be needed for brain reorganisation as indicated by neuroimaging experts (Guida et al., 2012). Consequently, the inferences of expertise theories regarding visual processing should be transferable to comparisons of correct and incorrect solvers.

Studies not only compared the differences between experts and non-experts to interpret their visual strategies. Eye movements were also used as predictors of performance (Klein, Küchemann, et al., 2019).

1.3.3. Predictivity of Eye Movements

Eye movements have been used to predict the answer correctness of students using statistical models (Becker et al., 2022; Chen et al., 2014; Klein, Küchemann, et al., 2019). For example, Chen et al. (2014) analysed 64 students' responses to computer-based textual or pictorial physics questions and found their performance could be predicted via their eye movements. Students seemed to look longer at their correct responses than at their incorrect responses. As a prediction model, Chen et al. (2014) used generalised estimating equations as extensions for generalised linear models (Liang & Zeger, 1986). There is also research on predicting students' performance on graph comprehension tasks using their eye movements: For example, using multiple regression, students' performance could be predicted for items of the Test of Understanding Graphs in Kinematics (TUG-K) (Beichner, 1994) via visual attention as indicated by total visit duration on relevant areas and those that were irrelevant to the solution of the tasks (Becker et al., 2022). Klein, Küchemann, et al., (2019) used the dwell time on concept-specific AOIs to predict solution correctness for graph comprehension tasks employing multiple ANOVAs as a statistical method.

Besides statistical methods, participants' performance on graph tasks can be predicted via machine-learning algorithms trained on the participants' eye movements (Dzsotjan et al., 2021; Küchemann et al., 2020, 2021). High-school students' (*N*=115) performance on items of the TUG-K could be predicted based on the students' visual behaviour, specifically the total dwell time (Küchemann et al., 2020) on and the number of transitions between AOIs on the graph (Küchemann et al., 2021). In a pilot study, the learning gain of 36 participants walking the shapes of acceleration-time graphs could be successfully predicted based on various eye-tracking metrics (Dzsotjan et al., 2021).

In contrast to statistical methods, which are used to evaluate known datasets, the goal of machine learning is to predict unknown outcomes based on known input data (Géron, 2019). In doing so, machine learning enables transfer to unknown and new datasets in a way that is not possible for statistical models. In an educational context, supervised learning algorithms are frequently used to predict students' performance (Namoun & Alshanqiti, 2021). Küchemann et al. (2020, 2021) used a Support Vector Machine (SVM) as a machine-learning algorithm. This is a supervised classification algorithm for categorising groups of an unknown test dataset based on the labels of the training dataset (Géron, 2019). SVM tries to find the best function separating the groups with the largest margin of error based on the training data (Géron, 2019). Dzsotjan et al. (2021) found that SVM was the best predictor compared to k-nearest neighbour and random forest. K-nearest neighbour groups data based on the data points' distance to each other (Theodoridis, 2020). Random forest uses an ensemble of decision trees for classification (Géron, 2019). The advantage of machine-learning algorithms, such as SVM, k-nearest neighbour and random forest, compared to statistical models is that they can also be used to predict complex non-linear relations between multiple features as they function independently of the data's distribution.

Which eye-tracking metrics should be used as input features can be analysed by evaluating feature importance. Using feature importance can also increase the interpretability of algorithms by indicating the importance of individual features for the prediction. For example, the error of the eye-tracker calibration was a relevant feature for the model trained by Caruso et al. (2022). They conducted a study with 147 university students investigating their reading comprehension and could predict this based on multiple eye-tracking metrics, such as fixation duration and saccades. In the context of graph comprehension, Dzsotjan et al. (2021) made similar analyses of feature importance. They found that adding more eye-tracking features does not necessarily improve the algorithms' performance.

Although machine learning has become more common in educational applications in recent years (Hilbert et al., 2021; Zawacki-Richter et al., 2019), using machine-learning methods can be difficult in an educational context, especially using eye movements as input features. This is due to the commonly small size of such datasets, frequently consisting of data of less than 100 participants (Hahn & Klein, 2022; Strohmaier et al., 2020). Small datasets can be difficult to analyse due to the limited number of data points (Rincón-Flores et al., 2022; Smith et al., 2014). A current study introduces a technique to judge the quality of machine-learning models and recommends a method to evaluate small datasets (Steinert et al., 2024): To reduce bias, machine-learning models are evaluated on various datasets using repeated nested cross-validation in combination with permutation tests. The latter established a probability of how generalizable the model is. The Matthews-Correlation-Coefficient (MCC) is the recommended metric for evaluating binary classification (Chicco & Jurman, 2020; Steinert et al., 2024). The MCC assesses the difference between actual and predicted variables and is not influenced by class imbalances which is valuable for doing machine learning with unbalanced datasets (Chicco & Jurman, 2020). This method could be particularly relevant for using machine-learning methods on educational eye-tracking datasets.

1.4. General Research Questions, Methodology, and Outline of the Studies

This thesis describes studies intended to contribute to facilitating graph use in educational practice by analysing how graphs are created as a tool in learning and problem-solving (research aim a), how graphs are visually processed by participants with varying expertise (research aim b), and what effect the study discipline has during problem-solving with graphs (research aim c). The theoretical foundation of all studies is the CTML (Mayer, 2014a) introduced in section 1.2.1 because all studies are concerned with learning with multiple representations including graphs. The first study is a systematic review of generating convention-based graphical representations of numerical data, defined as *graphing*, and connects to previous research on drawing (see section 1.2.2). Students often have difficulties during graphing (Clement, 1989; Leinhardt et al., 1990). It is in the nature of expertise, that experts have fewer difficulties. As expertise can be distinguished via eye movements (see section 1.3.2), a review of differences in visual processing between experts and non-experts during learning and problem-solving with graphs is presented as the second study of this thesis. The findings of the study support the information-reduction hypothesis (Haider & Frensch, 1999). This hypothesis was also supported by a study investigating graph comprehension in a physics context (Klein, Küchemann, et al., 2019). Although there does not seem to be an overarching definition of expertise in the context of graph comprehension (see section 3), physics students are a convenient comparison group for expertise research, because physics students are often considered experts compared to students of different disciplines (Brückner et al., 2020; Klein, Küchemann, et al., 2019; Susac et al., 2018). Furthermore, expertise is a common topic in physics education research (Hahn & Klein, 2022). The third study presents an eye-tracking study comparing first-year physics with medical students – both groups taking physics courses during the first semester. An analysis of the participants' eye movements indicates that there are differences in visual processing between correct and incorrect solvers. Results regarding performance differences between these groups are presented in Study 3 (see section 4).

The following sections provide short overviews of the three articles as well as brief clarifications of the specific theory and the methodology used. The detailed results are presented in the articles included in the following sections as well as in a brief overview of the findings of every study in section 5.1.

1.4.1. Outline of Study 1

Creating graphical representations is an important aspect of education (see section 1.2.2). Graphing falls under this description. Many studies have analysed graphing (see section 1.2.3). There are many educational contexts in which students can create graphs, for example, during problem-solving (Kanzaki & Miwa, 2011) or test-taking (Curcio, 1987), as well as during graphing instruction (Harsh & Schmitt-Harsh, 2016). Instruction might be combined with problem-solving activities, for example, by allowing students to revise graphs based on their ideas as part of an instruction regarding density concepts (Vitale et al., 2019). Graphing has also been investigated during lab activities (Adams & Shrum, 1990). For example, undergraduate students taking a physics lab course had difficulties finding the correct scale when constructing the frame for a graph in a study about finding best-fit lines (Nixon et al., 2016). Students also had trouble connecting the graph with the underlying concept, although they could successfully construct the line (Nixon et al., 2016). A special research focus has been on difficulties during graphing (e.g., McDermott et al., 1987), such as student misconceptions (Clement, 1985). Common student misconceptions include graph-as-picture errors, which, for example, is the case when students create a pictorial representation of a situation instead of a position-time graph based on data representing the situation (Gerard et al., 2012; Harrison et al., 2019). Extracting and appropriately representing data can, therefore, be difficult for students (Oslington et al., 2020). As graphs are an important representation not only in education and creating graphs is a relevant aspect of graphing competence (see section 1.2.3), it is important to know how graphing is instructed and what difficulties students may have.

However, there has not yet been a systematic review of the literature on this topic. Therefore, the first study aims to fill this research gap. The following research questions (RQ) were considered:

RQ 1: How is the graphing implemented in studies on this activity in K-12 STEM education?

RQ 2: What is the added value of graphing in K-12 STEM education?

RQ 3: Which difficulties can arise when graphing in K-12 STEM education?

To answer these research questions, a systematic literature review according to guidelines of preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 (Page et al., 2021) was conducted. First, terms in the relevant categories of graph creation, STEM education, and numerical data as the basis of graphs were used for a search of several databases (*ERIC, PsychInfo, Scopus*). The literature search found 10,296 articles. Two raters then screened the titles and abstracts of these articles in duplicate with a machine-learning-based software called ASReview (Utrecht University, 2021). This software can reduce the work needed for screening by approximately 80% (van de Schoot et al., 2021). Of the initially discovered articles, 394 were included in the full-text screening. This resulted in 41 included studies. Afterwards, additional forward- and backwards-snowball searches were conducted during which three additional eligible studies were found. Two coders coded relevant variables for all included articles to answer the research questions.

1.4.2. Outline of Study 2

The final research question of Study 1 focused on student difficulties during graphing. However, students not only have difficulties during graphing but also when learning and problem-solving with graphs (see section 1.2.3). Per definition, experts do not have the same problems as non-experts. This could be due to the more efficient information processing of experts compared to non-experts (see section 1.3.2). Cognitive processes can be inferred from eye movements (see section 1.3.1) and expertise differences can also be seen in visual processing (see section 1.3.2). Various reviews analysed differences in visual behaviour between experts and non-experts (Brams et al., 2019; Gegenfurtner et al., 2011; Sheridan & Reingold, 2017). This is also a common research topic in STEM education, such as physics (Hahn & Klein, 2022). As graphs are a common topic in STEM education, visual differences between experts and non-experts during learning and problem-solving with graphs have been a common research topic. An overview of visual differences between experts and non-experts in this area could help facilitate student learning as these can be used to develop support for students.

The second study presents a review of the literature investigating differences in the visual processing of experts and non-experts during learning and problem-solving with graphs. We had two research aims:

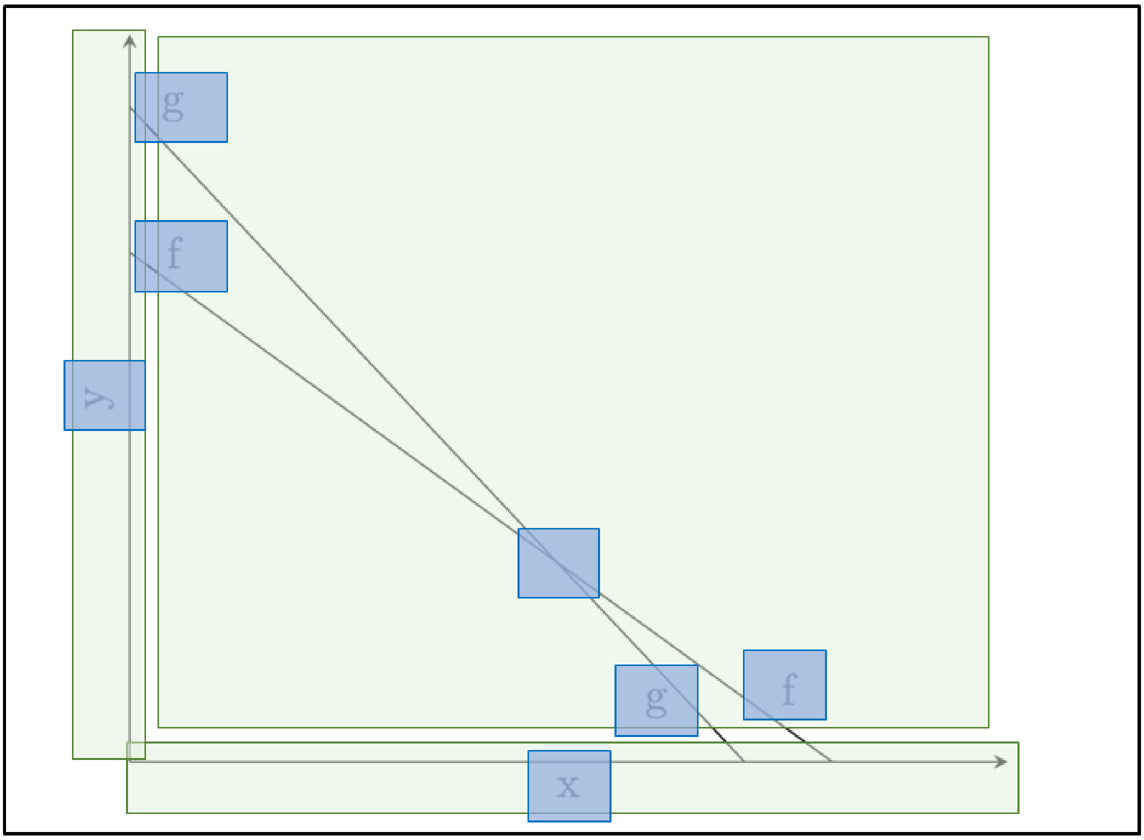
Research aim 1: Provide an overview of the eye-tracking metrics used to distinguish experts and non-experts.

Research aim 2: Provide an overview of differences in visual behaviour between experts and non-experts.

We analysed 32 studies published between 2003 and 2022 to answer these research questions. All studies compared the visual behaviour of experts and non-experts learning or problem-solving with graphs in STEM contexts. We coded relevant information including the graph subject, the type of graph used, the type of eye-tracking metric, and how many eye-tracking metrics were analysed. The literature review includes a summary and narrative analysis of these results. We distinguished between eye-tracking metrics and the size of the AOIs because the calculation of various eye movements, such as fixation duration, is based on the AOIs (see section 1.3.1). Three sizes can be distinguished: micro-level AOIs consist of very small areas, such as individual ticks on the x-axis of a graph; meso-level AOIs are bigger, for example, consisting of the entire x-axis of a graph; macro-level AOIs distinguish between large parts of a stimulus, such as between an entire graph and corresponding text (Andrá et al., 2015). An example of these AOIs is depicted in Figure 1: The entire graph is considered a macro-level AOI, the functions combined with the x-axes and y-axes depicted in green are meso-level AOIs. The function and axes labels as well as small task-relevant areas are micro-level AOIs.

Figure 1

Macro-Level AOI (Entire Graph), Meso-Level AOIs in Green, Micro-Level AOIs in Blue.



1.4.3. Outline of Study 3

Graph comprehension in the context of learning and problem-solving can be investigated with eye movements. Eye movements can indicate visual processing (see section 1.3.1) and expertise (see section 1.3.2). There are attempts to define expertise; for example, expert students should use knowledge intelligently and wisely, combining not only the application of knowledge but also the use of creative strategies and a successful transfer to practice (Sternberg, 2003). Other research considered specific disciplines, for example, physics experts might know how to approach a problem with an obvious solution without having to use the same problem-solving processes as physics novices (Maries & Singh, 2023). Further insights into problem-solving practices of various disciplines could lead to insights about expertise regarding the solution process of various types of problems, such as graph comprehension tasks, and how to facilitate those skills.

Several studies have compared physics students as experts with non-physics students as non-experts during problem-solving with graphs (Brückner et al., 2020; Klein, Küchemann, et al., 2019; Susac et al., 2018). All studies found that physics students outperformed non-physics students in graphs of physics and non-physics contexts (Brückner et al., 2020; Klein, Küchemann, et al., 2019; Susac et al., 2018). There were also differences in visual behaviour between correct and incorrect solvers (Klein, Küchemann, et al., 2019). With this study, we extend previous research by comparing physics students with medical students who also take physics courses during the first semester. We aimed to answer the following research questions:

RQ 1: Are there differences in learning gain between physics and medical students?

RQ 2: Are there differences in the visual behaviour of students solving tasks correctly or incorrectly?

We conducted a pretest-posttest study with first-semester medical and physics students. Students participated in November 2022 at the beginning of their first semester and again in March 2023 at the end of the first semester. They completed isomorphic items in physics and mathematics as in the previous studies (Brückner et al., 2020; Klein, Küchemann, et al., 2019) based on approved tests (Ceuppens et al., 2019; Susac et al., 2018). We designed similar isomorphic items for a medical context. The complete test material is available under https://osf.io/dgx3p/?view_only=515ffd3ec1bc474abfd7f1c2778d721e. The eye movements of the participants were recorded with Tobii Pro Nano eye trackers. We analysed dwell time to compare our results with previous studies (Brückner et al., 2020; Klein, Küchemann, et al., 2019; Susac et al., 2018).

Statistical analyses of learning gain and dwell time were conducted in R. Learning gain consisted of the difference between the percentage of correct items in the posttest and the percentage of correct items in the pretest. A factorial ANOVA was conducted to analyse differences in learning gain depending on the study subject (physics vs. medicine) and item context (mathematics vs. physics vs. medicine). Dwell time was analysed via multiple linear regression to analyse the effect of study discipline (physics vs. medicine), the specific AOIs (text vs. answers vs. graph vs. axes vs. axes labels), the context (mathematics vs. physics vs. medicine), the test (pretest vs. posttest), and the concept of the task (area vs. slope).

Previous studies found differences in visual behaviour, specifically in dwell times, between correct and incorrect solvers (Klein, Küchemann, et al., 2019). We aimed to replicate these findings with machine learning algorithms as this is assumed to be a suitable method to analyse eye movements (see section 1.3.3). Answer correctness via the participants' dwell times on the AOIs. Machine learning analyses were carried out in Python in the Jupyter Notebook environment. The machine learning algorithms used were optimised for small datasets (Steinert et al., 2024).

2. Study 1: A Systematic Review of Empirical Research on Graphing Numerical Data in K-12 STEM Education

Reference: Ruf, V., Thüs, D., Malone, S., Küchemann, S., Becker, S., Vogel, M., Brünken, R., & Kuhn, J. (2024). *A Systematic Review of Empirical Research on Graphing Numerical Data in STEM Education*. ArXiv. https://doi.org/10.48550/arXiv.2411.13195

A Systematic Review of Empirical Research on Graphing Numerical Data in K-12 STEM Education

Verena Ruf[ai], Dominik Thüs[bj], Sarah Malone[cj], Stefan Küchemann[di], Sebastian Becker-Genschow[e], Markus Vogel[f], Roland Brünken[gj], Jochen Kuhn[hi]

[a] Corresponding author. Email: v.ruf@physik.uni-muenchen.de, LMU, Edmund-Rumpler-Straße 13, 80939 Munich, Germany
[b] Email: dominik.thues@uni-saarland.de
[c] Email: s.malone@mx.uni-saarland.de
[d] Email: s.kuechemann@physik.uni-muenchen.de
[e] Digital Education Research, Faculty of Mathematics and Natural Sciences, University of Cologne, Herbert-Lewin-Straße 10, 50931 Cologne, Germany. Email: sebastian.becker-genschow@uni-koeln.de
[f] Fakultät für Natur- und Geisteswissenschaften, Pädagogische Hochschule Heidelberg, Keplerstraße 87, 69120 Heidelberg, Germany. Email: vogel@ph-heidelberg.de
[g] Email: r.bruenken@mx.uni-saarland.de
[h] Email: jochen.kuhn@physik.uni-muenchen.de
[i] Faculty of Physics / Chair of Physics Education Research, LMU Munich, Munich, Edmund-Rumpler-Straße 13, 80939 Munich, Germany.
[j] Department of Education, Saarland University, Campus Saarbrücken, Building A4 2, 66123 Saarbrücken, Germany.

Abstract

Graphs are essential representations in the professions and education concerning the science, technology, engineering, and mathematics (STEM) disciplines. Beyond their academic relevance, graphs find extensive utility in everyday scenarios, ranging from news media to educational materials. This underscores the importance of people's being able to understand graphs. However, the ability to understand graphs is connected to the ability to create graphs. Therefore, in school education, particularly in STEM subjects, not only the understanding but also the skill of constructing graphs from numerical data is emphasized. Although constructing graphs is a skill that most people do not require in their everyday lives and professions, it is a well-established student activity that has been empirically studied several times. Therefore, since a synthesis of the research findings on this topic has not yet been conducted, a summary of the studies investigating graphing via various viewpoints and differing methods could be a valuable contribution. To provide an overview of the empirical literature on this important topic, our systematic review identifies how the construction of convention-based graphical representations of numerical data, referred to as graphing, has been studied in previous research, how effective graphing is, and which types of difficulties are encountered by students. Based on these aspects, we defined inclusion criteria that led to 50 peer-reviewed empirical studies on graphing in K–12 STEM education found in SCOPUS, ERIC, and PsychInfo. Graphing instruction seemed to be beneficial for student learning, not only improving graph construction but also graph interpretation skills. However, the students experienced various difficulties during graphing, both during graph construction and the interpretation and usage of data.

Keywords: systematic review; STEM education; graphing; numerical data

Introduction

General mathematical skills, such as the competent handling of numerical data and their external representations (Chalkiadaki, 2018), are key skills in the 21st century for thinking critically and using information adequately (Program for International Student Assessment, 2021). Numerical data are often represented graphically to provide a comprehensive and easily accessible overview of a topic. Unlike informal representations, such as sketches, formal graphical data representations follow certain conventions, and to extract information correctly, it is necessary to be familiar with these conventions. We define *graphing* as the construction of convention-based graphical representations of numerical data, a key aspect of graphing competence (Glazer, 2011). *Graphing competence* contains both the interpretation and creation of convention-based graphical representations (Glazer, 2011) and is an important skill in K–12 education, especially in science, technology, engineering, and mathematics (STEM) disciplines. Consequently, both activities should be practiced extensively in class (Díaz-Levicoy et al., 2018; Dossey et al., 2016; Glazer, 2011; Kultusministerkonferenz, 2012). In addition, the ability to interpret numerical data is fundamental to understanding everyday statistical phenomena, such as stock markets, the progression of a pandemic, and risk evaluation (Schüller et al., 2019). Therefore, numerical data are highly relevant beyond STEM fields, and the ability to understand data is an essential part of everyday life (Friedrich et al., 2024). Several theoretical scientific works and multiple empirical studies have examined this graphing as a student activity in K–12 education. However, empirical research on this topic is diverse using different study designs and analysis methods. In this systematic review, we aim to synthesize existing empirical research on the use of student graphing of numerical data in K–12 STEM education and its effectiveness as an educational method by reviewing how this skill has been employed in various contexts, what its effectiveness has been in terms of different learning outcomes, and which difficulties students and teachers have reported.

Representations in STEM Education

Representations are common in STEM education (Rau, 2017) and serve multiple functions, such as memory support, facilitating inferences, or making discoveries (Tversky, 1997). Learning material can, for example, consist of text, equations, and graphical representations – such as pictorial or graphical representations of numerical data. Therefore, such learning material can be considered in the context of learning with multiple representations.

Multiple Representations in Education

Two prominent theoretical models that deal with multiple representations in learning are the *Integrated Model of Text and Picture Comprehension* (ITPC; Schnotz, 2005) and the *Cognitive Theory of Multimedia Learning* (CTML; Mayer, 2005, 2014). Both models, distinguish between symbolic and analogue representations (usually text and pictures) and predict the learning benefits of the simultaneous use of both types of representations owing to the construction of dual mental representations, which enhances the storage and retrieval of information (the multimedia effect; Mayer, 2005). Ott et al. (2018) showed that the multimedia effect also accounts for problem-solving using text and symbolic mathematical representations. Different types of representations accentuate different information (Zhang & Norman, 1994) thus activating different cognitive functions (Zhang, 1997). Understanding this thinking process is also important when considering graphical representations of data in STEM education (Duval, 2006). For example, data is interpreted differently depending on whether it is presented in a bar or a line graph (Shah & Freedman, 2011).

Besides their representational advantages, multiple representations can fulfil various other pedagogical functions. Ainsworth (1999) defined three different functions that provided or self-generated multiple representations can have during learning. First, two (or more) representations can complement each other. In this regard, Ainsworth (1999) distinguished between representations that involve complementary processes and those that contain complementary information. In complementary processes, diverse representations may help perform different tasks, support unique learners' characteristics, or accommodate specific strategies that learners may want to use. Multiple representations with complementary information are not redundant; rather, each individual representation provides unique essential information for learning. Second, multiple representations can constrain one another's interpretations to prevent misunderstandings—for example, by adding a familiar type of representation to the introduction of a new one. Moreover, representations constrain one another via their properties—for example, if one representation is more specific than another (Ainsworth, 1999). Third, representations can be used to deepen understanding via abstraction, extension, or relations. Abstraction, as a function of representations, entails the reorganization of information. If the reason for using multiple representations is extension, then knowledge is enhanced—for example, by being transferred to another context. Relating information between representations means translating between representations. Finally, representations can be used for more than one function simultaneously (Ainsworth, 2006).

These functions are also a part of Ainsworth's (2006) design, function, and task (DeFT) framework. This framework proposes that aspects of design, functions, and tasks should influence the creation of learning material that includes multiple representations. Apart from the aforementioned functions, DeFT considers the following five design parameters: the number of representations, how the information is distributed across representations, the types of representations, the order of representations, and the amount of support needed to translate between representations (Ainsworth, 2006). The cognitive tasks that learners are presented with when they encounter a (novel) representation include understanding how information is represented, how this information relates to the corresponding domain, how to choose a suitable representation, and how to construct a representation (Ainsworth, 2006). Self-generated graphs have the potential to serve all the functions from Ainsworth's approach, depending on the characteristics of the visualized information, such as design, the relation to other given representations, and learner characteristics. Moreover, graphing might include all cognitive tasks suggested by Ainsworth's framework (2006), as learners have to first choose and then correctly generate a graphical representation, which requires knowledge of how to correctly present information and relate the information to the domain as well as to other representations. Graphing thus leads to learners handling multiple external representations, which can promote learning in different ways.

Students need certain competencies to effectively learn with multiple representations, such as the ability to describe and use (multiple) representations appropriately, which involves the ability to choose a representation based on the context, task demands, as well as on personal ability and goals (Rau, 2017). This is an important instructional goal for students (diSessa, 2004), especially due to its relevance for scientific education (diSessa & Bruce, 2000) and practice (Gooding, 2004). The ability to use multiple representations is an essential skill for communicating information (de Vries & Masclet, 2013) and solving problems (Duval, 2006; Zhang, 1997). For example, sixth-grade students were more likely to solve math problems when they used schematic instead of pictorial representations (Hegarty & Kozhevnikov, 1999). Difficulties in dealing with representations might be due to lacking metarepresentational competence (diSessa, 2004). For example, Duval (2006) distinguishes between two different cognitive processes when transforming representations: transforming representations in one register (such as creating a graph based on a table, called *treatments*) and between registers (i.e., creating a function graph based on an equation, called *conversions*). This review considers only the first type of transformation.

Graphs of Numerical Data

Graphical representations of numerical data (graphs) are generally based on conventions, such as depicting the dependent variable on the y-axis (Lachmayer et al., 2007), which is particularly important for their function of conveying information (de Vries & Masclet, 2013). Graphs are widespread and frequently used representations that display the variance of a single variable (univariate) or the dependence of two (bivariate) or more (multivariate) variables (Eichler & Vogel, 2012). Numerical data often represent an observed phenomenon, including randomly caused variances, and functional graphs are used to model the structure of a bivariate (or multivariate) dataset via certain mathematical functions (Eichler & Vogel, 2012; Engel, 2018). Interpreting unknown graphs is an important skill because it denotes an act of learning – moving between abstract and concrete representations (Roth & Hwang, 2006). Functional graphs, which in a narrow sense are considered pure mathematical objects without contextual information, are not part of this review because such graphs can be constructed without numerical data using a mathematical equation alone. Graphing draws information from another source of numerical data (e.g., creating a line graph based on a data table). Generating such graphical representations is assumed to improve understanding, as the generation process requires engaging deeply with the structure of the represented numerical data, and the relevant characteristics of datasets can be communicated very effectively via graphical representations (Ainsworth et al., 2011). Therefore, we include studies on *self-generated* or *partially completed* graphical representations and exclude those that exclusively address the interpretation of or learning with *provided* graphs in our review.

Although numerous scientific studies have examined student graphing, to the best of our knowledge, scholars have not conducted a systematic review to synthesize the findings. The present review aims to synthesize the existing results, both positive and negative, regarding the construction of convention-based graphical representations based on numerical data.

Graphing in STEM Education

Graphing competence covers both the interpretation and creation of graphical representations (Glazer, 2011). Graphing competence is necessary for STEM learning—for example, students should be able to relate the data depicted in a graph to the phenomenon that it describes (Glazer, 2011). The definition of graphing competence also incorporates such concepts as "graph comprehension" (e.g., Curcio, 1987; Kanzaki & Miwa, 2012; Zacks & Tversky, 1999), "graph interpretation" (e.g., Biehler, 2006; Boels et al., 2019; Ergül, 2018; Gaona et al., 2021; Lachmayer et al., 2007; Nixon et al., 2016; Roth & Bowen, 2001), and selecting an appropriate graph type for a certain task (e.g., Baker et al., 2001; Kanzaki & Miwa, 2012; von Kotzebue et al., 2014). The term "graphical literacy" is also related to graphing competence and includes the skill of constructing graphs (Subali et al., 2017); in addition, it influences graph comprehension (Freedman & Shah, 2002).

Theoretical Framework of Graphing

Several theoretical approaches can be employed to predict the positive effects of graphing on learning. These approaches can be distinguished in terms of two aspects of graphing: graphing as learning using multiple representations (see above) and graphing as an active generative activity. The first perspective assumes that learning with more than one representation facilitates learning, as is the case when creating a second (graphical) representation that complements the first (often numerical) representation (Mayer, 2005, 2014; Schnotz, 2005). Theories on active generative activities argue that learners are more engaged when they generate representations (Chi & Wylie, 2014).

An educationally relevant aspect of graphing is that, for learners, it is an active generative process. Consequently, being located at the high end of Chi and Wylie's (2014) *Interactive, Constructive, Active*, and *Passive* (ICAP) framework, graphing has several advantages. The ICAP framework states that the effectiveness of learning activities decreases from interactive to constructive and from active to passive activities as learner engagement declines. According to the ICAP framework, generating a graphical representation that learners have not seen during learning and that is solely based on numerical data is a constructive activity and should be associated with considerable benefits.

This is supported by generative learning theory, in which Wittrock (1974) suggested that understanding is closely related to generation. He assumed that there are "organizational structures for storing and retrieving information" (p. 182) as well as mechanisms for retrieving stored (prior) knowledge and integrating new material. Wittrock (1992) explicitly stated that the aim of learning is not to store information but to build meaningful connections. This includes paying attention to and making sense of the learning material (Wittrock, 1992). Accordingly, instruction should focus on teaching learners to generate meaning. The overall benefit of generation activities, such as finding synonyms instead of merely reading a text, has also been called the "generation effect" and can be found in literature reviews (e.g., Bertsch et al., 2007). In their meta-analysis, Bertsch et al. (2007) found evidence that the generation effect increased with higher mental effort. When performing an active generation activity during learning, such as constructing a graphical representation based on the provided information, close attention should be paid to the presented information—for example, to the structure of the provided numerical data. This information is then connected to prior knowledge, such as previously learned conventions for interpreting graphical representations. Based on this research, one can assume that graphing is more advantageous to learners than merely looking at graphs and practicing graph comprehension. However, it is difficult to determine the effectiveness of graphing as an instructional method in general based on individual studies, especially because the way graphing is conducted (e.g., manually or via tools) might also influence the learning outcome.

There are numerous ways in which graphing activities can be implemented in education. In our review, we are interested in not only how graphing has been examined in studies but also whether and how authors embedded graphing into existing theories. As theoretical contextualization may depend on the purpose of education, analyzing the possible implications of this link could produce valuable insights for educators. The purpose and advantages of graphing in the STEM context may differ from the results found regarding the generation of external representations in general.

Previous Research on Graphing

Previous literature reviews have focused on students "drawing" or "sketching" during learning (Cromley, Du, & Dane, 2020; Fiorella & Zhang, 2018; van Meter & Garner, 2005; Wu & Rau, 2019), which, as generative processes, have certain similarities to graphing. According to van Meter and Garner (2005), sketches of various activities, such as swimming (Schmidgall et al., 2019), are representations of a mental image; therefore, the designs of representations can vary between learners. However, graphing follows general construction conventions, which means that graphing, when done correctly based on numerical data for a specific purpose, leads to comparable results regardless of the learner. Despite this, graphing is not a simple procedure, as even scientists sometimes do not choose the most appropriate graph type to display their data and, for example, seem to prefer line graphs to other types of graphs (Weissgerber et al., 2015).

The generation of representations is used to visualize and externalize information (Schmidgall et al., 2019). According to Stern et al. (2003), active graph generation supports an understanding of how graph design is generally related to the contents represented, compared with just exploring a given graph. This general understanding is transferable across domains. Furthermore, Stern et al. (2003) found that due to an enhanced examination of the material (generation effect), learners benefitted from generating graphs during a problem-solving task involving graphing in stock-keeping. In her review of graph interpretation, Glazer (2011) recommended instructing students in graph interpretation and graph creation (see also Cox, 1999). Such focused instruction can guide the generation of graphs, improve their accuracy, and highlight key points in graph design as well as the relationships between graphical representations (van Meter & Garner, 2005). Support during graphing can facilitate comprehension, although generation increases students' cognitive load compared to learning with provided representations (Zhang & Fiorella, 2021). Generating external representations also improves problem-solving (Cox, 1999), deep understanding, and knowledge transfer (Chi & Wylie, 2014).

In educational practice, generating graphical representations is relevant to several disciplines, most notably STEM subjects. Díaz-Levicoy et al. (2018) evaluated statistical graphs in mathematics textbooks for first- to sixth-grade students in primary education and found that generating graphs was the second most frequently taught activity (after basic mathematical operations, such as addition). A previous review by Leinhardt et al. (1990) of the literature on graphing and graph interpretation of functions also highlighted the importance of the ability to construct graphs because it "can be seen as one of the critical moments in early mathematics" (p. 2). Convention-based graphical representations of numerical data are used not only in mathematics but also in other contexts, such as physics, and research on graphing focuses on diverse learning objectives.

Using a pretest-posttest design, Mevarech and Kramarsky (1997) noticed a positive effect of graphing instruction on eighth-grade students' manual graphing abilities but also identified persistent difficulties. In a three-year-long study that analyzed students' reasoning when constructing line graphs, Wavering (1989) found that students' reasoning ability increased over time. This is particularly important because graphing is a relevant aspect of scientific inquiry (Gooding, 2010). For example, Schultheis et al. (2023) used authentic research experiences (Data Nuggets), which included graphing authentic datasets, in the context of biology and found that students improved in using scientific constructing scientific explanations. They were also more interested in STEM careers. Bahtaji (2020) compared undergraduate students' learning in the following three conditions, examining the conditions' effects on conceptual knowledge and graphing skills: (a) with provided graphs, (b) with self-constructed graphs, and (c) with self-constructed graphs made using explicit graphing instruction. He found that although all interventions facilitated conceptual knowledge acquisition, only explicit instruction developed graphing skills. Similarly, Harsh and Schmitt-Harsh (2016) used an instructional design to enhance graphing skills during a general education science course at a university and found that students improved between the pretest and the posttest. Angra and Gardner (2016) discovered differences between novices and experts in the construction of graphical representations based on tables. Along the same lines, experts' explanations were consistent with the graphs that they generated, which was not necessarily the case for novices (Kanzaki & Miwa, 2012). Based on a qualitative analysis, Angra and Gardner (2016) developed a step-by-step guide for teaching students how to choose a graph type, construct the graph after planning each step, and, finally, critically reflect on the graph choice. Based on these steps, teachers can adapt their instructions in relation to students' answers, and students can develop a clear process for graphing.

Furthermore, graphing can be performed manually as well as with the help of software tools. For example, in a qualitative analysis, Parnafes and Digoodi (2004) found that reasoning type was related to representation type when middle school students interacted in groups with the software environment NumberSpeed. The students used content-based reasoning when working with number–list representations and more model-based reasoning when working with spatial and dynamic motion representations. Nixon et al. (2016) noted that undergraduate students' understanding of best-fit lines changed depending on the physics lab activity. Biehler's (2006) results indicated that students had difficulties interpreting graphical representations in the context of a task during a qualitative analysis of students' group work with Fathom software.

Overall, previous studies in STEM education have revealed (a) a variety of educational settings in which the graphing of numerical data is useful (Bahtaji, 2020; Harsh & Schmitt-Harsh, 2016), (b) the cognitive aspects necessary for and during graphing (e.g., Mevarech & Kramarsky, 1997; Wavering, 1989), and (c) students' graphing strategies and their interpretations in the context of physics lab activities (Nixon et al., 2016). In the present review, we aim to systematically synthesize and summarize studies on the graphing of numerical data to facilitate the comparison of various methods across different STEM contexts in terms of their similarities, differences, and effectiveness.

Students' Difficulties

Understanding students' difficulties is important to facilitate graphing competence. In a literature review, Boels (2019) described conceptual difficulties when interpreting histograms related to the interpretation of data and distribution. Moreover, von Kotzebue et al. (2014) investigated the mistakes made by 437 science students when constructing diagrams in the context of biology. Students had trouble in all analyzed categories, such as choosing the correct diagram type, assigning variables to axes, and scaling. Nixon et al.

(2016) observed that determining the axes' scaling posed a problem for students when graphing during physics lab activities. Baker et al. (2001) asked eighth and ninth graders to select, interpret, and construct graphs. Most students' performance seemed to depend on their ability to transfer knowledge from bar graphs to other graphical representations of data. Bayri and Kurnaz (2015) examined eighth-grade students who constructed different types of graphical representations and concluded that students had difficulty shifting between different types of representations. The authors assumed this to indicate that, for students, various types of graphical representations could have different meanings. One reason for this could be that students see representations as analogue reflections of reality rather than representations of a symbolic character (graph-as-picture error; Clement, 1989).

These studies show how difficult graphing can be for students and how important it is for instructors to be aware of these difficulties. Therefore, it is important to consider not only how graphing is implemented in various studies and its effectiveness in these contexts but also the possible problems faced by students.

Research Questions

Based on previous research, our literature review aims to systematically examine graphing in STEM subjects in K-12 education. This is the context in which students learn how to construct graphs and where graphing is often assessed (i.e., in exams). This review therefore addresses the following questions:

(1) How is graphing implemented in studies on this activity in K-12 STEM education?

(2) What is the effectiveness of graphing as an instructional method in K-12 STEM education?

(3) Which difficulties can arise when graphing in K-12 STEM education?

Methods

This systematic review is documented and reported according to the guidelines for preferred reporting items for systematic reviews and meta-analyses 2020 (PRISMA, Page et al., 2021). In the following section, we present the exact method.

Inclusion and Exclusion Criteria

An article's suitability was judged based on the inclusion criteria. We were interested in all studies that examined the graphing in K–12 STEM education. Participants were students in K-12 education who participated in studies with a STEM topic. We were only interested in self-generated or partially completed graphical representations rather than pure graph interpretations. The representations should be based on convention; completely new or made-up representations were not included. Furthermore, the generated graphs should be based on numerical data. Studies that investigated graphing based on given verbal–textual or mathematical–symbolical information were excluded. All criteria are summarized in Table 1.

Table 1

Criteria for Inclusion and Examples of Exclusion

Criterion	Inclusion	Exclusion examples
Subjects	K-12 students	University students
STEM Education	STEM topic	Liberal arts topic
Graphing	Self-produced graph	Provided graphs
STEM practice	Convention-based graph	Self-invented representations
Numerical data	Data-based graph	Function graph not based on data

Publication language	English	Korean
Publication type	Empirical research	Theoretical research
Scientific evaluation	Peer review	Grey literature

Literature search

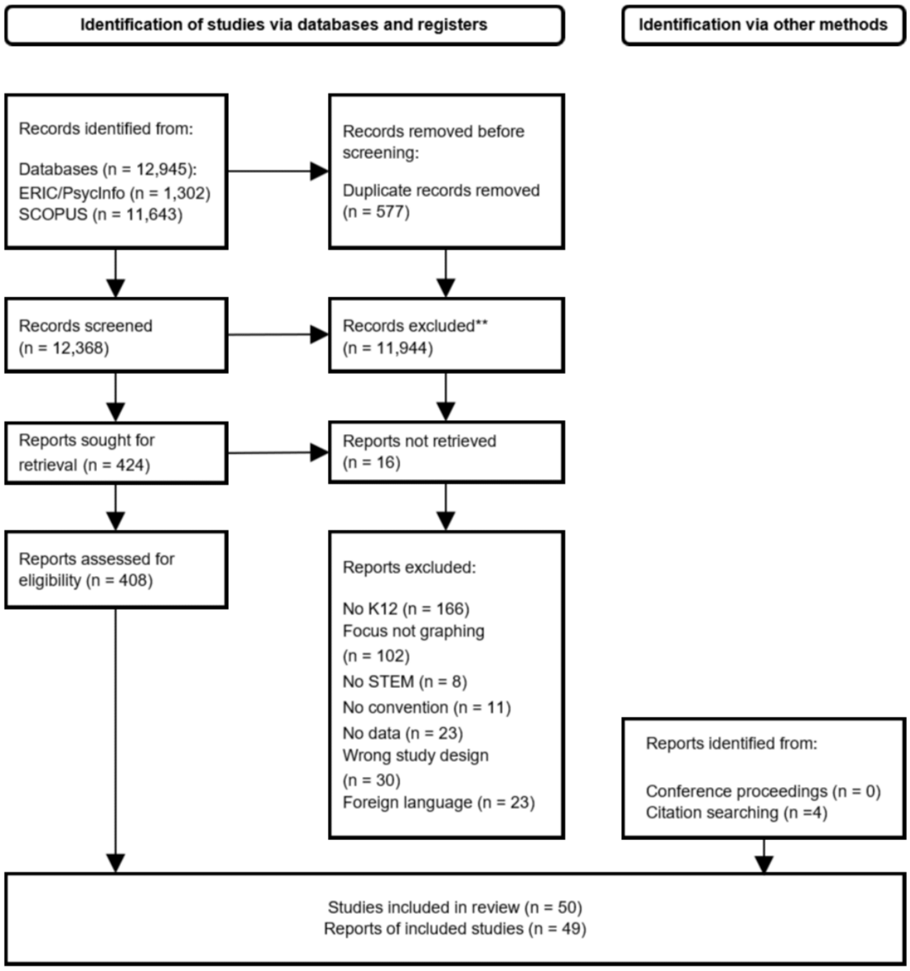
The search was conducted in March 2022 using the following three major scientific databases selected by our interdisciplinary review team due to the high quantity of educational research in various disciplines: Scopus, ERIC, and PsycInfo. ERIC and PsycInfo were accessed via EbscoHost. Relevant terms in the categories of Education, Graphing, and Data use were based on previous scoping searches to form a complete yet economical search term. The full search term was *(educat* OR student*) AND (graphing OR graph OR graphs OR plotting OR plot OR plots) AND (data* OR variable* OR construct*)*. We considered the title, abstract, and keywords in the systematic search.

Exported records were saved, managed, and deduplicated using the reference management software Mendeley. After deduplication, the records were screened. In the first step of the screening, only the title and abstract were considered. For this, we used ASReview (Utrecht University, 2021). ASReview speeds screening by using machine learning to prioritize the articles found during the systematic search (van de Schoot et al., 2021). The software continuously updates the order of the titles and shows the most relevant studies first. The algorithm was first trained using included and excluded articles, which were found during scoping searches and discussed by the review team. Each user decision (include/exclude) made in ASReview was used to improve the model during the entire screening process. Two coders from the review group trained the algorithm up to a termination criterion (<5 relevant titles in 100). The same two raters then manually screened the studies that were judged relevant when considering the full text using the software

Rayyan (Ouzzani et al., 2016). Any discrepancies were solved via discussions between the two raters and, if necessary, with the broader review group. Therefore, we reached complete agreement between the two raters. In the next step, relevant conferences in the psychological and educational contexts were reviewed for the relevant proceedings according to the following eligibility criteria: We looked for international conferences in a (STEM) education context, with English abstracts, and full papers published as part of the proceedings. The conferences should also be peer-reviewed. Seven conferences matched these criteria: the International Conference of the Learning Sciences (ICLS), the International Conference on Teaching Statistics (ICOTS), and the meetings of the American Educational Research Association (AERA), the European Association for Research on Learning and Instruction (EARLI), the European Science Education Research Association (ESERA), the Cognitive Science Society, and the National Association for Research in Science Teaching (NARST). We restricted our search to start from the year 2019 because we assumed that relevant research would be published as a paper that would be found in the normal search process. For articles included after the full-text screening, a forward and backward search for further potentially relevant articles was conducted. A second search for papers published after the initial search was conducted in April 2024 to update the results. A complete overview of the literature search and screening process can be seen in Figure 1.

Figure 1

PRISMA Flow Diagram Detailing the Study Selection



Data Extraction and Coding Schemes

In our literature review, we summarize and synthesize studies that deal with the contexts, effects, and difficulties of graphing in K–12 STEM education. For each included study, we extracted relevant data (see Table 2 for an overview of example codes). Two coders coded all studies, and possible disagreements were resolved via discussion.

We extracted information about the sample and the general study characteristics to structure the body of the research. We classified the selected studies according to the addressed topic, as research indicates that students' representation use and graph interpretation can differ between domains and contexts (Chang, 2018; Roth & Bowen, 2001). Furthermore, we are interested in the types of graphical representations that students generated to compare the graphing of univariate (e.g., bar graphs) and multivariate data (e.g., scatter plots). This could be relevant because, for example, bar and line graphs are associated with different concepts, such as discrete comparisons and trends (Zacks & Tversky, 1999). For graphing, it may be important whether the learners generated the data themselves, as described by Nixon et al. (2016), or whether they were provided with the data (e.g., Bahtaji, 2020). This might be relevant as self-generated and provided data have different learning benefits (Hug & McNeill, 2008). With the availability of computers and other technical tools, there might be differences due to technological difficulties (Chang et al., 2024) between manual graphing (e.g., Wavering, 1989; Subali et al., 2017) and the construction of graphical representations via programs, whose availability has increased in recent years (e.g., Kohnle et al., 2020; Lee & Lee, 2018). This might influence the perspective from which students view data because it might be easier to create representations of aggregate data, such as box plots, with tools than manually (Konold et al., 2015). Moreover, we would like to know whether a gender-balanced study design is comparable to, for example, an unbalanced design (e.g., Castro-Alonso et al., 2019). For each included study, the ratio of male to female participants was coded. This allowed us to assess whether gender was considered in the research, the extent to which the effects found might be generalized across genders, and whether participants' gender was a possible moderator of the effects. All themes, categories, and example codes for the coded variables are shown in Table 2. If variables were not reported, they were coded as "NR."

Table 2

Themes, Categories, and Example Codes From the Data Analysis

Theme	Category	Example Code
Population	Children	Pre-school, primary school (approx. age 5–10), high school (age >10/11)
General information	Gender	Balanced, unbalanced (female skew), unbalanced (male skew)
	Country	Germany, the USA, the Netherlands
	N	92,110
Topic	STEM	Biology, chemistry, physics, computer science, engineering, math, technology
Graphing	Method	Manual, tool-based
	Graph type	Histogram, box plot, line graph, scatter plot
	Guidance	Minimal, explicit, completing, comparison
Numerical data	Collection	Self-generated, given
	Data type	Univariate, bivariate, multivariate
Study design	Activity	Problem-solving, experiment, instruction
	Study setting	Lab, lesson, course
	Moderating variable	Prior knowledge, motivation, spatial, load
Results	Analysis	Quantitative, qualitative, mixed methods
	Overall effect of graphing	Positive, negative, inconclusive
	Effect size	*r, d*
	Hypothesis	Confirmed, unclear

Theory	Cognitive theory for graphing	CTML/ITPC, generative learning, ICAP, other[a]
Difficulties	Construction difficulties	Graph construction; Variable ordering; Data translation
	Theoretical difficulties	Connection to concepts; Interpretation

Note. CTML: Cognitive theory of multimedia learning (e.g., Mayer, 2014), ITPC: Integrated model of text and picture comprehension (Schnotz, 2005), ICAP: Interactive, active, constructive passive framework (Chi & Wylie, 2014).

Data Analysis

The data extracted from the included studies were saved and processed in Microsoft Excel and Access. In addition to answering the research questions, we performed a narrative comparison of the studies in terms of their use of univariate versus multivariate data.

Publication bias is indicated by the systematic difference between published and unpublished research (Vevea et al., 2019). As this is, to the best of our knowledge, the first literature review of graphing in STEM subjects, we wanted to ensure high quality by only including peer-reviewed studies. We believe that bias was sufficiently reduced by using multiple search engines and covering various topics and journals, as well as by including proceedings from conferences judged relevant by experts in various STEM fields.

Results

In the following sections, we present the results according to the codes (see Table 2). Sometimes, the authors did not explicitly mention the information encoded for our review; these studies are not reported below. An overview of all the studies included in the literature review can be seen in Table 3. A complete overview of all codes is provided in the tables of the supplementary material.

Table 3

Overview of the Relevant Codes for the Included Studies

Authors	Year	STEM Topic	Data Collection	Construction Difficulties	Theoretical Difficulties
Åberg-Bengtsson	2006	Math	Both[c]	GC[e]	Part vs. Whole
Adams & Shrum	1990	Biology	SG[d]	NR	NR
Arteaga et al.	2020	Math	Both	GC	NR
Ates & Stevens	2003	Chemistry	SG	NR	NR
Aydın-Güç et al.	2022	Math	Given	DT	IP[h]
Berg & Phillips	1994	NR[a]	Given	GC	Concept
Branisa & Jenisova	2015	Chemistry	Given	GC, DT[f]	IP, Concept
Brasell & Rowe	1993	Physics	Given	GC, VO[g]	Concept
Detiana et al.	2020	Math	SG	NR	NR
Dewi et al.	2018	Physics	NR	GC, VO	IP, Concept
Dimas et al.	2018	Physics	Given	GC	Concept
English & Watson	2015	Math	SG	GC	NR
English	2022	Physics	SG	NR	Concept
English	2023	Math	SG	NR	Concept, Part vs. Whole
Fielding-Wells	2018	Math	SG	NR	IP
García-García & Dolores-Flores	2019	Math	SG	DT	NR
Garcia-Mila et al. (Study 1)	2014	Math	Given	GC, VO, DT	NR

Garcia-Mila et al. (Study 2)	2014	Math	Given	GC, VO	NR
Gardenia et al.	2021	Math	SG	DT	NR
Gerard et al.	2012	NR	SG	NR	Concept
Gultepe & Kilic	2015	Chemistry	Given	GC, VO	NR
Gültepe	2016	Chemistry	NR	DT	IP, Concept, GT[i]
Harrison et al.	2019	NR	SG	VO	IP
Jackson et al.	1992	CS[b]	Given	GC	GT
Jackson et al.	1993	CS	Given	GC	IP, GT
Karplus	1979	Math	Given	VO	Concept
Kramarski	1999	Math	SG	GC, DT	NR
Meisadewi et al.	2017	Biology	NR	NR	NR
Mevarech & Kramarsky	1997	Math	SG	GC, VO	NR
Moritz	2003	Math	Both	GC, VO, DT	NR
Ng & Nicholas (Study 2)	2011	Biology	SG	NR	NR
Nurrahmawati et al.	2021	Math	SG	VR, DT	IP
Onwu	1993	Various	Given	GC, VO	IP
Oslington et al.	2020	Math	Both	GC, DT	Concept
Ozmen et al.	2020	Math	Given	GC, VO	GT
Padilla et al.	1986	Various	Given	GC, VO	NR
Pols	2019	Physics	SG	GC	Concept
Pospiech et al. (Study 1)	2019	Physics	Both	NR	NR
Pratt	1995	Math	SG	VO	IP
Rahmawati et al.	2020	Math	SG	NR	NR
Saldanha & McAllister	2016	Math	Given	NR	NR

Stephens	2024	Physics	SG	NR	IP, Concept
Struck & Yerrick	2010	Physics	SG	VO	NR
Tairab & Al-Naqbi	2004	Biology	Given	DT	IP
Vitale et al.	2019	Physics	Both	NR	Concept
Watson	2022	Physics	SG	NR	IP
Watson et al.	2023	Physics	SG	GC	NR
Wavering	1989	Math	Given	GC	NR
Webb & Boltt	1991	Biology	Given	NR	IP
Wu & Krajcik	2006	Various	SG	GC	IP

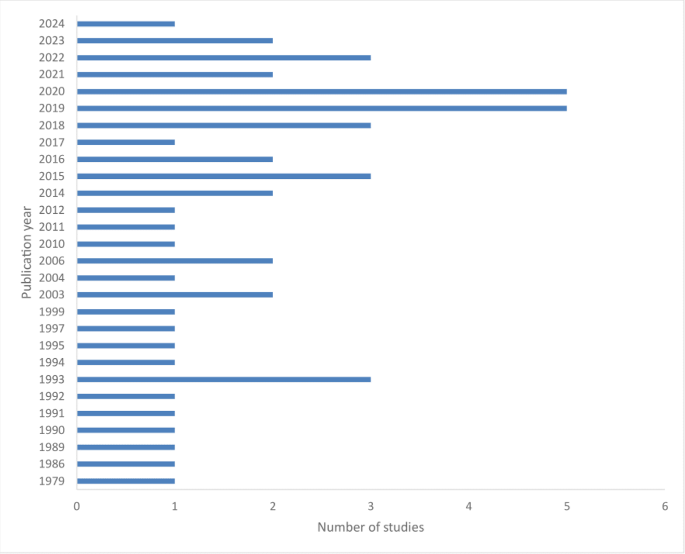
Note: [a] NR: None reported, [b] CS: computer science, [c] given as well as self-generated data: Both, [d] self-generated: SG, [e] graph construction: GC, [f] data translation: DT, [g] variable ordering: VO, [h] interpretation: IP, [i] graph type: GT

General Information

The studies included in the review were published between 1979 and 2024, which is the last year in which we identified relevant articles in the literature review (see Figure 2). The rate of publications on graphing seems to have reached its maximum in the years 2019 and 2020.

Figure 2

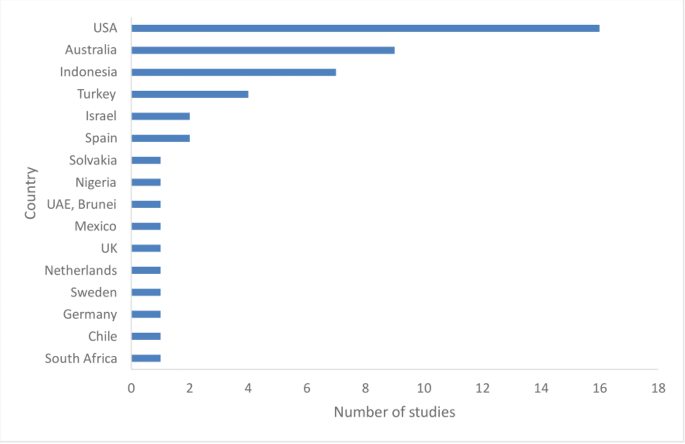
Overview of the Publication Years of the Included Studies



Most of the studies were conducted in the USA, followed by Indonesia and Australia (see Figure 3). Few studies were published in Europe, which is surprising because other studies reference, e.g., the German curriculum (Meisadewi et al., 2017).

Figure 3

Overview of the Countries Where the Studies were Published



Most studies did not base their research on a theory specific to graphing (n=23). Two studies referred to generative learning (Åberg-Bengtsson, 2006; Vitale et al., 2019). Additionally, 27 studies referred to other theories. For example, Ainsworth's (2006) design, function, and task (DeFT) framework, which is based on the functions that representations have during instruction, was referenced for this purpose (Dimas et al., 2018), but it was also used in one study as a theoretical basis for translating between two types of representations, such as the "mathematical representation translation from verbal to graph" (p. 401, Rahmawati et al., 2020). Similarly, English (2022) referred to metarepresentational competence and the need to connect graphical representations to the represented concept (see also Vitale et al., 2019), as well as transform types of representations (Garcia-Mila et al., 2014). Other tool-based studies have argued that computer-based graphing frees up cognitive resources (Jackson et al., 1992, 1993). Another perspective includes a theoretical background based on the progression of logical thinking abilities indicated by graphing (Berg & Phillips, 1994). Graphing has also been used as a good research tool to capture "students' thinking processes" (Ng & Nicholas, 2011, p. 79).

Another study used active learning as its theoretical basis (Pratt, 1995). One study used a pedagogical approach called *Active Graphing* (Åberg-Bengtsson, 2006), introduced in previous research (Ainley, 2000). For microcomputer-based laboratories (MBLs), authors based their research on the benefit of having a "genuine scientific experience" (p.778, Adams & Shrum, 1990) or on the view of "graphing as practice" (p.57, Ates & Stevens, 2003) and did not consider graphing from a cognitive perspective. This scientific (Fielding-Wells, 2018; Stephens, 2024; Watson et al., 2023; Wu & Krajcik, 2005) and practical (Aydın-Güç et al., 2022; Gültepe, 2016; Harrison et al., 2019; Oslington et al., 2020) view is stated similarly in other studies, for example, as part of problem-solving (Bransia & Janisova, 2015; Pospiech et al., 2018).

Implementation of Graphing

Study Design and Context

Most authors analyzed graphing in a problem-solving environment (n=36), although participants received graphing instruction during some studies (n=6). For example, Mevarech and Kramarsky (1997) asked 92 eighth-grade students to construct graphs before and after a graphing unit and qualitatively analyzed their progress based on the students' responses. Several studies investigated graphing in the context of experimentation (n=10), such as investigating tool-based graphing in the context of oscillation using a spring-mass simulation (Stephens, 2024). Meisadewi et al. (2017) found that lab-based activities could improve students' graphing skills [see also Struck & Yerrick (2009) for similar results, as well as Gerard et al. (2012) for a mixed-methods analysis]. Three studies used an instructional context combined with problem-solving (Åberg-Bengtsson, 2006; Gerard et al., 2012; Vitale et al., 2019), for example, Åberg-Bengtsson (2006) instructed elementary students on how to use the software Excel in a collaborative setting and investigated their reasoning during graph construction.

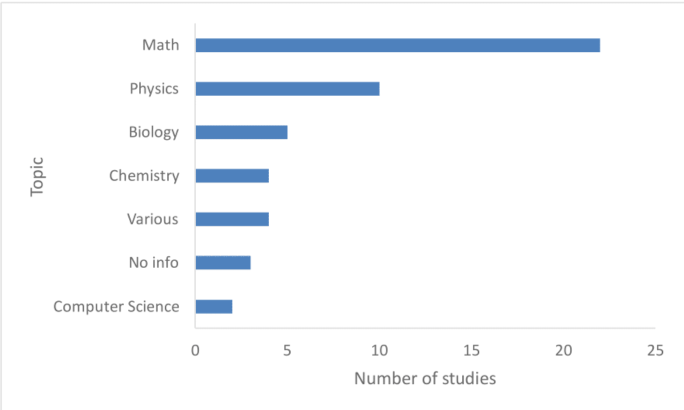
Studies were most often conducted in more than one lesson (n=22), followed by a single lesson (n=17) and interviews (n=5). Interviews were sometimes conducted concurrently with lessons (Aydın-Güç et al., 2022; Karplus, 1979); for example, Karplus (1979) asked 414 high-school students to solve "functionality puzzles" (p. 398) during a lesson and investigated the answers of 37 students during interviews to "clarify the written answers" (p. 398). Only one study was conducted in the researchers' lab (Pospiech et al., 2019).

The number of participants differed greatly between studies, from three (Detiana et al., 2020; Gardenia et al., 2021) to 745 (Arteaga et al., 2020). On average, 116 students participated (SD = 174). Students in high school participated in most studies (n=43), with only six studies analyzing graphing in primary schools. One study analyzed students from grades three to seven, spanning both primary and high school (Moritz, 2003). The gender of the students was often not reported (n=31). Six studies had an equal number of female and male participants and in nine studies more males than females participated. There were four studies where more females than males participated.

The STEM topics varied between studies (see Figure 4). Graphing was mostly conducted on the topic of mathematics (n=22); for example, Moritz (2003) examined how primary and high-school students constructed coordinate graphs during their mathematics lessons. Other studies were conducted in biology, physics, chemistry, and computer science. In three studies, students constructed various contexts (Onwu, 1993; Padilla et al., 1986; Wu & Krajcik, 2006) — for example, by "incorporat[ing] fundamental science concepts across several science disciplines" (p.66, Wu & Krajcik, 2006).

Figure 4

STEM Topics Used for Graphing

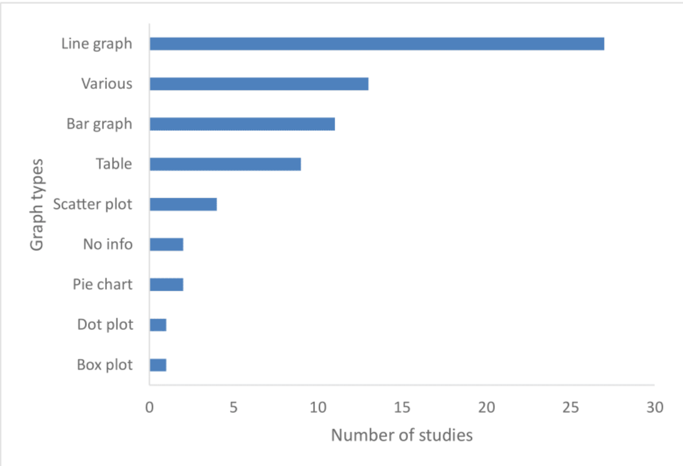


Graph Types, Graphing Method, and Graphing Guidance

Students constructed numerous types of graphs (see Figure 5). The most frequently constructed graph type was a line graph (n=27). Other studies let students decide the type of graphs, e.g., for a task during a test (Ozmen et al., 2020). Bar graphs and tables were also common types of graphs. Two studies did not specifically state which types of graphs they used (Branisa & Jenisova, 2015; Meisadewi et al., 2017).

Figure 5

Types of Graphs Constructed by Participants



Note: Multiple mentions are possible.

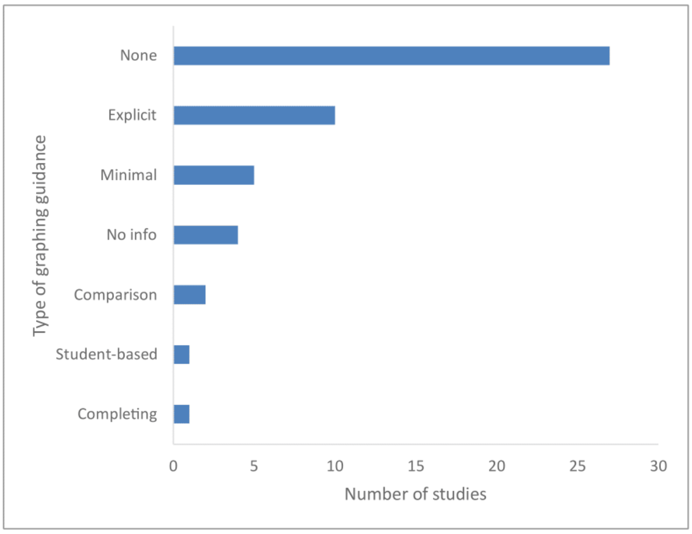
During most studies, participants created graphs by hand (n=29); for example, Nurrahmawati et al. (2021) investigated high-school students' errors when translating between mathematical representations, such as between tables and line graphs. Nine studies analyzed graphs created via tools (Gerard et al., 2012; Harrison et al., 2019; Jackson et al., 1992; Jackson et al., 1993; Ng & Nicholas, 2011; Pratt, 1995; Saldanha & McAllister, 2016; Stephens, 2024; Vitale et al., 2019). Ten studies examined graphs created both manually and via tools (Åberg-Bengtsson, 2006; Adams & Shrum, 1990; Ates & Stevens, 2003; Brasell &

Rowe, 1993; English & Watson, 2015; English, 2022; Kramarski, 1999; Watson et al., 2022; Wu & Krajcik, 2006). All studies examining tool-based graphing used various computer applications, such as TinkerPlots (English & Watson, 2015; Saldanha & McAllister, 2016; Watson et al., 2022; Watson et al., 2023), the web-based inquiry science environments WISE (Gerard et al., 2012; Harrison et al., 2019; Vitale et al., 2019), the web-based data exploration environment CODAP (Stephens, 2024), CricketGraph (Jackson et al., 1992, 1993), or Excel (Åberg-Bengtsson, 2006; English, 2022).

Participants often did not receive any guidance about how to create graphs, although ten studies provided explicit instructions (see Figure 6). For example, Detiana & Mailizar (2020) showed high-school students instructional video tutorials on function graphs before asking them to manually construct their own graphs. Participants received minimal instruction in five studies (Brasell & Rowe, 1993; Jackson et al., 1993; Karplus, 1979; Pratt, 1995; Webb & Boltt, 1991), which included providing participants with worksheets, including frameworks for the graphs (Webb & Boltt, 1991). In one study, the students were asked to complete a representation consisting of a table with corresponding graphs (Åberg-Bengtsson, 2006). In two other studies, students were asked to compare their representations to graphs created by fictional students (Harrison et al., 2019) or to compare their graphs with teacher-generated graphs constructed with TinkerPlots (Watson et al., 2022). Guidance was coded as “student-based” for a study in which participants were interviewed while they constructed graphs and received feedback based on their progress (García-García & Dolores-Flores, 2019).

Figure 6

Types of Guidance for Graphing



Numerical Data Used for Graphing

The data that the students used for graphing varied. Participants mostly created the data themselves (n=23), either during experiments by measuring variables or by creating the data for graphing, e.g., based on problem statements (Gardenia et al., 2021). Eighteen studies provided participants with the data to create graphs, and some studies used a combination of both (n=6). For example, Oslington et al. (2020) provided third-graders with temperature data and asked them to predict future temperatures before constructing a graphical representation of the data.

Most of the numerical data used for graphing was bivariate (n=34), compared to univariate (n=6) and multivariate (n=3) data. Two studies provided students with a variety of data; for example, Jackson et al. (1992) provided students with graphing instruction for *Cricket Graph*^{TM} in a computer science course using a variety of contexts and data.

Effectiveness of Graphing as an Instructional Method

Study Results and Moderators

Most studies reported no hypothesis (n=45) and no effect sizes (n=49) or results regarding the benefits of graphing or their instruction (n=38). This might be due to the number of qualitative (n=22) and mixed-methods analyses (n=19). Nine studies analyzed their data quantitatively (Adams & Shrum, 1990; Ates & Stevens, 2003; Harrison et al., 2019; Meisadewi et al., 2017; Onwu, 1993; Padilla et al., 1986; Struck & Yerrick, 2010; Wavering, 1989; Webb & Boltt, 1991).

Eleven studies reported positive effects (Adams & Shrum, 1990; Branisa & Jenisova, 2015; Karplus, 1979; Gerard et al., 2012; Gultepe & Kilic, 2015; Meisadewi et al., 2017; Mevarech & Kramarsky, 1997; Padilla et al., 1986; Struck & Yerrick, 2010; Vitale et al., 2019; Wu & Krajcik, 2006). Two studies reported positive effects for manual graphing (Adams & Shrum, 1990; Branisa & Jenisova, 2015). Branisa and Jenisova (2015) compared manual graphing with automatically generated graphs and found that students practiced manual graph construction based on experimental data performed better than students who were provided with computer-constructed graphs. Adams and Shrum (1990) found that conventional graphing instruction was better than instruction with microcomputers. They were the only authors who reported an effect size (−1.01). Some studies reported positive effects of various types of instruction (Gerard et al., 2012; Gultepe & Kilic, 2015; Meisadewi et al., 2017; Mevarech & Kramarsky, 1997; Struck & Yerrick, 2010; Wu & Krajcik, 2006) on graphing skills. For example, Gerard et al. (2012) compared drawing tools with probe-based tools and found that probe-based tools might be better for learning how not to see graphs as pictures but as representations of data. However, students who drew instead of using motion sensors "constructed more precise graphs and verbal interpretations" (Gerard et al., 2012, p. 569). Another possible moderator of graphing skills might be the level of education because

Karplus (1979) and Padilla et al., (1986) reported an increase in skills with a progression between grades. While Karplus (2006) compared students between sixth and eighth grade who constructed function graphs based on data pairs, his results correspond to the results of a mixed-methods approach by Mevarech and Kramarsky (1997), who found that instruction improved students' graphing performance. The goal of instruction varied between studies: some specifically wanted to improve students' graphing skills (Gerard et al., 2012; Meisadewi et al., 2017; Mevarech & Kramarsky, 1997), whereas others taught specific topics, such as kinetics (Struck & Yerrick, 2010) or water quality (Wu & Krajcik, 2006). One study aimed to improve students' scientific argumentation skills (Gultepe & Kilic, 2015). Furthermore, students benefited from graphing data that illustrated their ideas and revising their graphs based on scientific concepts (Vitale et al., 2019). Of the studies reporting positive effects, five analyzed their results quantitatively (Adams & Shrum, 1990; Meisadewi et al., 2017; Padilla et al., 1986; Struck & Yerrick, 2010) and one qualitatively (Wu & Krajcik, 2006). Six studies used a mixed-methods approach (Branisa & Jenisova, 2015; Gerard et al., 2012; Gultepe & Kilic, 2015; Karplus, 1979; Mevarech & Kramarsky, 1997; Vitale et al., 2019). One study reported inconclusive results (Ates & Stevens, 2003). None of the included studies reported any negative effects of graphing.

Most studies (n=37) did not document specific participant characteristics influencing graphing skills. Thirteen studies reported possible moderators. Most effects were reported for age or grade (n=7) and types of mathematical understanding (n=3). For example, Gardenia et al. (2020) found that students with high mathematical skills were better able to construct mathematical representations than students with low or medium skills. Other studies mentioned moderating effects due to cognitive development (n=4), such as reasoning skills (Ates & Stevens, 2014; Berg & Phillips, 1997; Wavering, 1989). One study described results dependent on gender with males performing better than females(Wavering, 1989), maybe due to advanced reasoning skills (Berg & Phillips, 1994).

Difficulties During Graphing

Difficulties with Graphing Conventions

Most studies reported that their participants had difficulties creating graphs that could be attributed to graphing conventions (n=34). Several studies noted more than one difficulty. As the studies varied extensively, we distinguished between three broad categories of difficulties. The first two types of difficulties are based on structural models that describe the graph construction process (Lachmayer et al., 2007). Graph construction difficulties are concerned with constructing the structure of the graphs, such as scaling the axes and assigning variables to the appropriate axes, whereas variable ordering difficulties describe problems, such as charting points at the correct locations in the graphs. For example, Watson et al. (2023) found that students sometimes had trouble scaling the axes when using TinkerPlots. Onwu (2014) reported that only 38% of the 366 junior high school students had could correctly determine the x- and y-coordinates of data points. An example graph for a variable translation difficulty can be seen in Figure 7. Data translation difficulties were found the least (n=11); they describe difficulties translating from one representation to another. For example, third graders seem to have trouble translating self-constructed tables into suitable graphical representations (Oslington et al., 2020). An overview of students' graphing difficulties due to the conventions can be seen in Figure 8.

Figure 7

Example of a graphical representation with variable ordering difficulties (based on Mevarech & Kramarsky, 1997)

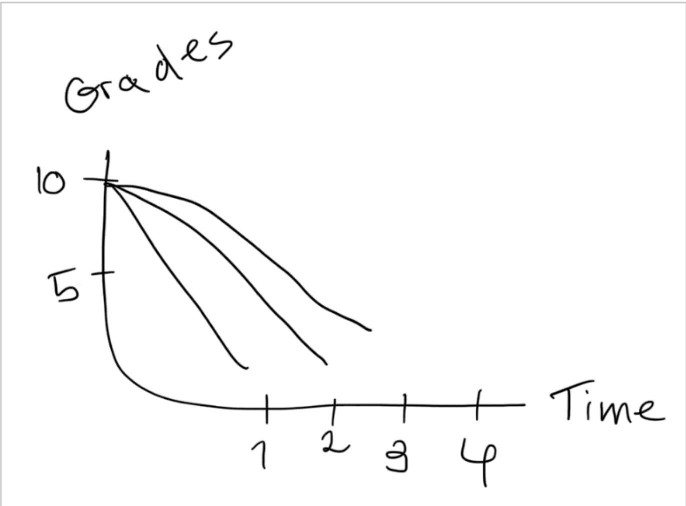
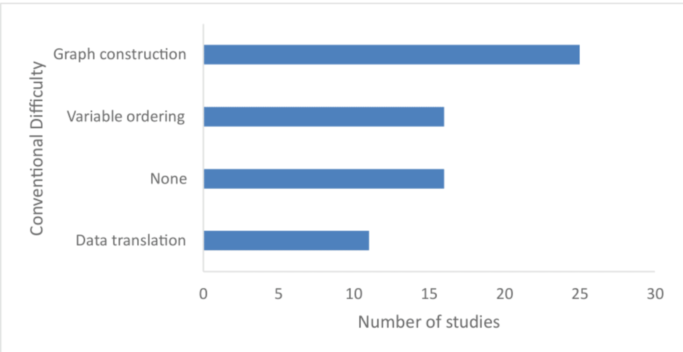


Figure 8

Overview of Students' Difficulties due to Graphing Conventions



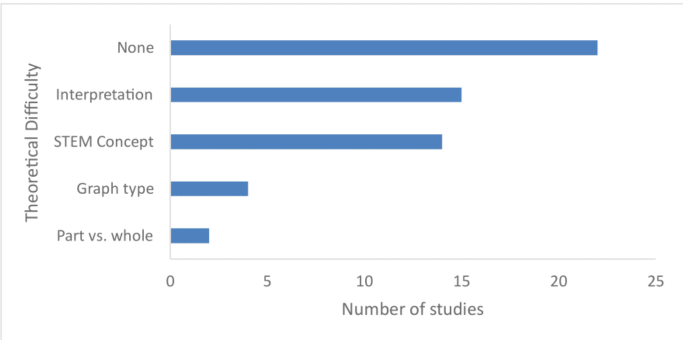
Note: Multiple mentions are possible.

Theoretical Difficulties During Graphing

Some students also faced theoretical difficulties (see Figure 9); however, almost half of the studies reported no difficulties (n=28). Conceptual understanding and interpretation were the most common theoretical difficulties described in the included studies. Fourteen studies reported difficulties connecting the graphs to the depicted STEM concept, such as graph-as-picture errors (Gerard et al., 2012). Other students had trouble interpreting the data, for example, to make predictions (Webb & Boltt, 1991). Four studies documented difficulties in choosing the correct graph type for a task (Gültepe, 2016; Jackson et al., 1992; Jackson et al., 1993; Ozmen et al., 2020), although this difficulty seemed to lessen with increasing experience (Karplus, 2006).

Figure 9

Overview of Students' Theoretical Difficulties



Note: Multiple mentions are possible.

Twenty studies reported both theoretical difficulties and difficulties associated with graph construction; for example, in a study with 10th-grade students, Dimas et al. (2018) reported that some students could not create a correct table for data from an oscillator experiment. The students also had trouble connecting the data to the concept of harmonic motion (Dimas et al., 2018). Interpretation difficulties were reported in combination with all types of difficulties. They were described in combination with graph construction difficulties (Branisa & Jenisova, 2015; Jackson et al., 1993; Wu & Krajcik, 2006), variable ordering difficulties (Harrison et al., 2019; Nurrahmawati et al., 2021; Pratt, 1995), or both (Dewi et al., 2018; Onwu, 1993). Interpretation-related difficulties were also found in combination with data translation difficulties (Aydın-Güç et al., 2022; Branisa & Jenisova, 2015; Gültepe, 2016; Nurrahmawati et al., 2021). Several studies reported difficulties in combination with the STEM concept, e.g., graph construction (Berg & Phillips, 1994; Dimas et al., 2018), variable ordering (Karplus, 1979), or both (Brasell & Rowe, 1993). Furthermore, conceptual difficulties were found in combination with data translation difficulties (Oslington et al., 2020). Two studies identified student difficulties related to finding the correct graph type in combination with graph construction and variable ordering difficulties (Jackson et al., 1992; Ozmen et al., 2020). One study reported a part vs. whole difficulty in combination with graph construction difficulties (Åberg-Bengtsson, 2006).

Discussion

This article presents a systematic review of empirical research on graphing in K–12 STEM education. A systematic search of three scientific databases found 12,945 records matching the search term. In total, we identified 50 relevant studies using our inclusion criteria, which were included in the review. Based on the codes presented above, we answer our research questions as follows. First, we summarize how graphing was implemented in the studies included in the literature review. Second, we condense the effectiveness of graphing as an instructional method, as reported in the studies. Third, we gather the reported student difficulties during graphing.

Implementation of Graphing in K-12 STEM Education

The theoretical foundation of graphing varied between studies: Many studies did not include a theoretical background, making an interpretation of the overarching results difficult because the study designs differed accordingly. A few studies included graphing from a cognitive perspective to construct a deeper understanding, e.g., by referring to the generation effect or the DeFT framework (Dimas et al., 2018; Rahmawati et al., 2020). In addition to the learning mechanisms during generative learning mentioned by Schmidgall et al. (2019), constructed graphs might offer additional cognitive functions by allowing learners to execute different cognitive strategies. For example, graphs can make inferences visible (Larkin & Simon, 1987). However, most studies investigate graphing owing to its relevance in professional or educational practices – that is, the aim is to learn graphing rather than using graphing as a learning method. Practices might vary between disciplines and between curricula, which in all likelihood played a role when investigating graphing in K–12 STEM education. Although most of the studies were conducted in the context of mathematics, a possible influence of the curriculum or the connection between graphing practices between disciplines could not be investigated in this review.

Regarding the specific aspects of graphing, line graphs were the most common type of graph constructed in our review. This could relate to the prevalence of bivariate data, which are often used to characterize relationships between two variables. Regarding the graphing method, there was no clear trend toward manual or tool-based graphing. However, a combination of tool-based and manual graphing seemed to facilitate graphing (Åberg-Bengtsson, 2006; Adams & Shrum, 1990; Ates & Stevens, 2003; English & Watson, 2015; Kramarski, 1999; Wu & Krajcik, 2006). Tool-based graphing was also used to analyze manually created graphs (Watson et al., 2022) or as an instructional method. For example, students who collected data via data acquisition probeware and used digital video analysis software in a physics course improved their graphing skills during the study (Struck & Yerrick, 2009).

Therefore, there was no typical study design that most of the studies used. However, an example of a common study design in this review would be an investigation of the effectiveness of graphing instruction for line graphs in the context of a regular class (during one or multiple lessons).

Effectiveness of Graphing as an Instructional Method

The second aim of this review was to outline the effectiveness of graphing as an instructional method, as described in the studies included in the literature review. This is challenging due to the differences in study design and methodology, e.g., tool-based vs. manual graphing and quantitative vs. qualitative analysis methods. Furthermore, several studies did not explicitly state a research hypothesis with clear results regarding the possible advantages of graphing.

Some studies reported that instruction helped students develop graphing skills (Gerard et al., 2012; Gultepe & Kilic, 2015; Meisadewi et al., 2017; Mevarech & Kramarsky, 1997; Struck & Yerrick, 2010; Wu & Krajcik, 2006). This is in line with Glazer (2011). However, the types of instruction varied between these studies and not all instructions referred specifically to graphing skills; other types of instructions, for example, in the context of specific topics (Struck & Yerrick, 2010; Wu & Krajcik, 2006) or scientific argumentation skills (Gultepe & Kilic, 2015), also facilitated students' graphing skills. This indicates that graphing skills can be improved using a broad range of instructions, provided that graphing is considered in some way during instruction.

Instruction was not only valuable to develop graphing skills but was also found to improve interpretation (Gerard et al., 2012; Gultepe & Kilic, 2015; Struck & Yerrick, 2010) and scientific process skills, such as developing hypotheses (Gultepe & Kilic, 2015).

Therefore, a focus on improving scientific process skills could be advantageous compared to traditional teaching (Gultepe & Kilic, 2015). Qualitative analyses reported similar results (Vitale et al., 2019; Wu & Krajcik, 2005). Due to the relevance of graphing for scientific inference (Gooding, 2010), this is an important aspect for teachers to consider during lesson planning.

Graphing skills seem to improve with grade level and age. In addition to Karplus (1979), six other studies considered this proposition (Garcia-Mila et al., 2014; Onwu, 1993; Padilla et al., 1986; Wavering, 1989; Webb & Boltt, 1991). This is related to mathematical understanding, which is also an important factor in graphing (Leinhardt et al., 1990). Two studies reported positive effects for manual graphing compared to tool-based graphing (Adams & Shrum, 1990; Branisa & Jenisova, 2015). Both studies compared manual graphing to graphing with (micro) computers. As there were 25 years between studies, and because technology evolved during this time, the benefit of manual graphing compared to letting a computer create a graph seems consistent, but the scant number of studies makes drawing conclusions difficult.

Several studies mention variables that might have a moderating effect on graphing skills, such as mathematical communication skills (Gardenia et al., 2021), statistical inference (Oslington et al., 2020), graph interpretation levels (Moritz, 2003), or cognitive development (Adams & Shrum, 1990), including mental structure (Berg & Phillips, 1994; Wavering, 1989) and scientific reasoning levels (Ates & Stevens, 2003). Unfortunately, due to the variance in the study designs and the low number of studies reporting moderators, a more detailed analysis of the effects of these moderators is not possible.

In conclusion, there seem to be multiple benefits of including graphing in K-12 STEM education (Glazer, 2011), such as improving graph interpretation skills. Graphing skills might also have the potential to facilitate scientific process skills. However, it should be noted that investigating the effects of graphing was not the main goal of most of the studies and there out of 50 included studies only nine had a control or comparison group. Most of the included studies investigated graphing in education from a practical perspective and focused on reporting the performance of their students.

Difficulties During Graphing

Out of 50 included studies, 42 reported graphing difficulties. Difficulties during graph construction were reported the most frequently. This included trouble with scaling (Åberg-Bengtsson, 2006) or labeling the axes (Berg & Phillips, 1994). Variable ordering, such as sketching data points at the correct coordinates (Mevarech & Kramarsky, 1997), was also common. Furthermore, data translation, for example, translating data from a table to a graph, seemed to cause students problems (Tairab & Khalaf Al-Naqbi, 2004). These results highlight the relevance of metarepresentational competence (diSessa, 2004; Rau, 2017). This is also reflected in students' strategies, such as constructing appropriate data visualizations using self-questioning and reflecting (Chang et al., 2024).

Theoretical difficulties were not reported as often but were also found repeatedly. The most common types of theoretical student difficulties were difficulties with interpretation (n=12) and concept (n=11). For example, students chose the wrong graph type for the data or STEM concept because they seemed to have problems connecting it to the context of the task and therefore could not construct a fitting graph (Jackson et al., 1992; Ozmen et al., 2020). Similarly, students seemed to have trouble determining the x- and y-coordinates (Dewi et al., 2018; Onwu, 1993), which could lead to "misunderstanding the graph" (p. 3, Dewi et al., 2018). A possible reason for missing graph interpretation and construction skills could be missing practice (Tairab & Al-Naqbi, 2004).

In total, 20 studies reported difficulties in both categories (see Supplementary Material). Therefore, theoretical difficulties might be related to construction difficulties. One study determined that "students with good levels of conceptual understanding were concluded to have strong graphing skills" (Gültepe, 2016, p. 53), whereas the opposite was found for students with low conceptual understanding. This connection between construction and theoretical difficulties is in line with Duval (2006) who considered translations within one register to be a possible cause of comprehension difficulties.

In summary, difficulties during the construction of a graph were consistently observed. Many of the included studies reported theoretical difficulties related to interpreting the data as well as convention-based difficulties during construction. This is in line with previous reviews that have also reported student difficulties; however, to our knowledge, none have specifically analyzed this connection (Clement, 1985; Leinhardt et al., 1990; Boels et al., 2019). These results indicate that students might have difficulty during graphing not only due to trouble understanding the conventions but also because they might not be able to correctly understand the data and its relevance and therefore might not know how to best display it.

Implications for Practice

Based on our review results, we can join previous research (e.g., Glazer, 2011; Leinhardt et al., 1990) in emphasizing the relevance of graphing and encouraging teachers to add graphing activities to their lessons. The difficulties students experience during graphing exemplify the importance of graphing instruction. Graphing activities during the instruction of scientific argumentation (Gultepe & Kilic, 2015) might be an effective tool for improving science learning (Gerard et al., 2012). Similarly, comparing students who plotted the given data with those who plotted imagined data indicated that the first activity led to more integration between students' ideas and scientific evidence and, therefore, to a deeper exploration of the graphs (Vitale et al., 2019). This relates to the relevance of using authentic, real-world data that students can relate to when instructing data literacy (Friedrich et al., 2024). However, students analyzing the provided data graphed more accurately (Vitale et al., 2019). This suggests that students might benefit (1) from graphing their perception of scientific principles before traditional instruction about a topic and (2) from an in-depth analysis of the provided data.

Limitations and Future Research

Limitations

This research has several limitations. The most relevant one might be the large number of studies found in the initial search. Due to the large number of studies, we had to limit the included studies to those with a precise focus on graphing. However, there are many more studies in which students construct graphs—for example, as part of studies more broadly analyzing scientific inquiry skills—that might have provided valuable contributions to the topic. Furthermore, we included only peer-reviewed studies and no gray literature as quality control, and all included studies were written in English. As the publication years ranged from 1979 to 2024, these criteria may have changed and influenced our study selection. The choice of codes could have also led to limitations. The included studies often reported specific student difficulties. Due to the number of specific difficulties reported, we considered only overarching categories of difficulties in this review. The coding of these difficulties was often a part of the discussion between raters. Although raters always reached an agreement, a more fine-grained analysis might provide further insights.

Future Research

Graphing as a method was often not based on theoretical research, but was justified due to its use in school (e.g., Moritz, 2003). We believe that more hypothesis-based testing grounded in theory in future research could provide valuable insights into the specific benefits of graphing in education. Additionally, more longitudinal research starting with younger students could generate a deeper understanding of the development of graphing skills. Further information about the influence of possible moderating variables could help improve instruction for students. Additionally, considering the function of graphing – whether students graph for themselves during learning, create a graph for others to explain something, or use a graph to compute a result during problem-solving – could play a role in interpreting the results. A future meta-analysis of empirical qualitative studies should take the effectiveness of graphing in these contexts into account.

Conclusion

Understanding graphical representations of data is an important skill, and graphing is a relevant part of graph interpretation competence. Therefore, graphing has been examined in many studies. In this systematic literature review of 50 studies, we aimed to provide an overview of current research findings on graphing in K-12 education. We focus on the possible benefits of and difficulties faced by students during graphing, as well as how graphing is implemented in research studies. Studies have frequently analyzed graphing in the context of a course and have often found graphing instruction beneficial for not only improving graphing skills but also graph interpretation. However, the students experienced various graphing difficulties, such as correctly sketching data points and interpreting the graph. Therefore, the difficulties encountered during the construction of graphs might be related to the theoretical understanding of the data. Consequently, both types of difficulties should be considered during instruction, for example, first by graphing students' perceptions of a scientific phenomenon and then independently revising the graph based on the data measured during an experiment.

List of Abbreviations

AERA: American Educational Research Association

CTML: Cognitive Theory of Multimedia Learning

DeFT framework: Design, Function, and Task framework

EARLI: European Association for Research on Learning and Instruction

ESERA: European Science Education Research Association

ICAP framework: Interactive, Constructive, Active, and Passive framework

ICLS: International Conference of the Learning Sciences

ICOTS: International Conference on Teaching Statistics

ITPC: Integrated Model of Text and Picture Comprehension

NARST: National Association for Research in Science Teaching

STEM: Science, Technology, Engineering, and Mathematics

PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

Declarations

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Competing interests

The authors declare they have no competing interests.

Funding

The project on which this report is based was funded by the German Federal Ministry of Education and Research (BMBF) under the funding code 16MF1006B. The responsibility for the content of this publication lies with the authors.

Authors' contributions

Conceptualization, V.R., S.M, S.K., S.B., M.V., R.B., and J.K. ; methodology, V.R., S.M, S.K., S.B., M.V., R.B., and J.K.; formal analysis, V.R.; investigation, V.R., D.T., validation, V.R., D.T., S.M, S.K., S.B., M.V., R.B., and J.K; data curation, V.R., D.T.; writing—original draft preparation, V.R.; writing—review and editing, V.R., D.T., S.M, S.K., S.B., M.V., R.B., and J.K; visualization, V.R.; supervision, S.M., S.K., and J.K.; project administration, S.M., funding acquisition, J.K..

References

*Åberg-Bengtsson, L. (2006). "Then you can take half… almost"—Elementary students learning bar graphs and pie charts in a computer-based context. *The Journal of Mathematical Behavior*, *25*(2), 116-135. https://doi.org/10.1016/j.jmathb.2006.02.007

*Adams, D. D., & Shrum, J. W. (1990). The effects of microcomputer-based laboratory exercises on the acquisition of line graph construction and interpretation skills by high school biology students. *Journal of Research in Science Teaching*, *27*(8), 777-787. https://doi.org/10.1002/tea.3660270807

Ainley, J. (2000). Transparency in graphs and graphing tasks: an iterative design process. *The Journal of Mathematical Behavior*, *19*(3), 365-384. https://doi.org/10.1016/s0732-3123(00)00052-3

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, *33*(2–3), 131-152. https://doi.org/10.1016/S0360-1315(99)00029-9

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183-198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, *333*(6046), 1096-1097. https://doi.org/10.1126/science.1204153

Angra, A., & Gardner, S. M. (2016). Development of a framework for graph choice and construction. *Advances in Physiology Education*, *40*(1), 123-128. https://doi.org/10.1152/advan.00152.2015

*Arteaga, P., Díaz-Levicoy, D., & Batanero, C. (2020). Chilean Primary School Children's Understanding of Statistical Graphs. *Acta Scientiae*, *22*(5), 2-24. https://doi.org/10.17648/acta.scientiae.5884

*Ates, S., & Stevens, J. T. (2003). Teaching Line Graphs to Tenth Grade Students Having Different Cognitive Developmental Levels by Using Two Different Instructional Modules. *Research in Science & Technological Education*, *21*(1), 55-66. https://doi.org/10.1080/02635140308339

*Aydın-Güç, F., Özmen, Z. M., & Güven, B. (2022). Difficulties scatter plots pose for 11th grade students. *The Journal of Educational Research, 115*(5), 298–314. https://doi.org/10.1080/00220671.2022.2128018

Bahtaji, M. A. A. (2020). Improving students graphing skills and conceptual understanding using explicit Graphical Physics Instructions. *Cypriot Journal of Educational Sciences*, *15*(4), 843-853. https://files.eric.ed.gov/fulltext/EJ1266848.pdf

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2001). Toward a model of learning data representations. In Moore, J. & Stenning, K. (Eds.), *Proceedings of the 23rd annual Conference of the Cognitive Science Society* (pp. 45-50). Lawrence Erlbaum Associates. https://escholarship.org/uc/item/2qb8m54p

Bayri, N. G., & Kurnaz, M. A. (2015). Investigation of the secondary school students' transition situations between different representation types. *Turkish Online Journal of Educational Technology*, *2015*(2), 337-341.

https://www.researchgate.net/profile/Mehmet-Kurnaz-2/publication/297766626_Investigation_of_the_secondary_school_students_transition_situations_between_different_representation_types/links/6094ff27299bf1ad8d81da71/Investigation-of-the-secondary-school-students-transition-situations-between-different-representation-types.pdf

*Berg, C. A., & Phillips, D. G. (1994). An investigation of the relationship between logical thinking structures and the ability to construct and interpret line graphs. *Journal of Research in Science Teaching, 31*(4), 323-344. https://doi.org/10.1002/tea.3660310404

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory and Cognition*, *35*(2), 201-210. https://doi.org/10.3758/BF03193441

Biehler, R. (2006). Working styles and obstacles: Computer-supported collaborative learning in statistics. In Rossman, A. & Chance, B. (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (pp. 1-6). International Association for Statistical Education, International Statistical Institute. https://www.ime.usp.br/~abe/ICOTS7/Proceedings/PDFs/InvitedPapers/2D2_BIEH.pdf

Boels, L., Bakker, A., Van Dooren, W., & Drijvers, P. (2019). Conceptual difficulties when interpreting histograms: A review. *Educational Research Review*, *28*, 1-23. https://doi.org/10.1016/j.edurev.2019.100291

*Branisa, J., & Jenisova, Z. (2015). The Use of Computer Aided Experiment in Enhancing the Ability of Students to Understand the Graphical Presentation of Chemical Processes. *Procedia - Social and Behavioral Sciences*, *197*, 2229-2235. https://doi.org/10.1016/j.sbspro.2015.07.364

*Brasell, H. M., & Rowe, M. B. (1993). Graphing Skills Among High School Physics Students. *School Science and Mathematics*, *93*(2), 63-70. https://doi.org/10.1111/j.1949-8594.1993.tb12196.x

Castro-Alonso, J. C., Wong, M., Adesope, O. O., Ayres, P., & Paas, F. (2019). Gender imbalance in instructional dynamic versus static visualizations: A meta-analysis. *Educational Psychology Review*, *31*(2), 361-387. https://doi.org/10.1007/s10648-019-09469-1

Chalkiadaki, A. (2018). A systematic literature review of 21st century skills and competencies in primary education. *International Journal of Instruction*, *11*(3), 1-16. https://doi.org/10.12973/iji.2018.1131a

Chang, H. Y. (2018). Students' representational competence with drawing technology across two domains of science. *Science Education*, *102*(5), 1129-1149. https://doi.org/10.1002/sce.21457

Chang, H.-Y., Chang, Y.-J., & Tsai, M.-J. (2024). Strategies and difficulties during students' construction of data visualizations. *International Journal of STEM Education, 11*(1). https://doi.org/10.1186/s40594-024-00463-w

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219-243. https://doi.org/10.1080/00461520.2014.965823

Clement, J. (1985). Misconceptions in graphing. *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education*, *1*, 369-375. http://people.umass.edu/clement/pdf/Misconceptions%20in%20Graphing.pdf

Clement, J. (1989). The concept of variation and misconceptions in Cartesian graphing. *Focus on Learning Problems in Mathematics*, *11*, 77-87. https://people.umass.edu/~clement/pdf/concept_of_variation.pdf

Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, *9*(4), 343-363. https://doi.org/10.1016/S0959-4752(98)00051-6

Cromley, J. G., Du, Y., & Dane, A. P. (2020). Drawing-to-learn: Does meta-analysis show differences between technology-based drawing and paper-and-pencil drawing? *Journal of Science Education and Technology*, *29*(2), 216-229. https://doi.org/10.1007/s10956-019-09807-6

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, *18*(5), 382-393. https://doi.org/10.5951/jresematheduc.18.5.0382

*Detiana, Johar, R., & Mailizar. (2020). Students' mathematical communication in discovering the concept of graphs of trigonometric functions through instructional videos. *Journal of Physics: Conference Series*, *1460*(1), 012020. https://doi.org/10.1088/1742-6596/1460/1/012020

de Vries, E., & Masclet, C. (2013). A framework for the study of external representations in collaborative design settings. *International Journal of Human-Computer Studies, 71*(1), 46–58. https://doi.org/10.1016/j.ijhcs.2012.07.005

*Dewi, A. R. C., Putra, N. M. D., & Susilo. (2018). Analysis of graphic representation ability in oscillation phenomena. *Journal of Physics: Conference Series*, *983*, 012024. https://doi.org/10.1088/1742-6596/983/1/012024

Díaz-Levicoy, D., Osorio, M., & Ruz, F. (2018). Activities and types of statistical graphs in textbooks of primary education in Perú. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan* (Vol. 10). International Statistical Institute. http://funes.uniandes.edu.co/12567/1/ICOTS10_C170.pdf

*Dimas, A., Suparmi, A., Sarwanto, & Nugraha, D. A. (2018). Analysis multiple representation skills of high school students on simple harmonic motion. *AIP Conference Proceedings, 2014*(1), 020131. https://doi.org/10.1063/1.5054535

diSessa, A. A. (2004). Metarepresentation: Native Competence and Targets for Instruction. *Cognition and Instruction, 22*(3), 293–331. https://doi.org/10.1207/s1532690xci2203_2

diSessa, A. A., & Sherin, B. L. (2000). Meta-representation: an introduction. *The Journal of Mathematical Behavior, 19*(4), 385–398. https://doi.org/10.1016/s0732-3123(01)00051-7

Dossey, J., McCrone, S., & Halvorsen, K. (2016). *Mathematics Education in the United States. A capsule summary fact book*. National Council of Teachers of Mathematics. Retrieved from https://www.nctm.org/uploadedFiles/About/MathEdInUS2016.pdf

Duval, R. (2006). A Cognitive Analysis of Problems of Comprehension in a Learning of Mathematics. *Educational Studies in Mathematics, 61*(1–2), 103–131. https://doi.org/10.1007/s10649-006-0400-z

Eichler, A., & Vogel, M. (2012). *Leitidee Daten und Zufall* (2nd edition). Springer Spektrum Wiesbaden. https://doi.org/10.1007/978-3-658-00118-6

Engel, J. (2018). *Anwendungsorientierte Mathematik: Von Daten zur Funktion*. Springer. https://doi.org/10.1007/978-3-662-55487-6

*English, L. D., & Watson, J. M. (2015). Exploring variation in measurement as a foundation for statistical thinking in the elementary school. *International Journal of STEM Education*, *2*(1). https://doi.org/10.1186/s40594-015-0016-x

*English, L. D. (2022). Fifth-grade Students' Quantitative Modeling in a STEM Investigation. *Journal for STEM Education Research, 5*(2), 134–162. https://doi.org/10.1007/s41979-022-00066-6

*English, L. D. (2023). Linking Probability and Statistics in Young Students' Reasoning with Chance. *Statistics Education Research Journal, 22*(2), 7. https://doi.org/10.52041/serj.v22i2.418

Ergül, N. R. (2018). Pre-service science teachers' construction and interpretation of graphs. *Universal Journal of Educational Research*, *6*(1), 139-144. https://doi.org/10.13189/ujer.2018.060113

*Fielding-Wells, J. (2018). Dot plots and hat plots: supporting young students emerging understandings of distribution, center and variability through modeling. *ZDM, 50*(7), 1125-1138. https://doi.org/10.1007/s11858-018-0961-1

Fiorella, L., & Zhang, Q. (2018). Drawing boundary conditions for learning by drawing. *Educational Psychology Review*, *30*(3), 1115-1137. https://doi.org/10.1007/s10648-018-9444-8

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2317*, 18-30. https://doi.org/10.1007/3-540-46037-3_3

Friedrich, A., Schreiter, S., Vogel, M., Becker-Genschow, S., Brünken, R., Kuhn, J., Lehmann, J., & Malone, S. (2024). What shapes statistical and data literacy research in K-12 STEM education? A systematic review of metrics and instructional strategies. *International Journal of STEM Education, 11*(1). https://doi.org/10.1186/s40594-024-00517-z

Gaona, J., Hernández, R., & Bravo, F. G. y V. (2021). *Influence of a function's coefficients and feedback of the mathematical work when reading a graph in an online assessment system.* arXiv. http://arxiv.org/abs/2107.11448

*García-García, J., & Dolores-Flores, C. (2019). Pre-university students' mathematical connections when sketching the graph of derivative and antiderivative functions. *Mathematics Education Research Journal*, *33*(1), 1-22. https://doi.org/10.1007/s13394-019-00286-x

*Garcia-Mila, M., Marti, E., Gilabert, S., & Castells, M. (2014). Fifth through eighth grade students' difficulties in constructing bar graphs: Data organization, data aggregation, and integration of a second variable. *Mathematical Thinking and Learning*, *16*(3), 201-233. https://doi.org/10.1080/10986065.2014.921132

*Gardenia, N., Herman, T., Juandi, D., Dahlan, T., & Kandaga, T. (2021, March). Analysis of mathematical communication skills of class 8 students on two-variable linear equation systems (SPLDV) concept. In *Journal of Physics: Conference Series (1806, 1)*, 1-7. IOP Publishing. https://doi.org/10.1088/1742-6596/1806/1/012073

*Gerard, L., Zertuche, A., & Linn, M. (2012, July 2 - 6). *Learning to Graph: A Comparison Study of Using Probe or Draw Tools in a Web-Based Learning Environment* [Conference poster]. International Conference of the Learning Sciences 2012, Sydney, Australia. https://repository.isls.org/bitstream/1/2380/1/569-570.pdf

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, *47*(2), 183-210. https://doi.org/10.1080/03057267.2011.605307

Gooding, D. (2004). Cognition, Construction and Culture: Visual Theories in the Sciences. *Journal of Cognition and Culture, 4*(3–4), 551–593. https://doi.org/10.1163/1568537042484896

Gooding, D. C. (2010). Visualizing Scientific Inference. *Topics in Cognitive Science, 2*(1), 15–35. Portico. https://doi.org/10.1111/j.1756-8765.2009.01048.x

*Gultepe, N., & Kilic, Z. (2015). Effect of Scientific Argumentation on the Development of Scientific Process Skills in the Context of Teaching Chemistry. *International Journal* *of Environmental and Science Education*, *10*(1), 111-132. https://files.eric.ed.gov/fulltext/EJ1060989.pdf

*Gültepe, N. (2016). Reflections on High School Students' Graphing Skills and Their Conceptual Understanding of Drawing Chemistry Graphs. *Educational Sciences: Theory & Practice*, *16*, 53-81. https://files.eric.ed.gov/fulltext/EJ1101172.pdf

*Harrison, E., McBride, B., & Linn, M. C. (2019). Revision Analysis of Students' Position-Time Graphs. In K. Lund, G. P. Niccolai, E. Lavoué, C. E. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *International Society for the Learning Sciences, Computer-Supported Collaborative Learning (CSCL) Annual Meeting*, *2*, 927-928. Lyon, France: International Society of the Learning Sciences. https://par.nsf.gov/servlets/purl/10181648

Harsh, J. A., & Schmitt-Harsh, M. (2016). Instructional strategies to develop graphing skills in the college science classroom. *The American Biology Teacher*, *78*(1), 49-56. https://doi.org/10.1525/abt.2016.78.1.49

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*(4), 684–689. https://doi.org/10.1037//0022-0663.91.4.684

Hug, B., & McNeill, K. L. (2008). Use of First-hand and Second-hand Data in Science: Does data type influence classroom conversations*?. International Journal of Science Education, 30*(13), 1725-1751. https://doi.org/10.1080/09500690701506945

*Jackson, D. F., Berger, C. F., & Edwards, B. J. (1992). Computer-assisted thinking tools: Problem solving in graphical data analysis. *Journal of Educational Computing Research*, *8*(1), 43-67. https://doi.org/10.2190/ECB0-MV0A-8AP9-N47Q

*Jackson, D. F., Edwards, B. J., & Berger, C. F. (1993). Teaching the design and interpretation of graphs through computer-aided graphical data analysis. *Journal of Research in Science Teaching*, *30*(5), 483-501. https://doi.org/10.1002/tea.3660300507

Kanzaki, N., & Miwa, K. (2012). An experimental investigation of consistency of explanation and graph representation. In Myiaki N., Peebles, D., & Cooper, R. P. (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34), 557-562. Cognitive Science Society. https://escholarship.org/uc/item/5rd400c8

*Karplus, R. (1979). Continuous Functions: Students' Viewpoints, *European Journal of Science Education, 1*(4), 397-415. http://dx.doi.org/10.1080/0140528790010404

Kramarski, B. (1999). The Study Of Graphs By Computers: Is Easier Better? Educational Media International*, *36*(3), 203-209. https://doi.org/10.1080/0952398990360306

Kultusministerkonferenz (KMK) (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife* [Educational standards for the subject mathematics for the general qualification for university entrance]. Berlin: KMK. Retrieved from https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf

Kohnle, A., Ainsworth, S. E., & Passante, G. (2020). Sketching to support visual learning with interactive tutorials. *Physical Review Physics Education Research*, *16*(2), 1-21. https://doi.org/10.1103/PhysRevPhysEducRes.16.020139

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2014). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325. https://doi.org/10.1007/s10649-013-9529-8

Lachmayer, S., Nerdel, C., & Prechtl, H. (2007). Modellierung kognitiver Fähigkeiten beim Umgang mit Diagrammen im naturwissenschaftlichen Unterricht [Modelling of cognitive abilities regarding the handling of graphs in science education]. *Zeitschrift für Didaktik der Naturwissenschaften*, *13*, 161-180. https://archiv.ipn.uni-kiel.de/zfdn/pdf/008_Lachmayer_13.pdf

Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, *11*(1), 65-100. https://doi.org/10.1016/S0364-0213(87)80026-5

Lee, J. J., & Lee, T. R. (2018). Developing a mobile learning course of elementary statistics for k-mooc by using modules of dynamic graphical software for teaching statistics, eStat. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan* (Vol. 10). International Statistical Institute. https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_9M1.pdf?1531364303

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, *60*(1), 1-64. https://doi.org/10.3102/00346543060001001

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning*, *41*, 31-48. https://doi.org/10.1017/CBO9780511816819

Mayer, R. (Ed.). (2014). *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

*Meisadewi, N., Anggraeni, S., & Supriatno, B. (2017, March). Improving Students' Graphing Skills through Quantitative-Based Lab Activities. In *IOP Conference*

Series: Materials Science and Engineering, 180(1), 1-5. IOP Publishing. https://doi.org/10.1088/1757-899X/180/1/012245

*Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics*, *32*(3), 229-263. https://doi.org/10.1023/A:1002965907987

*Moritz, J. (2003). Constructing Coordinate Graphs: Representing Corresponding Ordered Values with Variation in Two-Dimensional Space. *Mathematics Education Research Journal*, *15*(3), 226-251. https://files.eric.ed.gov/fulltext/EJ776330.pdf

*Ng, W., & Nicholas, H. (2011). Insights into Students' Thinking with Handheld Computers. *Mobile Technologies and Handheld Devices for Ubiquitous Learning*, 79-98. https://doi.org/10.4018/978-1-61692-849-0.

Nixon, R. S., Godfrey, T. J., Mayhew, N. T., & Wiegert, C. C. (2016). Undergraduate student construction and interpretation of graphs in physics lab activities. *Physical Review Physics Education Research*, *12*(1), 1-19. https://doi.org/10.1103/PhysRevPhysEducRes.12.010104

*Nurrahmawati, N., Sa'dijah, C., Sudirman, S., & Muksar, M. (2021). Assessing students' errors in mathematical translation: From symbolic to verbal and graphic representations. *International Journal of Evaluation and Research in Education (IJERE), 10*(1), 115. https://doi.org/10.11591/ijere.v10i1.20819

*Onwu, G. O. M. (1993). Line graphing ability of some Nigerian secondary science students. *International Journal of Mathematical Education in Science and Technology*, *24*(3), 385-390. https://doi.org/10.1080/0020739930240308

*Oslington, G., Mulligan, J., & Van Bergen, P. (2020). Third-graders' predictive reasoning strategies. *Educational Studies in Mathematics*, *104*, 1-20. https://doi.org/10.1007/s10649-020-09949-0

Ott, N., Brünken, R., Vogel, M., & Malone, S. (2018). Multiple symbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic. *Learning & Instruction, 58*, 88–105. https://doi.org/10.1016/j.learninstruc.2018.04.010

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews, 5*(1). https://doi.org/10.1186/s13643-016-0384-4

*Ozmen, Z. M., Guven, B., & Kurak, Y. (2020). Determining the graphical literacy levels of the 8th grade students. *Eurasian Journal of Educational Research*, *20*(86), 269-292. https://dergipark.org.tr/en/download/article-file/1079986

*Padilla, M. J., McKenzie, D. L., & Shaw, E. L. (1986). An Examination of the Line Graphing Ability of Students in Grades Seven Through Twelve. *School Science and Mathematics*, *86*(1), 20-26. Portico. https://doi.org/10.1111/j.1949-8594.1986.tb11581.x

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 1-9. https://doi.org/10.1016/j.ijsu.2021.105906

Parnafes, O., & Disessa, A. (2004). Relations between types of reasoning and computational representations. *International Journal of Computers for Mathematical Learning*, *9*(3), 251-280. https://doi.org/10.1007/s10758-004-3794-7

*Pols, C. F. J. (2019). Students' ability to analyse empirical data in practical work. *Journal of Physics: Conference Series*, *1287*(1), 012001. https://doi.org/10.1088/1742-6596/1287/1/012001

*Pospiech, G., Geyer, M. A., Ceuppens, S., De Cock, M., Deprez, J., Dehaene, W., Michelini, M., & Stefanel, A. (2019). Role of graphs in the mathematization process in physics education. *Journal of Physics: Conference Series*, *1287*(1), 012014. https://doi.org/10.1088/1742-6596/1287/1/012014

*Pratt, D. (1995). Young children's active and passive graphing. *Journal of Computer Assisted Learning*, *11*(3), 157-169. https://doi.org/10.1111/j.1365-2729.1995.tb00130.x

Program for International Student Assessment (PISA). (2021). PISA 2021 Rahmenkonzeption für Mathematik [PISA 2021 Framework guidelines for mathematics]. Retrieved from https://pisa2022-maths.oecd.org/

*Rahmawati, D., & Ahmad, J. (2020). Characteristics of Mathematical Representation Translation from Verbal to Graph. *International Journal of Innovation, Creativity and Change*, 401-420. https://www.ijicc.net/images/vol_13/Iss_2/SC47_Rahmawati_2020_E_R.pdfhttps://www.ijicc.net/images/vol_13/Iss_2/SC47_Rahmawati_2020_E_R.pdf

Rau, M. A. (2016). Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. *Educational Psychology Review, 29*(4), 717–761. https://doi.org/10.1007/s10648-016-9365-3

Roth, W., & Bowen, G. M. (2001). Professionals read graphs: A semiotic analysis. *Journal for Research in Mathematics Education*, *32*(2), 159-194. https://doi.org/10.2307/749672

Roth, W.-M., & Hwang, S. (2006). On the relation of abstract and concrete in scientists' graph interpretations: A case study. *The Journal of Mathematical Behavior, 25*(4), 318–333. https://doi.org/10.1016/j.jmathb.2006.11.005

*Saldanha, L., & McAllister, M. (2016). Building Up the Box Plot as a Tool for Representing and Structuring Data Distributions: An Instructional Effort Using Tinkerplots and Evidence of Students' Reasoning. *The Teaching and Learning of Statisti*cs, 235-245. https://doi.org/10.1007/978-3-319-23470-0_29

Schmidgall, S. P., Eitel, A., & Scheiter, K. (2019). Why do learners who draw perform well? Investigating the role of visualization, generation and externalization in learner-generated drawing. *Learning and Instruction*, *60*, 138-153. https://doi.org/10.1016/j.learninstruc.2018.01.006

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 49-69). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819

Schultheis, E. H., Kjelvik, M. K., Snowden, J., Mead, L., & Stuhlsatz, M. A. M. (2022). Effects of Data Nuggets on Student Interest in STEM Careers, Self-efficacy in Data Tasks, and Ability to Construct Scientific Explanations. *International Journal of Science and Mathematics Education, 21*(4), 1339–1362. https://doi.org/10.1007/s10763-022-10295-1

Schüller, K., Busch, P., & Hindinger, C. (2019). Future Skills: Ein Framework für Data Literacy [Future Skills: A framework for data literacy]. *Hochschulforum Digitalisierung*, *47,* 1-116. Retrieved from: https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr_47_DALI_Kompetenzrahmen_WEB.pdf

Shah, P., & Freedman, E. G. (2011). Bar and Line Graph Comprehension: An Interaction of Top-Down and Bottom-Up Processes. *Topics in Cognitive Science, 3*(3), 560–578. Portico. https://doi.org/10.1111/j.1756-8765.2009.01066.x

*Stephens, A. L. (2024). From graphs as task to graphs as tool. *Journal of Research in Science Teaching, 61*(5), 1206–1233. Portico. https://doi.org/10.1002/tea.21930

Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction*, *13*(2), 191-203. https://doi.org/10.1016/S0959-4752(02)00020-8

*Struck, W., & Yerrick, R. (2010). The effect of data acquisition-probeware and digital video analysis on accurate graphical representation of kinetics in a high school physics class. *Journal of Science Education and Technology*, *19*, 199-211. https://doi.org/10.1007/s10956-009-9194-y

Subali, B., Rusdiana, D., Firman, H., Kaniawati, I., & Ellianawati, E. (2017). Computer-based experiment of free fall movement to improve the graphical literacy. *Jurnal Pendidikan IPA Indonesia*, *6*(1), 41-48. https://doi.org/10.15294/jpii.v6i1.8750

*Tairab, H. H., & Khalaf Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs? *Journal of Biological Education*, *38*(3), 127-132. https://doi.org/10.1080/00219266.2004.9655920

Tversky, B. (1997). Cognitive principles of graphic displays. In *AAAI 1997 Fall Symposium on Reasoning with Diagrammatic Representations* (pp. 116-124). https://cdn.aaai.org/Symposia/Fall/1997/FS-97-03/FS97-03-015.pdf

Utrecht University. (2021). ASReview Lab (0.18) [computer software]. Retrieved from https://asreview.nl

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Wie jdema, F., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, *3*(2), 125-133. https://doi.org/10.1038/s42256-020-00287-7

van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*(4), 285-325. https://doi.org/10.1007/s10648-005-8136-3

Vevea, J. L., Coburn, K., & Sutton, A. (2019). Publication bias. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *The Handbook of Research Synthesis and Meta-analysis* (3rd ed.) (pp. 383-429). New York: Russell Sage Foundation. https://doi.org/10.7758/9781610448864

*Vitale, J. M., Applebaum, L., & Linn, M. C. (2019). Coordinating between graphs and science concepts: Density and buoyancy. *Cognition and Instruction*, *37*(1), 38-72. https://doi.org/10.1080/07370008.2018.1539736

von Kotzebue, L., Gerstl, M., & Nerdel, C. (2014). Common mistakes in the construction of diagrams in biological contexts. *Research in Science Education*, *45*(2), 193-213. https://doi.org/10.1007/s11165-014-9419-9

*Watson, J., Fitzallen, N., Wright, S., & Kelly, B. (2022). Characterizing Student Experience of Variation Within a STEM Context: Improving Catapults. *Statistics Education Research Journal, 21*(1), 11. https://doi.org/10.52041/serj.v21i1.7

*Wavering, M. J. (1989). Logical reasoning necessary to make line graphs. *Journal of Research in Science Teaching*, *26*(5), 373-379. https://doi.org/10.1002/tea.3660260502

*Webb, P., & Boltt, G. (1991). High school pupils' and first-year university students' responses to questions based on data analysis of an ecological case study. *Journal of Biological Education*, *25*(2), 119-122. https://doi.org/10.1080/00219266.1991.9655189

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology, 13*(4), e1002128. https://doi.org/10.1371/journal.pbio.1002128

Wittrock, M. C. (1974). A generative model of mathematics learning. *Journal for Research in Mathematics Education*, *5*(4), 181-196. https://doi.org/10.2307/748845

Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist*, *27*(4), 531-541. https://doi.org/10.1207/s15326985ep2704_8

*Wu, H. K., & Krajcik, J. S. (2006). Inscriptional practices in two inquiry-based classrooms: A case study of seventh graders' use of data tables and graphs. *Journal of Research in Science Teaching, 43*(1), 63-95. https://doi.org/10.1002/tea.20092

Wu, S. P., & Rau, M. A. (2019). How students learn content in science, technology, engineering, and mathematics (STEM) through drawing activities. *Educational Psychology Review*, *31*(1), 87-120. https://doi.org/10.1007/s10648-019-09467-3

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory and Cognition*, *27*(6), 1073-1079. https://doi.org/10.3758/BF03201236

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*(2), 179–217. https://doi.org/10.1016/s0364-0213(99)80022-6

Zhang, Q., & Fiorella, L. (2021). Learning by drawing: When is it worth the time and effort? *Contemporary Educational Psychology*, *66*, 101990. https://doi.org/10.1016/j.cedpsych.2021.101990

Zhang, J., & Norman, D. A. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science, 18*(1), 87–122. Portico. https://doi.org/10.1207/s15516709cog1801_3

3. Study 2: A Literature Review Comparing Experts' and Non-Experts' Visual Processing of Graphs during Problem-Solving and Learning

Reference: Ruf, V., Horrer, A., Berndt, M., Hofer, S. I., Fischer, F., Fischer, M. R., Zottmann, J. M., Kuhn, J., & Küchemann, S. (2023). A Literature Review Comparing Experts' and Non-Experts' Visual Processing of Graphs during Problem-Solving and Learning. *Education Sciences, 13*(2), 216. https://doi.org/10.3390/educsci13020216 education sciences

MDPI

Review

A Literature Review Comparing Experts' and Non-Experts' Visual Processing of Graphs during Problem-Solving and Learning

Verena Ruf [1,2,*], Anna Horrer [3], Markus Berndt [3], Sarah Isabelle Hofer [4], Frank Fischer [4], Martin R. Fischer [3], Jan M. Zottmann [3], Jochen Kuhn [2] and Stefan Küchemann [2]

1 Department of Electrical and Computer Engineering, RPTU Kaiserslautern, 67663 Kaiserslautern, Germany
2 Faculty of Physics, LMU Munich, 80539 Munich, Germany
3 Institute of Medical Education, University Hospital, LMU Munich, 80336 Munich, Germany
4 Department of Psychology, LMU Munich, 80539 Munich, Germany
* Correspondence: v.ruf@physik.uni-muenchen.de

Abstract: The interpretation of graphs plays a pivotal role in education because it is relevant for understanding and representing data and comprehending concepts in various domains. Accordingly, many studies examine students' gaze behavior by comparing different levels of expertise when interpreting graphs. This literature review presents an overview of 32 articles comparing the gaze behavior of experts and non-experts during problem-solving and learning with graphs up to January 2022. Most studies analyzed students' dwell time, fixation duration, and fixation count on macro- and meso-, as well as on micro-level areas of interest. Experts seemed to pay more attention to relevant parts of the graph and less to irrelevant parts of a graph, in line with the information-reduction hypothesis. Experts also made more integrative eye movements within a graph in terms of dynamic metrics. However, the determination of expertise is inconsistent. Therefore, we recommend four factors that will help to better determine expertise. This review gives an overview of evaluation strategies for different types of graphs and across various domains, which could facilitate instructing students in evaluating graphs.

Keywords: literature review; eye tracking; STEM education; graphical representation; expertise

Citation: Ruf, V.; Horrer, A.; Berndt, M.; Hofer, S.I.; Fischer, F.; Fischer, M.R.; Zottmann, J.M.; Kuhn, J.; Küchemann, S. A Literature Review Comparing Experts' and Non-Experts' Visual Processing of Graphs during Problem-Solving and Learning. *Educ. Sci.* **2023**, *13*, 216. https://doi.org/10.3390/educsci13020216

Academic Editor: Fred Paas

Received: 23 December 2022
Revised: 10 February 2023
Accepted: 16 February 2023
Published: 19 February 2023

1. Introduction

Interpreting data presented in graphs is essential to understanding concepts across domains [1,2], especially for learning mathematics [3], to interpret and represent data [4,5], as well as to use media [6]. Therefore, graph interpretation was highlighted as a valuable skill in PISA and as a 21st-century workforce skill [7]. Graph-comprehension skills differ across individuals and depend on multiple factors: (1) graphical literacy, meaning the ability to interpret information represented in graphical form, for instance, identifying relevant features in any context [8–10]; (2) domain knowledge about the represented topic [9,11]; (3) prior knowledge about the underlying mathematical concepts of the graph [8]; (4) task knowledge, such as using a graph to solve a problem or identifying specific data points [12]. It is reasonable to assume that experts should have higher levels of graph-comprehension skills than non-experts. However, the determination of expertise can differ (see section Determination of Expertise). This is an important aspect to keep in mind, as the interpretation of differences in the visual behavior of experts and non-experts may depend on how expertise is determined. This holds true for this review when comparing the visual behavior of experts and non-experts during problem-solving and learning with graphs.

Visual processing of the graph is very important for graph comprehension. We use the term visual processing to emphasize that not only seeing the relevant information, but also actively processing is important for comprehending the depicted information. There is evidence that the visual processing of external representations changes with increasing expertise [13,14]. The underlying assumption is that people mentally process the information they look at [15] (eye-mind hypothesis).

There are various theories why the way we distribute attention might change with increasing expertise. Several of those theories have been supported by eye-tracking studies and literature reviews. For example, the holistic model of image perception states that experts can process an image more efficiently than non-experts [16,17]. This is explained by the enhanced parafoveal processing of experts [16,18]. Experts can analyze an entire image and fixate relevant information earlier than non-experts [16]. Furthermore, experts seem to process information faster than non-experts, as evidenced by shorter fixation durations (see the meta-analysis of Gegenfurtner et al. [17]). This supports the theory of long-term working memory. This theory states that experts learn how to store and retrieve information more effectively, which results in enhanced short-term memory processing [19]. Additionally, the findings of Gegenfurtner et al. [17] support the information-reduction hypothesis [20]. With increasing practice, participants focused more on task-relevant information and less on information that was not relevant to the task [20]. This is called selective attention [10]. These results suggest differences in the visual behavior between experts and non-experts when viewing external representations, such as graphs.

The difference between experts and non-experts' viewing behaviors can be important in the context of education. For example, experts' eye movements could be used as visual instructions to help learners make sense of external representations [21]. Knowledge about how experts read graphs could also be used to facilitate students' information processing [22] or to identify student difficulties in problem-solving or learning with graphs. However, the theories mentioned above use various eye-tracking metrics, such as time to first fixation [16], the fixation count [20], total viewing time [10], or fixation duration [17]. There are similarities between different metrics, e.g., a correlation between total viewing times and fixation count [10] (see also [23] for similar results), but there are also conflicting relations between theoretical models and eye-tracking metrics. For instance, the theory of long-term memory predicts a shorter fixation duration for experts. This, however, is only consistent with the information-reduction hypothesis if experts fixate shorter on irrelevant areas, as this hypothesis predicts more fixations on task-relevant areas for experts than for non-experts [17]. Such possible inconsistencies make it more difficult to interpret how these metrics relate to the differences between experts and non-experts in viewing graphs or diagrams. Furthermore, the way experts and non-experts are defined should be acknowledged, especially in the context of education.

There have been previous literature reviews of eye-tracking in education with various research foci, for example, to summarize the eye-tracking research in physics education [24], to review the scenarios of eye tracking in mathematics education research [25], to compare experts and novices' gaze behavior in sports and medical education research [26], to present a summary of eye-tracking research within the "Psychology of Mathematics Education" conference [27], to investigate the relation between eye movements and cognitive processes during multimedia learning [28], or to provide an overview of the applications of eye tracking in education [29]. None of these review articles focuses on a single type of representation, and regarding the pivotal role of graphs in education, we intend to fill this gap with our review.

We hence aim to (1) provide an overview of eye-tracking metrics that have been used to compare the visual processing of experts and non-experts during problem-solving and learning with graphs. We also (2) summarize the previously found differences in visual behavior between experts and non-experts during learning or problem-solving with graphs.

Knowing how experts view graphs can provide guidelines to support students' visual processing of graphs. For instance, it allows the identification of suitable eye movement modelling examples [30] or relevant areas for signaling support [31]. Moreover, such knowledge can be used to evaluate students' fluency in the visual processing of graphs [32]. In this literature review, we provide an overview of the domains, the types of graphs, the eye-tracking metrics, and how experts are distinguished in the studies.

2. Materials and Methods

A literature review typically consists of three parts. First, the literature search. This is followed by the data extraction, which is then analyzed in the third part of a literature review. In the following we first present the search process of our literature review and then continue with the method used for data extraction. The results based on the data extraction are shown in the Results section.

2.1. Literature Review

The literature search aimed to find articles that analyzed visual behavior when looking at graphs in the context of problem-solving and learning in Science, Technology, Engineering, and Math (STEM) subjects. All included articles should fulfil the following criteria:

- Comparison of experts vs. non-experts (population)
- STEM subject (domain)
- Learning or problem-solving with graphs, diagrams, or functions (intervention)
- Analysis of visual behavior via eye-tracking metrics (outcome)
- Empirical study
- Full text available in English

This resulted in the following categories and terms (see Table 1). In the search string, categories were linked with the Boolean operator AND and terms with the Boolean operator OR.

Table 1. Categories and terms used for searching.

Categories	Terms
Visual behavior	"eye tracking", "viewing behavior", "visual attention"
Graphs	"graph", "diagram", "function"

To identify relevant articles, we searched for titles and abstracts in the databases ERIC, Scopus, Pedocs, and SpringerLink. One possible search string for Scopus would be ("eye tracking" OR "viewing behavior" OR "visual attention") AND ("graph" OR "diagram" OR "function"). As search algorithms differed between databases, key terms in the search string were sometimes replaced with corresponding adjectives or adverbs to include alternative phrasings. This search was conducted in February 2022. Therefore, the publication deadline for relevant publications was 31 January 2022. After the screening process, 24 empirical studies met the inclusion criteria and were included. We then conducted a backwards snowball search using Google Scholar for all included articles and found eight more articles. In total, 32 articles were included in this review.

2.2. Data Extraction

Once the search was completed, relevant data were extracted. Based on our research focus on the differences in visual behavior between experts and non-experts during problem-solving or learning with graphs, we extracted the following data:

- Year of publication
- STEM subject in which the study was conducted
- Type of graph
- Eye-tracking metrics
- Areas of interest (AOIs) used for the analysis of eye-tracking metrics
- Expertise determination
- Key findings

To analyze differences in visual behavior between experts and non-experts, we coded the way authors determined expertise. Furthermore, we coded the domain (STEM subject) and type of graph, as well as the analyzed eye-tracking metrics. To analyze eye-tracking data, the stimuli are split into areas of interest (AOIs). This is useful to investigate the distribution of eye movements across relevant and irrelevant areas. The distribution of eye movements can give insights into the relevance of a representation's components. Depending on the research aim, an AOI can consist of an entire representation, such as a graph, or smaller components, for example, the axes. Furthermore, the analysis of eye-tracking metrics depends on the granularity of the AOIs.

In this review, we differentiate between macro- and meso-level AOIs and micro-level AOIs when analyzing the gaze behavior of experts and non-experts [33]. We used this distinction to code AOIs based on descriptions in the included studies. Macro-level AOIs consist of an entire graph. These AOIs can be useful to research how graphs are embedded in the learning material, e.g., between questions and answers. Meso-level AOIs divide the graph into large components, such as dividing the graph from the x- and y-axes. This means that more than one AOI covers the graph area, but there are separate information sources, such as single-axis values that are included in the same AOI. Micro-level AOIs split a comprehensive representation into particular elements, that can be based, for example, on specific information that is relevant to study specific sections of a graph, such as an axis with separate numbers on it.

3. Results

We identified 32 articles in our review, that analyzed the visual behavior of experts and non-experts when looking at graphs in the context of problem-solving and learning. An overview of all included studies can be found in Table 2. This table surveys authors, publication years, subjects, graph types, measurements to determine expertise and analyzed eye-tracking metrics.

Table 2. Overview over studies included in the literature review, including eye-tracking metrics (FD: fixation duration, FC: fixation count; DT: dwell time; S: saccades; FG: first gaze; PS: pupil size; T: transitions; NRV: number of revisits; AOI: area of interest; SD: standard deviation).

Reference	Year of Publication	Subject	Graph Type	Determination of Expertise	Eye-Tracking Metrics
Ahmed et al.	2021	Engineering	Line graphs	Professionals	FD (average, total), FC (average, total)
Atkins and McNeal	2018	Geoscience	Line and bar graphs	Pre-test	FD (normalized, total)
Brückner et al.	2020	Physics, Economics	Line graphs	Domain	DT (total, on relevant AOIs)
Dzsotjan et al.	2021	Physics	Line graphs	Learning gain	Multiple features including DT (total, mean; SD of both)
Harsh et al.	2019	Biology	Line graphs, diagrams	Level of study	FC (normalized), DT (normalized), S (normalized)
Huang and Chen	2016	Physics	Diagram	Spatial working memory	DT (average), FC (total stimulus, on AOIs), FG, PS, S
Ho et al.	2014	Biology	Line graphs	Prior knowledge	FD (total), T, NRV
Kekule	2014	Physics	Line graphs	Performance	Heat maps based on FC
Keller and Junghans	2017	Medicine	Line graphs	Numeracy	FD (relative), FC (relative)
Kim et al.	2014	Math	Line graphs	Dyslexia	DT, FG.
Kim and Wisehart	2017	Math	Bar graphs	Dyslexia	DT, T

Table 2. *Cont.*

Reference	Year of Publication	Subject	Graph Type	Determination of Expertise	Eye-Tracking Metrics
Klein et al.	2019	Physics, Finance	Line graphs	Domain	DT (total; AOI and entire stimulus), FC (average; AOI), S
Klein et al.	2020	Physics	Line graphs	Performance	DT
Kozhevnikov et al.	2007	Physics	Line graphs	Spatial ability	FD (relative)
Küchemann et al.	2020	Physics	Line graphs	Performance	DT
Küchemann et al.	2021	Physics	Line graphs	Performance	DT (total, relative), T
Madsen et al.	2012	Physics	Diagrams, line graphs	Performance	FD (normalized; overall, first two seconds)
Okan et al.	2016a	Medicine	Line and bar graphs	Graph literacy	FD (total)
Okan et al.	2016b	Medicine	Line and bar graphs	Graph literacy	FD
Peebles and Cheng	2003	Economics	Line graphs	NA †	Not applicable
Richter et al.	2021	Economics	Line graphs	Prior knowledge	DT, FG, T, PS
Rouinfar et al.	2014	Physics	Diagram	Performance	Domain relative ration (relative dwell time /relative area of AOI)
Skrabankova et al.	2020	Physics	Line graphs	Teacher's opinion	T, FC
Strobel et al.	2019	Various topics	Bar graphs	Working memory capacity	FD (total)
Susac et al.	2018	Physics, Finance	Line graphs	Domain	DT
Tai et al.	2006	Various topics	Line graphs	Domain	FD, DT, S
Toker et al.	2013	Evaluating student performance	Bar and radar graphs	Working memory capacity, visualization experience	FD (total, relative, mean, SD), FC (total, relative), S,T
Toker and Conati	2014	Data analysis	Bar graphs	Perceptual speed, working memory	FC, FD, S
Viiri et al.	2017	Physics	Line graphs	Performance	Heat maps
Vila and Gomez	2016	Economics	Bar graphs	Performance	DT
Yen et al.	2012	Physics, various topics	Line graphs	Domain	DT (normalized), FC
Zhu and Feng	2015	Math	Line graphs	Performance	T
Viiri et al.	2017	Physics	Line graphs	Performance	Heat maps
Vila and Gomez	2016	Economics	Bar graphs	Performance	DT
Yen et al.	2012	Physics, various topics	Line graphs	Domain	DT (normalized), FC
Zhu and Feng	2015	Math	Line graphs	Performance	T

† Comparison with a scanpath assumed optimal for the task.

An overview of the analyzed variables can be seen in the graphs depicted in Figure 1. The included experiments are described in more detail regarding the individual variables in the following sections, starting with the publication period of the included studies.

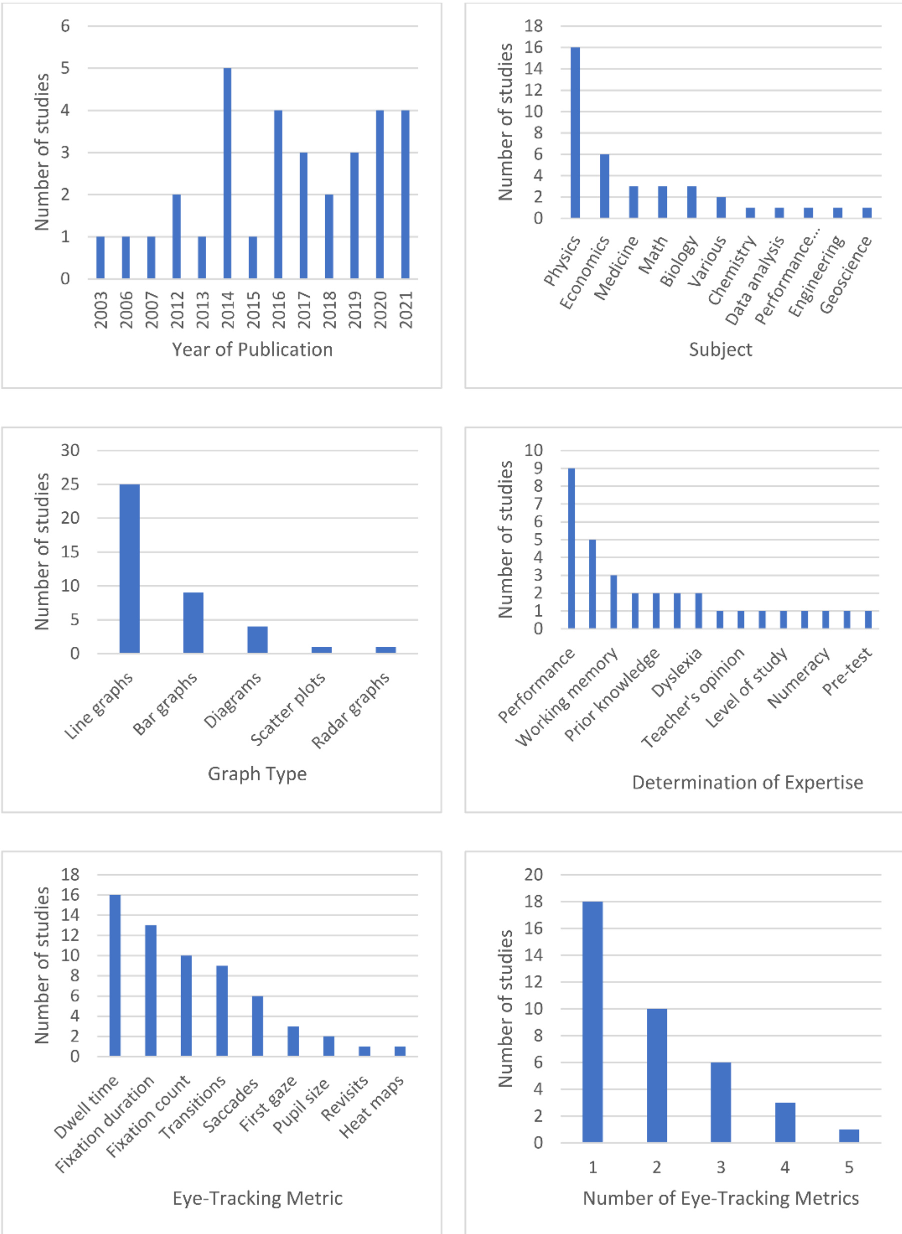


Figure 1. Overview of the number of studies related to the visual behavior of experts and non-experts' during learning and problem-solving with graphs per year (**top left**); number of studies using graphs of a certain subject (multiple mentions are possible, **top right**); types of graphs used in the studies (**middle left**); overview of the measure for determining expertise (multiple mentions are possible, **middle right**); overview of eye-tracking metrics used in the studies included in the literature review (**low left**); number of eye-tracking metrics used for analyzing visual behavior when looking at graphs (**low right**).

3.1. Publication Period

Of the 32 included articles, the first study was published in 2003 (see Figure 1, top left). In the first decade starting from 2003, only a limited number of six studies were published, whereas most studies (n = 26) were published after 2013. The number of studies in our review did not increase uniformly, as we identified six years between 2003 and 2013 in which no studies with eye tracking that examined the visual behavior of experts and non-experts during problem-solving and learning with graphs were published. After 2014, we could see an increase in the number of publications about visual behavior when looking at graphs, with five publications in 2014 and four each in 2016, 2020, and 2021. Starting in 2018, a more constant number of studies comparing experts and non-experts when learning or solving problems with graphs were published.

This distribution is comparable to other reviews about eye tracking in education. Before 2006, only a few eye-tracking studies were published in math education research [25], increasing until the year 2018. The authors stated that this increase could be due to the technical advances in eye-tracking technology and therefore easier usage [25]. Correspondingly, more terms related to eye tracking ("eye[-]tracking", "eye[-]movement", "gaze[-]tracking", "gaze[-]movement") were identified via content analysis in the proceedings of the International Group for the Psychology of Mathematics Education, indicating an increased relevance of eye-tracking technology in education research [27].

3.2. Domains and Types of Graphs

In education research, eye-tracking studies about experts and non-experts learning and solving problems with graphs have been conducted in various STEM subjects (see Figure 1, top right). Out of 32 studies, 16 presented graphs based on the subject of physics, for example, works by Dzsotjan et al. [34] or Kozhevnikov et al. [35]. Out of these, three articles compared physics with economics graphs [1,11,36]. Following physics and economics, the second most studies were conducted in medicine [10,37], mathematics [38–40], and biology [41–43] with three published experiments per subject.

Most of the studies (n = 25) used line graphs (Figure 1, middle left). This finding holds when looking at specific STEM subjects. For example, 13 out of the 16 physics studies presented line graphs. This corresponds to the common topic of kinematics [44,45]. Studies on visual behavior in graphs in a biological domain also used line graphs exclusively [41–43]. Studies in a mathematics and medical domain also mostly used line graphs (math: [38,40]; medicine: [10,37]). However, Okan et al. [10] analyzed the visual processing of line and bar graphs in a medical domain. Likewise, bar graphs in combination with line graphs were the focus of studies in a geoscience domain [23]. Furthermore, bar graphs were used in combination with radar graphs [46]. Studies using only bar graphs ranged in topic from economics [47] to data analysis [48].

3.3. Determination of Expertise

To compare the visual behavior of participants of various expertise levels during problem-solving and learning with graphs, researchers classified their participants according to different measures. An overview of the measures used for expertise determination across all experiments can be seen in Figure 1 (middle right). An overview of the expertise determination in individual studies can be found in Table 2. Please note that we cannot identify potential differences and overlaps between the measures used to determine expertise because not all test materials were publicly available. In the Introduction we presented four factors that are often used to determine expertise: (1) graphical literacy, (2) domain knowledge, (3) mathematical prior knowledge, and (4) task knowledge. However, some of the measures used to determine expertise in the studies examined in this review cannot be categorized as one of these four. Clear discrimination between these factors may not always be possible and mapping them with the indicators of expertise used in the studies is complex. For example, an item in which students solve a problem with a graph may contain information about students' prior knowledge in both domain and math contexts as well as a certain level of graphical literacy skills and task knowledge. In such a case, the performance when solving the item would measure all four factors. Similarly, learning gain [34], teacher's opinion [49], level of study [41], comparison with professionals [50] and pretest score (e.g., in graph understanding [23]) may all cover the four factors. In contrast, working memory [51], spatial abilities [35], and dyslexia [38,39] do not address any of these factors, whereas the remaining determinators cover only parts of the factors, although one might argue that the latter contains the factor of task knowledge, as dyslexic participants had trouble with reading.

Most researchers determined expertise post-hoc based on participants' performance in the learning or problem-solving task (e.g., [52–54]). Determining expertise a priori based on their domain of study was performed when comparing students of different subjects (e.g., [1,11]) or science with non-science students [55].

Moreover, some authors used multiple measures, for example, working memory capacity and subjective assessments of visualization experience [46].

Although there was a clear preference to use performance and domain to determine experts, other—sometimes unusual—measures were also employed. Many studies compared experts and non-experts via students' performance on specific tasks, where expertise might be located on a continuous scale, instead of comparing groups with clear distinctions. The variety of ways expertise was determined should be kept in mind when interpreting the eye-tracking metrics and comparing experts and non-experts as described in the next sections.

3.4. Eye-Tracking Metrics

Previous studies used various eye-tracking metrics to compare the visual processes of experts and non-experts during problem-solving and learning with graphs. In the following, we aim to provide an overview of the analyzed eye-tracking metrics in the included studies (research aim 1).

Figure 1 (lower left) shows the eye-tracking metrics that the authors of the 32 included studies used to compare the visual behavior of experts and non-experts. Eye-tracking metrics can be grouped into static and dynamic metrics. The sum of static metrics or the average of eye movements over time, for example, attained by calculating the duration someone fixated on a stimulus for the entire time the stimulus, is shown. Dynamic metrics include information about the change in visual attention over time, e.g., the number of eye-movement switches from one part of the stimulus to another (gaze transitions) or the duration between two fixations (saccadic duration). Static eye-tracking metrics in the included studies were based on fixations. These metrics were evaluated by most studies, e.g., mean fixation duration (e.g., [23,56]) or the fixation count. Another popular static metric was dwell time, which describes the sum of total fixation durations and the total duration of saccades within an AOI [57]. However, definitions of dwell time in the articles differ. Whereas some defined it as the "viewing time" [36] (p. 4), others used more specific definitions, such as "eye movements below an acceleration of 8500°/s2 and a velocity below 30°/s" [11] (p. 5). In some cases, we coded metrics as dwell time based on the description in the papers (e.g., "gaze duration", p. 335, [58]), although, in general, we classified the used eye-tracking metrics based on the terms the authors used. Dwell time was also used to calculate new metrics, such as the so-called domain-relative attention, which is defined by dividing the relative dwell time of an AOI by the relative area of the AOI [59]. Other static eye-tracking metrics were the mean time to first fixation on an AOI [60], the pupil size (e.g., [58]), and the number of revisits on AOIs [42]. Dynamic eye-tracking metrics that were used to distinguish the visual processing of experts and non-experts during problem-solving and learning with graphs included transitions (e.g., [40]), and saccades (gaze jumps between two fixations, e.g., saccade duration [43]; absolute saccadic direction [1]). One study qualitatively analyzed heat maps without specifying on what metric they were constructed [54].

Since there are noticeable differences in the type of metric between studies, we also analyzed how many eye-tracking metrics were used in each study. We found that most studies examined more than one eye-tracking metric (M = 1.92, SD = 0.9) but this value differed across domains (see Figure 1, lower right). An exception was one study that used five metrics (fixation duration, fixation count, initial gaze, pupil size, and saccade counts [58]). Three studies used four eye-tracking metrics, e.g., for analyzing individual user characteristics when evaluating student performance (fixation count, fixation duration, saccades, and transitions [46]).

As physics is the most common domain in this review (n = 16, see section Domains and Types of Graphs), we wanted to take a closer look at the eye-tracking metrics used in physics studies. An overview of the metrics used to compare experts and non-experts' visual behaviors when looking at graphs in the domain of physics can be seen in Figure 2. As studies usually collected several eye-tracking metrics (e.g., [34]), the reported number of metrics exceeds the actual number of studies. In all these studies, participants were supposed to solve problems. One exception was a study that analyzed differences in gaze behavior between experts and non-experts before walking the shape of a graph [34]. Static metrics were used to analyze differences in the visual attention of experts and non-experts on relevant and irrelevant areas [1,56]. Comparable to the overall distribution, most studies analyzed dwell time, often comparing physics and non-physics students [1,36]. Both studies found that physics students looked longer at the graph (see section Gaze Behavior below for a closer analysis). Dynamic metrics, such as transitions, were used to predict the performance of students solving the Test of Understanding Graphs in Kinematics (TUG-K [53]).

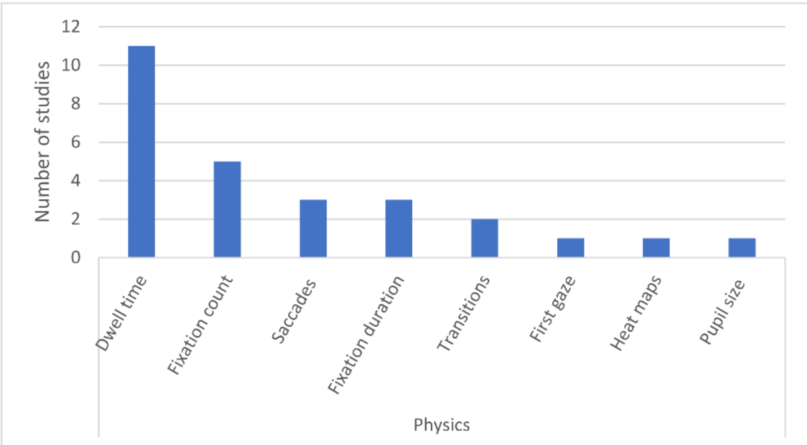


Figure 2. The number and types of eye-tracking metrics used in studies investigating the visual behavior of experts and non-experts learning or problem-solving with physics graphs.

3.5. Gaze Behavior of Experts and Non-Experts

To summarize the previously found differences in visual behavior between experts and non-experts during problem-solving or learning with graphs (research aim 2), we differentiated the analysis of eye-tracking metrics, whether static or dynamic, depending on the granularity of the AOIs. We therefore consider results based on the way AOIs are defined: at macro- or meso-level and at micro-level (see also section Data Extraction). We first present the results based on the bigger macro- and meso-level AOIs and then go on to the smaller micro-level AOIs.

3.5.1. Macro- and Meso-Level

Macro- and meso-level AOIs consist of an entire graph or analyze mid-sized sections of a graph, such as the axes and the graph. Results of studies using meso- and micro-level AOIs can be seen in Table 3.

Table 3. Overview of findings of studies analyzing eye-tracking metrics based on meso- and macro-level AOIs.

Dependent Variable	Findings and References
Fixation duration	Experts have longer average fixation durations, but spend a shorter time on the graph than non-experts [50] Experts have the same fixation duration on a graph as non-experts [55,58] Experts fixate less on seductive details [54] Experts pay more attention to trends than non-experts, but non-experts pay more attention to the title and the axes [23] Experts look longer at the graph than non-experts ([42]; [10], experiment 2, only for conflicting graphs) Experts look longer at relevant areas (experiment 1 [10]; [59]) Experts look less at irrelevant axes' labels [54,55]
Fixation count	On average, experts fixate less often on graphs than non-experts [43,58] Experts and non-experts make the same number of fixations [49] Experts look less often at irrelevant regions [55]
Transitions	Experts transitioned less often between a graph and text [39,51] Experts switch more often between graphs and between graphics and text than non-experts [42] Experts made "more strategic transitions among AOI triples" [40] (p. 1) Experts made fewer transitions than non-experts on harder tasks [48] Experts made the same relative number of transitions as non-experts (experiment 1 [10])
First gaze/fixation	Experts initially spend more time on the graph than non-experts [58] Experts look at the graph data later than non-experts [60]
Dwell time	Non-experts spend more time on the graph than experts [36,38] There are no differences in total dwell time between experts and non-experts [11] Experts look longer at the correct answer [45] Experts (i.e., students without dyslexia) paid less attention to the x-axis [39]
Saccades	Experts make fewer saccades than non-experts [43]
Revisits	Experts visit the graph more often than non-experts [42]

Regarding the analysis of meso- and macro-level AOIs, there were varying results when looking at fixation duration, the fixation count, transitions, and dwell time (see Table 3). First, we look at the static metrics that many studies analyzed: fixation duration, fixation count and dwell time. In general, it seems as if experts pay more attention to relevant areas than non-experts (experiment 1 [10], [23,45,59]) and less attention to irrelevant areas [51,54,55]. Experts might also attend less to the graph than non-experts [36,38,43,50,58], although this finding is unclear, as other studies found no differences [11,49,55,58] or concluded that experts look longer at the graph than non-experts ([42]; [10], experiment 2, only for conflicting graphs).

One study with results that contradict other studies in several instances is the one by Huang and Chen [58]. In this case, expertise was based on gender under the assumption that the gender difference in spatial working memory might influence the integration between text and diagram [58]. However, the authors did not find gender differences in this task. Additionally, only one of the three diagrams analyzed together was a graph [58]. The operationalization of expertise could also not be categorized based on the four factors. The results of this experiment might not match the others due to differences in determining expertise. Similarly, another experiment compared the expertise as determined by the teacher [49]. The authors also concluded that the teacher's opinion was not well suited for grouping students according to performance [49]. The same might hold true for using dyslexia as a determinator of expertise [38]. The reasons for the varying results of the other studies are less clear. Some compared science and non-science students [11,36,55]. Brückner et al. [11] compared physics and economics students, whereas Susac et al. [36] compared physics and psychology students. Although these student groups had different domain knowledge, one could assume that economics students might have more experience with reading graphs (factor graph literacy) as well as more experience with math lectures (factor math prior knowledge). Economics students might have been more similar to physics students than psychology students in this regard, leading to varying results. Tai et al. [43] compared biology, chemistry, and physics students. Besides the differences in expertise determination, the sample sizes might also play a role in the results (e.g., $N = 6$ [43]).

There were not as many experiments analyzing dynamic eye-tracking metrics as there were for static eye-tracking metrics (see Table 2). Since transitions were the most used dynamic eye-tracking metric, we will take a closer look at them. Two studies found that experts transitioned less often than non-experts between graphs and text [39,58], whereas others found the opposite [42]. An explanation could be that the transitions of experts were more strategic during problem-solving [40], which could lead to experts making the same relative number of transitions as non-experts, taking the total number of transitions into account [10] (experiment 1).

3.5.2. Micro-Level

In contrast to macro- and meso-level AOIs, AOIs at the micro-level are very small and include specific parts of the graph, for example, certain sections of the x-axis. In this section, we will consider experts' strategies solely on the graph area (i.e., without the question or answer choices). To get an understanding of experts' strategies at this level, a finer classification of AOIs in the graph domain is warranted, typically considering individual values separately. The results of studies using these types of AOIs can be seen in Table 4.

Table 4. Overview of findings of studies analyzing eye-tracking metrics based on micro-level AOIs.

Dependent Variable	Findings and References
Fixation duration	Experts spend more time on graph information (such as title and variables) than non-experts [41,46] Experts look at the entire graph [1] Experts spend more time on relevant areas [1,37,47]
Fixation count	Experts fixate on the axes more often [35] Experts visit graph information (such as title and variables) more often than non-experts [41] Experts fixate more often on task-relevant AOIs [37]
Transitions	Experts transition more often between conceptually relevant areas [53]
Revisits	Experts study the axes, axes labels and line segments more often [35]
Dwell time	Experts look longer at conceptually relevant areas [52,53,56] Experts spend less time on areas that can be used to calculate the solution [53] Experts spend less time on areas found relevant for non-experts [56]
Saccades	Experts look along the graph slope [1]

Similarly, to meso- and macro-level AOIs, regarding static eye-tracking metrics, experiments analyzing micro-level AOIs also found that experts paid more attention to relevant AOIs [1,37,47,52,53,56], including graph information [35,41,46]. Furthermore, experts looked at the entire graph [1]. Moreover, experts seemed to systematically distribute their gaze not only spatially but also temporally [41]. In one example, a faculty member analyzed a graph and the authors showed that efficient information processing meant specifically evaluating graph information and related data at the beginning of viewing. In contrast, inexperienced students jumped between information sources and especially back to the task and the answer choices in no particular order [41].

Few experiments analyzed dynamic eye-tracking metrics in a micro-level analysis (see Table 4). It is difficult to draw a conclusion from such a small sample. In the following section we therefore aim to summarize the visual strategies of experts and non-experts during problem-solving and learning with graphs over the bigger (meso- and macro-level) and smaller (micro-level) AOIs.

4. Discussion

The aim of the present literature review was twofold: (1) We wanted to give an overview of the eye-tracking metrics used to compare experts and non-experts when problem-solving and learning with graphs. Furthermore, we focused on the visual strategies of experts and non-experts guided by the research foci of the identified research articles (2). We further categorized AOIs based on their size, as it might influence the analysis of whether the AOIs are at the bigger meso- and macro-level or at the smaller micro-level.

4.1. Summary of Experts' and Non-Experts' Visual Strategies

To analyze the visual strategies of experts and non-experts during problem-solving and learning with graphs, we first summarize the eye-tracking metrics used in the studies and the according experiments included in this literature review (research aim 1). As there were differences between meso-/macro- and micro-level eye-movement analyses of eye-tracking metrics, we examine those separately before summarizing the visual strategies of experts and non-experts (research aim 2). Finally, we discuss the various ways expertise was determined and how this might influence the interpretation of eye-tracking results.

4.1.1. Overview of Eye-Tracking Metrics

Most experiments compared static metrics, such as dwell time, and fixation duration or fixation count, to analyze visual behavior (n = 39). In comparison, only 15 experiments analyzed dynamic eye-tracking metrics, such as transitions and saccades. Static metrics are useful to analyze the visual behavior over the entire time participants looked at stimuli (e.g., see section Eye-Tracking Metrics). Dynamic metrics can be used to analyze the (temporal) strategy of participants when looking at a stimulus. Although many studies only measured one metric (n = 18), researchers analyzed two eye-tracking metrics on average. Four out of 32 experiments used four or more eye-tracking metrics.

Fixation duration and fixation count were useful for both small and large AOIs. Using two or more (uncorrelated) metrics might give researchers more insight into the visual behavior, especially in a combination of static and dynamic metrics. Regarding transitions between AOIs, we recommend a micro-level analysis, because it is very sensitive to differences between experts and non-experts in more detail. As was common in most studies, we also recommend distinguishing between task- or conceptually relevant and irrelevant AOIs.

4.1.2. Meso-and Macro- vs. Micro-Level AOIs

The distinction between relevant and irrelevant AOIs was quite common in the experiments included in the literature review. However, there might be differences when taking the size of the AOIs into account.

In general, the findings between macro- or meso-level AOIs and micro-level AOIs were very similar (e.g., for fixation duration and fixation count, see Tables 3 and 4), but there were contrary findings when analyzing transitions at different levels. At the meso- and macro-level, experts seemed to make fewer transitions than non-experts (see Table 3). In contrast, at the micro-level, experts made more transitions than non-experts between conceptually relevant AOIs (see Table 4). On a micro-level analysis, experts transitioned more between AOIs, whereas experts seemed to make fewer transitions between AOIs when looking at macro- and meso-level AOIs. One reason could be that experts seemed to pay closer attention to the relevant details of the graph (e.g., [52,53,56]). However, only one experiment analyzed transitions at the micro-level [53]; consequently, it will be necessary to confirm these results before a conclusion can be drawn. We nevertheless can make some statements taking previous theories into account.

As mentioned in the Introduction, there are several theories why there are differences in the visual behavior of experts and non-experts. The results of some experiments included in the literature review support several of those theories. For example, at macro- and meso-level AOIs, Okan et al. [10] demonstrated the so-called information-reduction hypothesis [17,20] in a comparison of participants with high and low graph literacy as investigated in a pre-test. In two studies, AOIs were defined at the meso-level and experts were classified using a graph comprehension test [10]. Consistent with the information-reduction hypothesis, the authors observed that experts were better at identifying task-relevant areas in a graph, which allowed them to spend a greater relative amount of time evaluating relevant information. Specifically, the authors showed that participants with high graph understanding reviewed axes' labels and scaling more frequently to avoid errors [10] (experiment 1). This corresponds to results by Rouinfar and colleagues [59], who found that participants who solved the problem correctly paid more attention to relevant areas of a diagram than incorrect solvers. Rouinfar et al. compared the influence of color highlighting on information extraction with 80 physics students and they stressed the importance of the ability to organize and integrate information to solve a problem correctly. This result confirmed that the improved performance was caused by a learned automatism in task performance (automatism hypothesis) and not by increased awareness of the relevant domains [59] (priority hypothesis). Similarly, Okan et al. [10] observed that the highest number of transitions seemed to occur between the graph region and the question and between the graph region and the axes [10] (experiment 1), which are relevant areas as well. At the micro-level, experts also paid more attention to relevant AOIs, which is in line with the results at the meso- and macro-level and the information-reduction hypothesis [17].

There were not enough experiments to conclusively identify distinct differences between experts and non-experts for specific measures. However, taken together the results of these experiments are in line with existing hypotheses. We therefore believe that we can make some statements about the visual strategies of experts and non-experts during problem-solving and learning with graphs that we will present in the following.

4.1.3. Visual Strategies of Experts and Non-Experts during Problem-Solving and Learning with Graphs

Based on our results, we can make a statement about what distinguishes visual expertise in problem-solving and learning with graphs. Experts systematically looked at relevant information, such as scales as well as labels (e.g., experiment 1 [10]), and performed more integrative eye movements within a graph in terms of dynamic metrics (see Table 3, transitions, revisits, saccades). Therefore, in addition to the formation of chunks [61], information reduction [17,20] is central to expertise related to graphs.

There were some conclusions regarding differences between experts and non-experts viewing specific AOIs. First, experts seemed to spend a relatively short amount of time on the task and answer choices during problem-solving [36,43,50,53], which might also be attributed to the fact that experts did not (or hardly) perform comparisons between answer choices [45]. Instead, experts paid more relative attention to axis scaling, axis labels, and graph progression [53], as well as to conceptually relevant AOIs [37].

Moreover, at least in data extraction, an order of information extraction appeared by comparing several works [41,46,62]. The most efficient order of information extraction seemed to emerge when participants looked at the given variables early on (if this indication existed) and directly identified them in the graph [41,46]. Thereby, a recognition of the respective axis and its scaling could take place (experiment 1 [10]; [53]), followed by a jump back to the task [41] to identify the target variable, which is then looked for directly in the graph [62]. Depending on cognitive abilities and the task difficulty, one may jump back to variable information [41]. The expertise seems easily transferable to other styles of graphs (e.g., linear vs. radial) but not (or only by further training) to other types of graphs (e.g., line and bar graphs) [63].

However, possible deviations from this strategy at high expertise have not been identified yet. Furthermore, influences or trade-offs that lead to deviation from this optimal strategy in experts remain unclear. In addition, the optimal temporal sequence for more complex tasks was not determined. A complex task would be, for example, determining the slope or the area underneath a graph. So far, in two tasks, it seemed that students with correct solutions looked longer along the graph (when determining the slope) and into the areas below and above the graph (when determining the area) [52].

There have also been some inconsistencies in our results (see Gaze Behavior of Experts and Non-Experts). These might be due to the determination of expertise in individual studies. As mentioned in the beginning, four factors are important when determining expertise in this area: (1) graphical literacy [8–10]; (2) domain knowledge [9,11]; (3) math prior knowledge [8]; (4) task knowledge [12].

In our review, performance, learning gain, level of study, comparison with professionals, and a pretest were measures used to determine expertise that may have fulfilled all four factors of graph-comprehension skills. A teacher's opinion may also consider all four factors. However, this did not prove to be a good indicator of expertise. Of these measures, performance was the most common one (Figure 1, middle right). Learning gain, level of study, comparison with professionals, and pretest were only used to determine expertise in one study, respectively (see Table 2). A direct comparison between studies using the same expertise determinator is generally possible, but the nine studies using performance vary strongly regarding AOI sizes and eye-tracking metrics, which makes them unsuitable for direct comparison. However, there are no conflicts in the findings. In sum, we recommend using objective measures for determining expertise and using tests that explicitly address all four factors to allow for replicability and comparability.

4.2. Limitations

Our review of the literature about visual processing comparing experts and non-experts during problem-solving and learning with graphs has several limitations. First, we did not concentrate on one specific definition of expertise determination. Therefore, studies used various measures to define and compare groups of varying expertise. This could be one reason for the contrasting results. It also made drawing overarching conclusions difficult.

Second, there were some inconsistencies in using terms for eye-tracking metrics. For example, the difference between dwell time and viewing time was not always clear. In one case, the basis for the calculation of heat maps was not reported [54].

Third, in analyzing the various articles on eye tracking during learning and problem-solving with graphs, the resolution of the eye-tracking systems was not considered. This means that the accuracy with which the results were reported may be subject to variation. An increase in spatial and temporal resolution, as well as accuracy, over the period studied may well be expected due to technological advancements in eye-tracking devices.

We do not claim completeness for the studies included in our review. Our search process was not entirely systematic, which might have led to an incomplete list of included studies. We also did not include grey literature, which might have resulted in a publication bias towards positive and significant results. Moreover, results were only coded by the first author; we could therefore not assess the validity of our codes. However, the codes were straightforward, apart from the eye-tracking metrics concerning dwell time, which made coding relatively easy.

4.3. Future Research

We aimed to examine relevant articles that investigated gaze behavior during problem-solving and learning with graphs. One of the main limitations of this literature review was the differing definitions of expertise determination. We therefore suggest the consideration of the four factors (1) graphical literacy, (2) domain knowledge, (3) mathematical prior knowledge, and (4) task knowledge. For example, expertise is sometimes established only based on the study progress [41]. This leaves it unclear to the reader to what extent participants are truly experts. Ideally, a criterion based on an assessment that tests the four factors would be established. In addition to these four factors, efficiency in visual processing, if applicable, may also be used as a criterion of expertise determination [32].

In general, it is probably the best idea to find a field consensus for the definition of experts. In the case of graphs, it might be difficult to identify the specific field to which graphs belong, and to find a consortium of researchers that represents all relevant fields. Therefore, we suggest an iterative empirical approach: Due to the lack of consensus for the definition of experts, we propose a research-informed and domain-independent identification of a group of experts. As a next step, it is necessary to verify and consequently to refine such identification of experts, which in turn needs to be tested again.

In the case of graphs, we believe that the most important variables are the AOIs that experts used to solve the task for various types of graphs and domains, how long they need to focus on it, and how they connect these areas (in terms of gaze transitions). Once there is such validated definition of experts, the visual processes of those experts would be a great implementation for teaching the understanding and efficient processing of graphs, how to approach graphs in unknown fields, i.e., to transfer the skills to other domains, how to best implement information in graphs, and how to design graphs.

We assumed that the articles identified in this review would be largely limited to stationary eye-tracking systems, as graphs in experiments in education research are primarily presented digitally on a computer screen. In fact, only three studies examined gaze behavior during problem-solving or learning with mobile eye-tracking systems [34,38,50]. This observation could be expected given the more diverse technological solutions and easier feasibility of stationary eye-tracking studies. As most studies with mobile eye tracking were published recently, we believe that their number will increase in the future. In terms of analysis of eye-tracking metrics, graphs mainly analyze spatial distributions of gaze. We could identify only one paper [41] that evaluated a temporal sequence of attention in problem-solving with graphs. However, others made the first steps, such as looking at the total fixation time on an AOI vs. the fixation time in the first two seconds in an AOI [56]. Accordingly, the evidence on expert strategies is also limited only to the spatial distribution of gaze. It would be interesting to see whether there are also temporal differences between experts and non-experts during problem-solving or learning with graphs.

We found two papers that depicted an evolution in subjects' gaze behavior while problem-solving or learning with graphs [11,59]. In both cases, there was no specific instruction to influence gaze behavior. Accordingly, the extent to which learning gains in graph comprehension are associated with changes in gaze behavior is currently under research. Furthermore, studying whether the results of problem-solving activities are transferable to learning would be very valuable. In this way, it would also be interesting to analyze the various phases of problem-solving separately. As mentioned above, there could be an ideal strategy to extract information from graphs and a closer look at these phases could be interesting.

Visual processing during problem-solving and learning might also depend on the education level of the participants. Most studies were conducted with college or university students; there are currently only three studies that investigate the gaze behavior of high school students during graph viewing [40,45,52]. Consequently, most papers have investigated an advanced stage of gaze behavior in graphs; there were no studies that analyzed the gaze behavior of children just learning about graphs. An account of the gaze behavior of students, who are just acquiring the understanding of graphs, and appropriate instructional suggestions based on this, are therefore currently missing. Our sample might also be biased towards physics because half of the included experiments (n = 16) used graphs in this domain. Although some studies compared various STEM contexts (e.g., biology, chemistry, and physics [43]), future research would benefit from comparisons in more domains as well as more types of graphs, since most experiments analyzed line graphs. Due to our limited sample, replication studies of the experiments presented here, for example with differing eye-tracking metrics or in other domains, might further strengthen the current evidence base.

5. Conclusions

Experts and non-experts differ in the way they interpret graphs. We reviewed 32 articles about experts and non-experts solving problems and learning with graphs. Most commonly examined eye-tracking metrics were static, such as fixation duration and fixation count. Experts seemed to focus longer on relevant areas and to identify the relevant variables in the graphs faster than non-experts. Their visual processing also seemed to be more systematic than that of non-experts: first identifying the given variables and then directly looking for the target variable in the task and the graph. Regarding dynamic process metrics, we suggest studying transitions between small areas of interest, and we encourage considering temporal metrics in future research. Furthermore, expertise was determined in different ways across studies, which are partially not in line with previous determinators of expertise in graph comprehension, limiting the replicability and comparability of findings. As a starting point for future research, we therefore recommend a clear definition of expertise and propose four factors of graph-comprehension skills as a starting point for consideration: (1) graphical literacy, (2) domain knowledge, (3) mathematical prior knowledge, and (4) task knowledge.

Author Contributions: Conceptualization, J.K. and S.K.; methodology, S.K. and V.R.; formal analysis, S.K. and V.R.; investigation: S.K. and V.R., validation: M.B., F.F., M.R.F., A.H., S.I.H., J.K., S.K., V.R. and J.M.Z.; data curation, V.R.; writing—original draft preparation, S.K. and V.R.; writing—review and editing, M.B., F.F., M.R.F., A.H., S.I.H., J.K., S.K., V.R. and J.M.Z.; visualization: V.R.; supervision, J.K. and S.K.; project administration, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Klein, P.; Küchemann, S.; Brückner, S.; Zlatkin-Troitschanskaia, O.; Kuhn, J. Student understanding of graph slope and area under a curve: A replication study comparing first-year physics and economics students. *Phys. Rev. Phys. Educ. Res.* **2019**, *15*, 1–17. [CrossRef]
2. Stern, E.; Aprea, C.; Ebner, H.G. Improving cross-content transfer in text processing by means of active graphical representation. *Learn. Instr.* **2003**, *13*, 191–203. [CrossRef]
3. Duval, R. A cognitive analysis of problems of comprehension in a learning of mathematics. *Educ. Stud. Math.* **2006**, *61*, 103–131. [CrossRef]
4. Gutiérrez, F.; Seipp, K.; Ochoa, X.; Chiluiza, K.; De Laet, T.; Verbert, K. LADA: A learning analytics dashboard for academic advising. *Comput. Hum. Behav.* **2018**, *107*, 105826. [CrossRef]
5. Leinhardt, G.; Zaslavsky, O.; Stein, M.K. Functions, Graphs, and Graphing: Tasks, Learning, and Teaching. *Rev. Educ. Res.* **1990**, *60*, 1–64. [CrossRef]
6. Gould, R. Data literacy is statistical literacy. *Stat. Educ. Res. J.* **2017**, *16*, 22–25. [CrossRef]
7. Program for International Student Assessment (PISA). PISA 2022 Mathematics Framework. 2022. Available online: https://pisa2022-maths.oecd.org/ (accessed on 22 September 2022).
8. Curcio, F.R. Comprehension of Mathematical Relationships Expressed in Graphs. *J. Res. Math. Educ.* **1987**, *18*, 382–393. [CrossRef]
9. Freedman, E.G.; Shah, P. Toward a model of knowledge-based graph comprehension. In *Diagrammatic Representation and Inference*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2002; Volume 2317, pp. 18–30. [CrossRef]

10. Okan, Y.; Galesic, M.; Garcia-Retamero, R. How people with low and high graph literacy process health graphs: Evidence from eye-tracking. *J. Behav. Decis. Mak.* **2015**, *29*, 271–294. [CrossRef]
11. Brückner, S.; Zlatkin-Troitschanskaia, O.; Küchemann, S.; Klein, P.; Kuhn, J. Changes in Students' Understanding of and Visual Attention on Digitally Represented Graphs Across Two Domains in Higher Education: A Postreplication Study. *Front. Psychol.* **2020**, *11*, 1–20. [CrossRef]
12. Friel, S.N.; Curcio, F.R.; Bright, G.W. Making sense of graphs: Critical factors influencing comprehension and instructional implications. *J. Res. Math. Educ.* **2001**, *32*, 124. [CrossRef]
13. Kok, E.M.; Jarodzka, H.; de Bruin, A.; BinAmir, H.A.N.; Robben, S.G.F.; Van Merriënboer, J.J.G. Systematic viewing in radiology: Seeing more, missing less? *Adv. Health Sci. Educ.* **2015**, *21*, 189–205. [CrossRef]
14. Mudrick, N.V.; Azevedo, R.; Taub, M. Integrating metacognitive judgments and eye movements using sequential pattern mining to understand processes underlying multimedia learning. *Comput. Hum. Behav.* **2019**, *96*, 223–234. [CrossRef]
15. Just, M.A.; Carpenter, P.A. A theory of reading: From eye fixations to comprehension. *Psychol. Rev.* **1980**, *87*, 329–354. [CrossRef] [PubMed]
16. Kundel, H.L.; Nodine, C.F.; Conant, E.F.; Weinstein, S.P. Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology* **2007**, *242*, 396–402. [CrossRef] [PubMed]
17. Gegenfurtner, A.; Lehtinen, E.; Säljö, R. Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educ. Psychol. Rev.* **2011**, *23*, 523–552. [CrossRef]
18. Sheridan, H.; Reingold, E.M. The holistic processing account of visual expertise in medical image perception: A review. *Front. Psychol.* **2017**, *8*, 1620. [CrossRef]
19. Ericsson, K.A.; Kintsch, W. Long-Term Working Memory. *Psychol. Rev.* **1995**, *102*, 211–245. [CrossRef]
20. Haider, H.; Frensch, P.A. Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* **1999**, *25*, 172. [CrossRef]
21. Xie, H.; Zhao, T.; Deng, S.; Peng, J.; Wang, F.; Zhou, Z. Using eye movement modelling examples to guide visual attention and foster cognitive performance: A meta-analysis. *J. Comput. Assist. Learn.* **2021**, *37*, 1194–1206. [CrossRef]
22. Noroozi, O.; Alikhani, I.; Järvelä, S.; Kirschner, P.A.; Juuso, I.; Seppänen, T. Multimodal data to design visual learning analytics for understanding regulation of learning. *Comput. Hum. Behav.* **2019**, *100*, 298–304. [CrossRef]
23. Atkins, R.M.; McNeal, K.S. Exploring differences among student populations during climate graph reading tasks: An eye tracking study. *J. Astron. Earth Sci. Educ. (JAESE)* **2018**, *5*, 85–114. [CrossRef]
24. Hahn, L.; Klein, P. Eye tracking in physics education research: A systematic literature review. *Phys. Rev. Phys. Educ. Res.* **2022**, *18*, 013102. [CrossRef]
25. Strohmaier, A.R.; MacKay, K.J.; Obersteiner, A.; Reiss, K.M. Eye-tracking methodology in mathematics education research: A systematic literature review. *Educ. Stud. Math.* **2020**, *104*, 147–200. [CrossRef]
26. Brams, S.; Ziv, G.; Levin, O.; Spitz, J.; Wagemans, J.; Williams, A.M.; Helsen, W.F. The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychol. Bull.* **2019**, *145*, 980. [CrossRef]
27. Lilienthal, A.J.; Schindler, M. Current Trends in Eye Tracking Research in Mathematics Education: A PME Literature Review: A PME Survey. In Proceedings of the Annual Meeting of the International Group for the Psychology of Mathematics Education (PME-43), Pretoria, South Africa, 7–12 July 2019; Volume 4, p. 62.
28. Alemdag, E.; Cagiltay, K. A systematic review of eye tracking research on multimedia learning. *Comput. Educ.* **2018**, *125*, 413–428. [CrossRef]
29. Lai, M.-L.; Tsai, M.-J.; Yang, F.-Y.; Hsu, C.-Y.; Liu, T.-C.; Lee, S.W.-Y.; Lee, M.-H.; Chiou, G.-L.; Liang, J.-C.; Tsai, C.-C. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educ. Res. Rev.* **2013**, *10*, 90–115. [CrossRef]
30. Jarodzka, H.; Balslev, T.; Holmqvist, K.; Nyström, M.; Scheiter, K.; Gerjets, P.; Eika, B. Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instr. Sci.* **2012**, *40*, 813–827. [CrossRef]
31. Van Gog, T. 11 The Signaling (or Cueing) Principle in Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*; Cambridge University Press: Cambridge, UK, 2014; pp. 263–278.
32. Rau, M.A. Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educ. Psychol. Rev.* **2016**, *29*, 717–761. [CrossRef]
33. Andrá, C.; Lindström, P.; Arzarello, F.; Holmqvist, K.; Robutti, O.; Sabena, C. Reading mathematics representations: An eye-tracking study. *Int. J. Sci. Math. Educ.* **2015**, *13*, 237–259. [CrossRef]
34. Dzsotjan, D.; Ludwig-Petsch, K.; Mukhametov, S.; Ishimaru, S.; Kuechemann, S.; Kuhn, J. The Predictive Power of Eye-Tracking Data in an Interactive AR Learning Environment. In *UbiComp/ISWC 2021—Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 467–471. [CrossRef]
35. Kozhevnikov, M.; Motes, M.A.; Hegarty, M. Spatial visualization in physics problem solving. *Cogn. Sci.* **2007**, *31*, 549–579. [CrossRef]
36. Susac, A.; Bubic, A.; Kazotti, E.; Planinic, M.; Palmovic, M. Student understanding of graph slope and area under a graph: A comparison of physics and nonphysics students. *Phys. Rev. Phys. Educ. Res.* **2018**, *14*, 020109. [CrossRef]
37. Keller, C.; Junghans, A. Does guiding toward task-relevant information help improve graph processing and graph comprehension of individuals with low or high numeracy? An eye-tracker experiment. *Med. Decis. Mak.* **2017**, *37*, 942–954. [CrossRef]

38. Kim, S.; Lombardino, L.J.; Cowles, W.; Altmann, L.J. Investigating graph comprehension in students with dyslexia: An eye tracking study. *Res. Dev. Disabil.* **2014**, *35*, 1609–1622. [CrossRef]
39. Kim, S.; Wiseheart, R. Exploring Text and Icon Graph Interpretation in Students with Dyslexia: An Eye-tracking Study. *Dyslexia* **2017**, *23*, 24–41. [CrossRef] [PubMed]
40. Zhu, M.; Feng, G. An exploratory study using social network analysis to model eye movements in mathe-matics problem solving. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, Poughkeepsie, NY, USA, 16–20 March 2015; pp. 383–387.
41. Harsh, J.A.; Campillo, M.; Murray, C.; Myers, C.; Nguyen, J.; Maltese, A.V. "Seeing" Data Like an Expert: An Eye-Tracking Study Using Graphical Data Representations. *CBE—Life Sci. Educ.* **2019**, *18*, ar32. [CrossRef] [PubMed]
42. Ho, H.N.J.; Tsai, M.-J.; Wang, C.-Y.; Tsai, C.-C. Prior knowledge and online inquiry-based science reading: Evidence from eye tracking. *Int. J. Sci. Math. Educ.* **2013**, *12*, 525–554. [CrossRef]
43. Tai, R.H.; Loehr, J.F.; Brigham, F.J. An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *Int. J. Res. Method Educ.* **2006**, *29*, 185–208. [CrossRef]
44. Kekule, M. Students' approaches when dealing with kinematics graphs explored by eye-tracking research method. In Proceedings of the Frontiers in Mathematics and Science Education Research Conference, FISER, Famagusta, North Cyprus, 1–3 May 2014; pp. 108–117.
45. Klein, P.; Lichtenberger, A.; Küchemann, S.; Becker, S.; Kekule, M.; Viiri, J.; Kuhn, J. Visual attention while solving the test of understanding graphs in kinematics: An eye-tracking analysis. *Eur. J. Phys.* **2020**, *41*, 1–16. [CrossRef]
46. Toker, D.; Conati, C.; Steichen, B.; Carenini, G. Individual user characteristics and information visualization: Connecting the dots through eye tracking. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; pp. 295–304.
47. Vila, J.; Gomez, Y. Extracting business information from graphs: An eye tracking experiment. *J. Bus. Res.* **2016**, *69*, 1741–1746. [CrossRef]
48. Toker, D.; Conati, C. Eye tracking to understand user differences in visualization processing with highlighting interventions. In Proceedings of the International Conference on User Modeling, Adaptation, and Personalization, Aalborg, Denmark, 7–11 July 2014; Springer: Cham, Switzerland, 2014; pp. 219–230.
49. Skrabankova, J.; Popelka, S.; Beitlova, M. Students' Ability to Work with Graphs in Physics Studies Related to Three Typical Student Groups. *J. Balt. Sci. Educ.* **2020**, *19*, 298–316. [CrossRef]
50. Ahmed, A.; Hurwitz, D.; Gestson, S.; Brown, S. Differences between Professionals and Students in Their Visual Attention on Multiple Representation Types while Solving an Open-Ended Engineering Design Problem. *J. Civ. Eng. Educ.* **2021**, *147*, 04021005. [CrossRef]
51. Strobel, B.; Grund, S.; Lindner, M.A. Do seductive details do their damage in the context of graph comprehension? Insights from eye movements. *Appl. Cogn. Psychol.* **2018**, *33*, 95–108. [CrossRef]
52. Küchemann, S.; Klein, P.; Becker, S.; Kumari, N.; Kuhn, J. Classification of Munich, 10 February 2023 Students' Conceptual Understanding in STEM Education using Their Visual Attention Distributions: A Comparison of Three Machine-Learning Approaches. In In Proceedings of the 12th International Conference on Computer Supported Education (CSEDU), Prague, Czech Republic, 2–4 May 2020; Volume 1, pp. 36–46.
53. Küchemann, S.; Becker, S.; Klein, P.; Kuhn, J. Gaze-Based Prediction of Students' Understanding of Physics Line-Graphs: An Eye-Tracking-Data Based Machine-Learning Approach. In *Computer Supported Education: 12th International Conference, CSEDU 2020, Virtual Event, May 2–4, 2020, Revised Selected Papers 12*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 450–467. [CrossRef]
54. Viiri, J.; Kekule, M.; Isoniemi, J.; Hautala, J. Eye-tracking the Effects of Representation on Students' Problem Solving Approaches. In Proceedings of the FMSERA Annual Symposium, Finnish Mathematics and Science Education Research Association (FMSERA), Joensuu, Finland, 27–28 October 2016; pp. 88–98.
55. Yen, M.H.; Lee, C.N.; Yang, Y.C. Eye movement patterns in solving scientific graph problems. In Proceedings of the International Conference on Theory and Application of Diagrams, Canterbury, UK, 2–6 July 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 343–345.
56. Madsen, A.M.; Larson, A.M.; Loschky, L.C.; Rebello, N.S. Differences in visual attention between those who correctly and incorrectly answer physics problems. *Phys. Rev. Spéc. Top.-Phys. Educ. Res.* **2012**, *8*, 010122. [CrossRef]
57. Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; Van de Weijer, J. *Eye Tracking: A Compre-Hensive Guide to Methods and Measures*; OUP Oxford: Oxford, UK, 2011.
58. Huang, P.-S.; Chen, H.-C. Gender differences in eye movements in solving text-and-diagram science problems. *Int. J. Sci. Math. Educ.* **2015**, *14*, 327–346. [CrossRef]
59. Rouinfar, A.; Agra, E.; Larson, A.M.; Rebello, N.S.; Loschky, L.C. Linking attentional processes and conceptual problem solving: Visual cues facilitate the automaticity of extracting relevant information from diagrams. *Front. Psychol.* **2014**, *5*, 1094. [CrossRef]
60. Richter, J.; Wehrle, A.; Scheiter, K. How the poor get richer: Signaling guides attention and fosters learning from text-graph combinations for students with low, but not high prior knowledge. *Appl. Cogn. Psychol.* **2020**, *35*, 632–645. [CrossRef]
61. Simon, H.A. How big is a chunk? By combining data from several experiments, a basic human memory unit can be identified and measured. *Science* **1974**, *183*, 482–488. [CrossRef] [PubMed]

62. Peebles, D.; Cheng, P.C.-H. Modeling the effect of task and graphical representation on response latency in a graph reading task. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2003**, *45*, 28–46. [CrossRef] [PubMed]
63. Goldberg, J.; Helfman, J. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Inf. Vis.* **2011**, *10*, 182–195. [CrossRef]

4. Study 3: Comparison and AI-based Prediction of Graph Comprehension Skills based on the Visual Strategies of First-Year Physics and Medicine Students

Reference: Ruf, V., Dinc, Y., Küchemann, S., Berndt, M., Steinert, S., Kugelmann, D., Bortfeldt, J., Schreiber, J., Fischer, M. R., & Kuhn, J. (2024). Comparison and AI-based prediction of graph comprehension skills based on the visual strategies of first-year physics and medicine students. *Physical Review Physics Education Research, 20*(2), 1-15. https://doi.org/10.1103/physrevphyseducres.20.020138

PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH **20,** 020138 (2024)

Comparison and AI-based prediction of graph comprehension skills based on the visual strategies of first-year physics and medicine students

Verena Ruf,[1,*] Yavuz Dinc,[1] Stefan Küchemann,[1] Markus Berndt,[2] Steffen Steinert,[1,3] Daniela Kugelmann,[4] Jonathan Bortfeldt,[5] Jörg Schreiber,[5] Martin R. Fischer,[2] and Jochen Kuhn[1]

[1]*Chair of Physics Education Research, Faculty of Physics, Ludwig-Maximilans-Universität München, Munich, Germany*
[2]*Institute of Medical Education, Ludwig-Maximilians-Universität University Hospital, Ludwig-Maximilans-Universität München, Munich, Germany*
[3]*Department of Electrical and Computer Engineering, University of Kaiserslautern-Landau, Kaiserslautern, Germany*
[4]*Institue of Anatomy, Faculty of Medicine, Ludwig-Maximilans-Universität München, Munich, Germany*
[5]*Chair of Medical Physics, Faculty of Physics, Ludwig-Maximilans-Universität München, Garching, Germany*

(Received 18 October 2023; accepted 3 September 2024; published 4 November 2024)

Graphical representations of data are common in many disciplines. Previous research has found that physics students appear to have better graph comprehension skills than students from social science disciplines, regardless of the task context. However, the graph comprehension skills of physics students have not yet been compared with (veterinary) medicine students, both of which are disciplines that require multiple science, technology, engineering, and mathematics (STEM) courses. This study extends previous research on this subject by exploring whether physics majors possess superior graph comprehension skills due to their study discipline. Here, participants solved 24 graph comprehension tasks across various subjects, including mathematics, physics, and medicine; these tests were conducted at the beginning and end of their first semester. Graph comprehension gain was calculated based on the percentage of correct and incorrect answers in the pretest and the post-test. In addition to these comparisons, we replicated previous research that successfully distinguished correct and incorrect solvers based on their visual behavior by using a novel machine-learning method tailored to small datasets. Through this replication of statistical analyses, we aim to ensure the reliability of adaptive learning systems in the future, regardless of data size, using the same machine-learning method. Physics and medical students were found to exhibit relatively similar graph comprehension gain; this is in contrast to previous research comparing physics and non-STEM students. Our results also revealed that both physics and medical students use similar visual strategies to solve these tasks. However, correct and incorrect solvers could be distinguished via machine-learning methods regardless of their discipline. Our research suggests that visual behavior is a good predictor of graph comprehension skills.

DOI: 10.1103/PhysRevPhysEducRes.20.020138

I. INTRODUCTION

Graphical representations are common in daily life. They are also often used in educational contexts, most notably in science, technology, engineering, and mathematics (STEM) education. Graphical representations are frequently found in learning environments and their benefits to learners are well documented both theoretically and empirically [1]. Graphical representations of data (i.e., graphs) are important, particularly in disciplines where the use of data is common, such as the aforementioned STEM disciplines. As dealing with information is one of the key competencies of the 21st century [2], the ability to handle information is a crucial learning point in students' education. However, many students can encounter difficulties when dealing with graphical representations of data, such as interpreting the represented data [3–5].

Previous research has shown that physics students have a definitive advantage when solving problems associated with graphical information compared to students from social sciences disciplines [6–8]. As these studies compared

[*]Contact author: v.ruf@physik.uni-muenchen.de

physics majors with students from non-STEM disciplines, the specific reasons for the advantages exhibited by physics majors are unclear. For example, physics students could be better at solving graphical problems due to their chosen discipline or other factors (e.g., because they study a STEM discipline that requires the more frequent handling of data compared to social science disciplines). Other factors, such as higher mathematical skills, could also improve graph comprehension skills.

To extend previous research on this subject, we conducted a pretest-post-test study comparing physics majors to a group of non-STEM students that consistently engage with several STEM courses: veterinary medicine and medicine students. Both physics and (veterinary) medicine students are voluntarily enrolled in disciplines with many courses that engage with data and require a high level of mathematical knowledge. In this study, we investigate whether the discipline itself—i.e., whether the difference between physics majors, which consist mainly of mathematics and physics courses, and (veterinary) medicine majors, which contain physics, biology, and chemistry courses—plays a key role in the difference in students' graph comprehension skills.

Previous research also compared the eye movements of students from physics majors with those of students of other disciplines as they were solving graphical tasks from both their respective disciplines [6–8]. The results revealed that correct and incorrect solvers could be distinguished based on their eye movements using statistical analyses [7].

Eye movements can also be valuable inputs for adaptive learning systems [9]. These systems will probably become more common, especially with the advent of web-based eye tracking [10]. Adaptive learning systems use machine-learning methods to analyze data instead of statistical analyses and can provide immediate feedback and scaffolding to students during the learning and problem-solving processes [11]. Process measures, such as eye-tracking data, are ideally suited as input for machine-learning models because they not only act as indicators of a student's performance [12] but also provide insights into their problem-solving strategies [13]. However, machine-learning methods must be able to accurately and reliably replicate the results of statistical analyses to ensure the reliability of adaptive learning systems. As a first step, this study attempts to previous results by using machine-learning techniques to predict students' performance based on their eye movements.

This manuscript is organized as follows: we first introduce the relevant literature (Sec. II) before describing our methodology (Sec. III). We then present our results on the differences in the participants' graph comprehension gain as well as an analysis of the eye movement data collected during the study (Sec. IV). We proceed to discuss the results, possible implications for educational practice, and the study's limitations (Sec. V). Finally, we present our conclusions (Sec. VI).

II. LITERATURE REVIEW

A. Learning with graphs

Many STEM disciplines use graphical representations of data within their learning environments, such as velocity-time diagrams in physics courses. Previous research suggests that such representations facilitate data processing and comparison [14]. Consequently, graph comprehension is a common research topic on which multiple reviews have been published [3–5,15]. In the literature, two terms referring to the same skill are often used interchangeably: graph interpretation (e.g., [3]) and graph comprehension (e.g., [16]). This skill describes one's ability to "obtain meaning from graphs" [4] (p. 190), which either can be constructed by oneself or provided by others. Graph comprehension is a crucial ability for many students [2,4] and often requires domain knowledge, knowledge about the context of the graph, graphical skills (e.g., knowledge about the axes), and explanatory skills [17]. Shah and Hoeffner [15] described three general steps of graph comprehension: (i) identifying the relevant visual features, (ii) relating those features to the "conceptual relations that are represented by those features" [15] (p. 53), and (iii) connecting this information to the appropriate domain concepts. Prior knowledge can also influence a viewer's interpretations of a graph [4,15].

Interpreting graphs is a challenging activity for students [4]. The most common difficulties faced by students include interpreting a graph as a picture (graph-as-picture errors) [5,18] or understanding functions (e.g., translating between a function graph and the corresponding equation) [5]. For example, 10th-grade students, who were asked to create a position-time graph, did not recognize graph-as-picture errors even when revising the line of their graph [18]. In physics, additional difficulties include identifying the relevant features of a graph or interpreting the data depicted in a graph [19]. Graph comprehension difficulties are also common in medical fields: for example, participants had trouble identifying conflicting information in graphs that presented medical information, such as treatment effects [20]. Furthermore, relating mathematical concepts with their contextual meaning is often not intuitive [21].

The context and characteristics of a discipline can influence graph comprehension [16]. Due to its importance across many disciplines, it is important to ensure that students can interpret graphical representations independently of their context. *Representational competence* refers to a student's knowledge of how content is presented in visual representations. It involves visual understanding, including the ability to identify relevant variables and information, as well as the ability to connect these representations to their respective concepts [22]. The ability to use representations and understand their meaning independently of the context is a part of *meta-representational competence* [23]—this is a key aspect of representational competence [22] and is a valuable skill in the context of graph comprehension [24]. However, students gain content knowledge at the same time as they acquire representational competence [22]; this prior knowledge and experience can contribute to learning. This is often assumed to be the case when utilizing previously learned procedures in a new situation [25], such as when students use the same method to calculate the slope of a function in both math and physics problems. Ideally, students should be able to apply their prior knowledge of problem-solving routines across multiple disciplines, such as calculating the slope of a line graph in both math and physics using the correct units. Once again, this requires representational competence, such as the ability to correctly identify which variables should be present on each axis, their units, and domain knowledge about the relationship between those variables. Cognitive processes during problem solving, such as identifying relevant features in a graph and the relationship between those features, can be inferred from eye movements [26].

B. Eye movements

Eye movement analyses, such as the length of fixations, are often used in educational research [26,27] because they can indicate attentional processes (eye-mind hypothesis, [28]). Education studies frequently analyze the differences between experts and nonexperts in the context of STEM education; for example, in terms of domain knowledge, e.g., physics education research [29] or graph comprehension skills [30]. In particular, information processing can differ between experts and nonexperts. There are several theories as to why this might be the case: these include the theory of long-term working memory [31], the model of holistic image processing [32], and the information-reduction hypothesis [33]. All theories imply that experts engage in more efficient information processing, which can be investigated via their eye movements. According to the theory of long-term working memory, experts store larger chunks of information in working memory compared to nonexperts [34] and they access these chunks via retrieval cues [31]. Consequently, experts' fixation durations should be shorter due to the reduced time needed to retrieve information [35]. In contrast, the holistic model of image perception approaches the issue from a perceptual level: experts process images globally [32], which leads to more efficient searches [35]. This is assumed to lead to quicker fixations on relevant information [35,36]. Finally, the information-reduction hypothesis states that experts can focus on relevant information by ignoring information irrelevant to the current task [33]; this suggests that experts should fixate on relevant information longer [35].

Such differences in expertise were also found in the visual strategies used during graph comprehension [30,35,36]. Experts paid more attention to relevant information than nonexperts during learning and problem-solving involving graphs [30]. This is consistent with the information-reduction hypothesis [33]. Previous research compared participants with varying levels of expertise as they solved science-related problems that used conventional representations of data: they found that participants of differing expertise levels used different visual strategies [37]. Okan *et al.* [20] observed that participants with high graph literacy looked at relevant graphical information for longer periods when making their interpretations. Differences in expertise may also be influenced by specific disciplines. In a study on 131 high-school students, Becker *et al.* [38] found that the participants had problems applying mathematical knowledge to a kinematic context, especially for line graphs with negative gradients. Similar results were reported by Ceuppens *et al.* [39]. Furthermore, students tended to look along the line of the function when attempting math problems while instead choosing to pay closer attention to the axes when attempting kinematics problems [38]. In previous studies with higher-education participants, physics students solved graphical problems better than both economics students [6,7] and psychology students [8], regardless of the subject. Susac *et al.* [8] measured and compared the performance and viewing times of psychology ($N = 45$) and physics ($N = 45$) students as they attempted to solve isomorphic physics and finance problems. An analysis of students' strategies indicated that physics students relied on equations for a solution, while psychology students used more "common-sense strategies" [8] (p. 1) (i.e., rise over run) that were more likely to lead to errors, such as confusing points and intervals [8]. Klein *et al.* [7] replicated these results and found that physics students ($N = 29$) performed better than economics students ($N = 40$) on the same isomorphic test items even though the finance problems were, from a content perspective, more closely related to the field of study of the economics students. They also found that students who solved the tasks correctly focused longer on concept-specific areas than those who did not [7]. This was consistent with Susac *et al.* [8] who found that physics students analyzed the actual graph more carefully compared to psychology students. Students were also asked to report their confidence in their answers—these scores revealed that physics students were better judges of the correctness of their answers compared to economics students, although the scores did not differ significantly between groups [7]. These results were replicated in a postreplication study employing a pretest-posttest design [6]. Both physics ($N = 20$) and economics ($N = 21$) students solved the same problems as participants in the previous studies both at the beginning and the end of their first semester. Importantly, the participants were a matched sample to the participants in the pretest study conducted by Klein *et al.* [7]. In the post-test, it was once again observed that physics students performed better than economics students. However, "students from both domains showed a similar increase in the overall test score" [6] (p. 8).

In addition, there were no differences in visual behavior between the correct and incorrect solvers from different disciplines, although physics students tended to be overconfident in their answers during the post-test [6].

C. Machine learning using eye movements

Previous studies used statistical methods to compare correct and incorrect solvers' eye movements and visual processing. For example, Klein *et al.* [7] used ANOVAs to compare the eye movements of physics and economics students. Statistical tests, such as regression analyses (i.e., ANOVAs), can be used to predict the influence of independent variables, such as eye movements, on the dependent variable, such as problem-solving performance [40]. Whereas statistical tests are a well-established method to analyze data for hypothesis testing [41], adaptive learning systems often use large datasets with many independent variables and possibly complex correlations [42]. Therefore, such adaptive systems frequently employ machine-learning techniques, because they outperform statistical methods [42]. This is probably due to the advantages of machine learning when analyzing complex nonlinear relations between variables. Such considerations need to be kept in mind, considering the increasing number of personalized and adaptive learning systems [43].

Statistical tests, such as ANOVAs—that were previously used to analyze eye movements [7], assume that data are based on underlying stochastic models [44]. In contrast, algorithmic models employing machine-learning methods assume that the underlying data distribution is unknown [44]. Machine-learning methods can use training data to generate generalizable and scalable models, which can be applied to unknown datasets of different sizes [45]. Such models are more robust to changes if the distribution of new incoming data differs strongly from the original data on which the model is trained. Machine learning can also be used to investigate complex (e.g., nonlinear or unknown) correlations [45], allowing users to find new patterns in the data without making previous assumptions regarding the data's distribution. An overview comparing statistical and machine-learning methods is presented in Table I. Combining the two approaches—statistical modeling and machine learning—in a hybrid approach can have various advantages, such as investigating causal effects while leveraging the benefits of machine learning by identifying the most predictive variables [41]. In this way, causality can be examined by comparing models trained with multiple variables and based on various assumptions; additionally, the predictive power of scientific theories can be tested [41]. This approach is suitable due to the increasing use of machine-learning techniques for complex analyses with real-time predictions on large datasets and the benefits of statistical methods employed to test hypotheses. Currently, datasets in education are often small (i.e., 29 physics students and 40 economics students [7]), whereas we believe that large amounts of data will be available to future adaptive learning systems. At the moment, small datasets make applying machine-learning techniques more difficult due to the limited amount of training data, which increases the necessity of testing the suitability of machine learning in a hybrid approach—including the benefits of both statistical and machine-learning analyses. Using an optimized method specifically designed for small datasets (i.e., Ref. [46]) in such an approach can lead to even more reliable results.

Most machine-learning methods used in educational contexts are supervised machine-learning models [42]; this refers to algorithms that can predict the outcome of unknown data based on labeled training data [45]. Only a portion of the collected data is used for training; in general, only 80% of the data is used as training data, with the remaining 20% withheld as test data [45]. Machine-learning tools can be applied to evaluate eye-tracking data: Supervised machine-learning methods have previously been employed to predict performance based on eye-tracking metrics [12,49,50]. Küchemann *et al.* [12] successfully predicted the performance of 11th-grade high-school students' ($N = 115$) on the Test of Students' Understanding of Graphs in Kinematics [51] based on the total visit duration on areas of interest (AOIs) and the frequency of gaze switches between AOIs with a support vector machine (SVM). Mozaffari *et al.* [52] and Rebello *et al.* [53] achieved similar results when students attempted

TABLE I. An overview of statistical and machine-learning methods.

	Statistical methods	Machine-learning methods
Purpose	Hypothesis testing [41] Statistical inference [41]	Model complex relationships [44,45] Discover new patterns
Focus	Investigating the influence of relation between individual variables [44]	Making accurate predictions [41,44]
Measurement of variable importance	Exact coefficients [47]	Feature importance based on the trained model [45]
Proposed application	Theory-based research [41,44]	Adaptive learning systems [42,48]
Recommended use cases	Simple assumptions Complex assumptions	Large dataset with complex assumptions

physics-based tasks. In addition to duration metrics, machine-learning models can predict performance using other eye-tracking metrics, such as the number of saccades and pupil size [54]. From an educational perspective, this research is highly relevant to adaptive learning systems, which have become more relevant due to the increasing use of artificial intelligence in education [48]. For example, adaptive learning systems can support students during problem-solving tasks by prompting them if they do not look at relevant areas.

In this study, we compare several machine-learning algorithms, including an SVM. SVMs have already been successfully applied to learning analytics [55] and have also been used in studies attempting to predict task performance based on visual behavior [12,49,52,53]. SVMs can be used as binary classifiers, capable of separating two groups by defining a boundary that attempts to maximize the margin of error on either side [45]. They can also be used in real time [55,56]. Other classification algorithms include the k-nearest neighbor (KNN), which forms groups based on the distance between data points and random forest (RF) models, which separate data points into groups based on an ensemble of decision trees [45]. Dzsotjan *et al.* [49] compared the accuracy of various machine-learning algorithms trained to predict performance based on visual behavior in an interactive augmented reality environment finding that SVMs returned high F1 scores. This suggests that eye movements are suitable inputs that can be used to predict students' performance on graph comprehension tasks using machine-learning techniques such as SVMs.

D. Research questions

Previous studies have compared physics students with non-STEM students in terms of their performance at graph comprehension tasks and found that physics students outperformed non-STEM students regardless of context [6–8]. In this study, we extend this research comparing physics majors with non-STEM students [6–8] by comparing the graph comprehension skills of physics majors with non-STEM students who also take STEM-related courses. To the best of our knowledge, no studies have compared the visual behavior of physics majors (STEM students) with that of (veterinary) medicine students (non-STEM students) who take STEM courses, such as anatomy, chemistry, and physics. In particular, while mathematics is a key subject for physics majors, medical students do not take any math classes. Based on the results of previous studies, we pose the following research questions:

(1) Are there differences in graph comprehension gain of physics and non-physics majors?
(2) Can differences in students' performance be accurately identified via machine-learning models trained on students' visual behavior?

If physics students perform better due to their discipline rather than the amount or type of STEM courses taken, then physics students should outperform medicine students as was found in previous studies comparing physics students to social sciences students [6–8]. However, if STEM courses are relevant to graph comprehension capabilities in general, then both the physics and veterinary medicine/medical students in this study should exhibit equal graph comprehension.

To answer the second research question, we replicated the results of Klein *et al.* [7] and assessed the visual differences between correct and incorrect solvers with machine-learning methods. Eye movements can be used as input for adaptive learning systems [9] and as a predictor of performance (Sec. II C). However, machine-learning models must be both accurate and reliable in order for them to be successfully employed as personalized systems in educational environments. In other words, it must be possible to replicate the statistical findings of machine-learning methods, even for small datasets common in educational contexts.

III. METHODS

We recruited first-semester students as participants in our pretest-post-test study. The participants' demographics, study design, test materials, apparatus used in the study, and machine-learning techniques used in this study are described in the following sections.

A. Participants

Twenty-four first-semester students from the Ludwig-Maximilians-Universität München (LMU Munich) voluntarily participated in the study at the beginning and the end of the semester. The students were enrolled in the physics ($N = 9$), physics education ($N = 3$), medicine ($N = 4$), or veterinary ($N = 8$) programs. A typical first semester in physics includes only STEM courses: experimental physics, mathematics, and physics labs. Both medical and veterinary students take the chemistry and anatomy STEM courses as well as an additional STEM course (medicine: biology, veterinary medicine: physics) and non-STEM course (medicine: medical terminology, veterinary medicine: zoology), depending on their vocation. In particular, mathematics is typically not part of the medical curriculum. Physics majors ($\mu = 19.25$ years) were the same age as the veterinary and medical students ($\mu = 19.75$ years); $t(15.177) = -0.62$, $p = 0.55$. Participants received €20 for their participation in both the pretest and the post-test. Most students were native German speakers; one student spoke German at the C1 level. The participants' mean A-level grade was 1.5 (i.e., the middle grade between A and B in the United States). There was no significant difference between the grades of physics majors ($\mu = 1.72$) and medical students ($\mu = 1.34$), $W = 44$, $p = 0.30$.

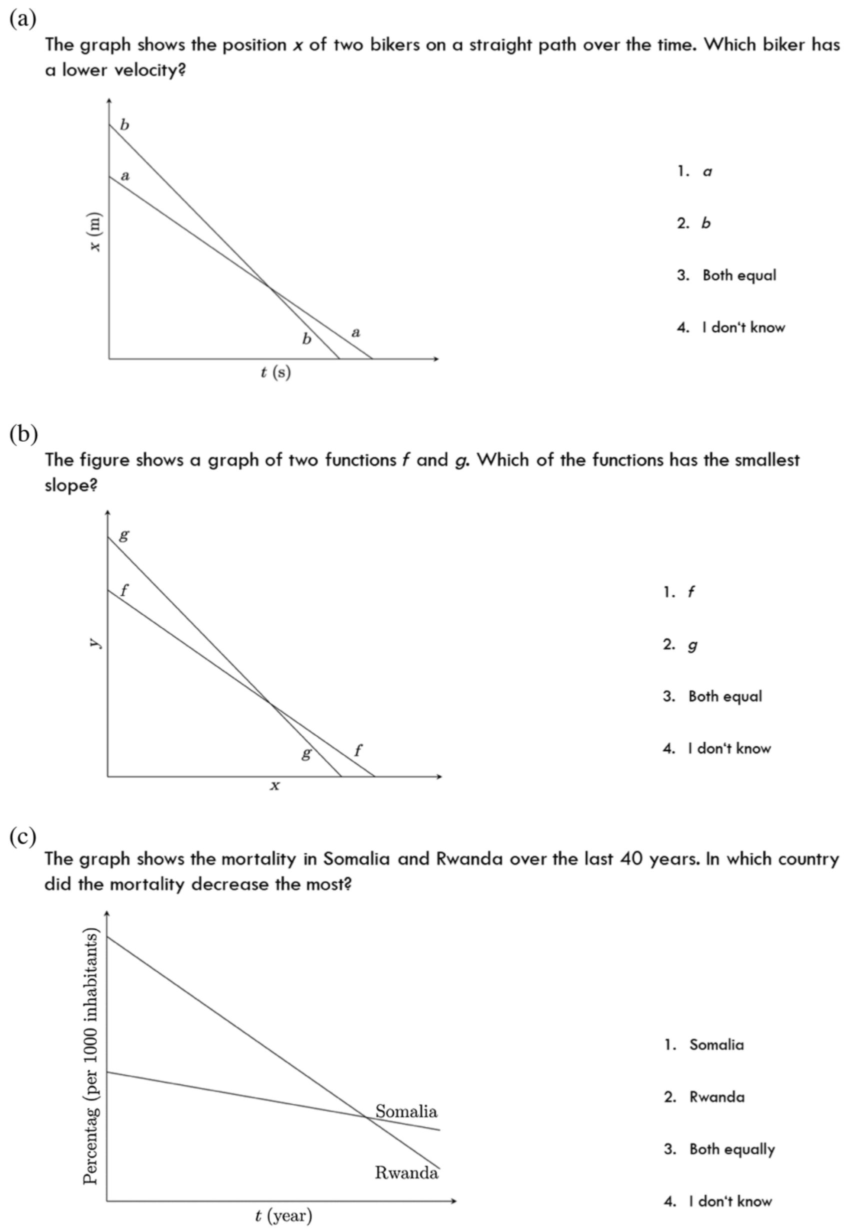


FIG. 1. Example math (a) and physics (b) items from Ceuppens *et al.* [39] as well as (c) the isomorphic medical item. The text is translated from the original German (Fig. 2).

B. Procedures and study design

The ethics committee of the mathematics, computer science, and statistics faculty found the pretest and the post-test to be ethically sound (EK-MIS-2022-122, EK-MIS-2023-143). Students from introductory physics and medicine courses at LMU Munich were recruited in the winter semester of the 2022/23 academic year. Participants answered the same questions at the beginning and end of the semester. The pretest took place in November 2022, while the post-test was conducted in February 2023.

Students signed a consent form and created a pseudonymized code they used to take the pretest and the post-test. The pretest also included a demographic questionnaire. Following this, the students attempted the given tasks.

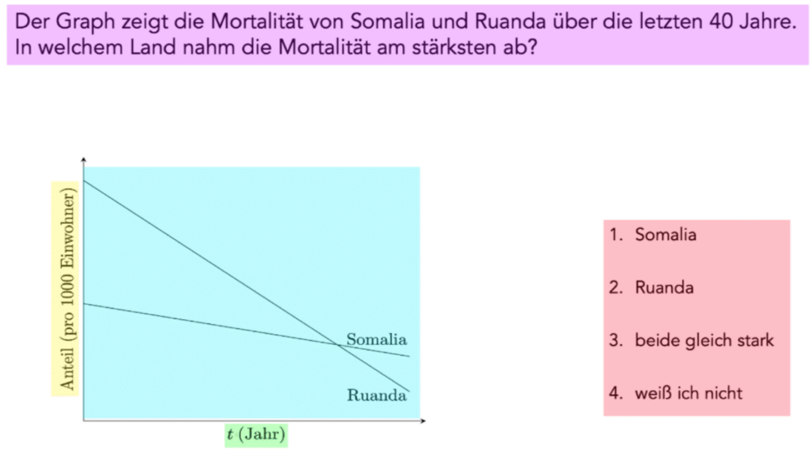


FIG. 2. The areas of interest (AOIs) on an exemplary test item. Selected AOIs include the text, answers, graph, labels, and axes (item presented in German, see Fig. 1(c) for the English translation).

C. Test material

The test material was based on instruments used in previous studies [8,39] and focused on assessing graph comprehension (Sec. II A). The test consisted of 24 items, composed of eight mathematics, physics, and medical items (*subjects*), respectively. Physics and medical items were posed in the context of their respective fields, while mathematical items were not presented in the form of a real-world scenario. Of the eight items in each of the three subjects, four were concerned with the area under the graph, while the other four were related to the slope of the graph (*concepts*). Example items are presented in Fig. 1. The complete test is available in Supplemental Material [57]. The participants' answers were rated as either correct (1) or incorrect (0).

D. Eye-tracking apparatus and measurements

Participants sat in front of a 24-inch computer screen with a resolution of 1920×1200 pixels and a refresh rate of 75 Hz. The distance to the screen was about 60 cm. A Tobii Nano eye tracker with a sampling frequency of 60 Hz and an ideal accuracy of 0.30° of the visual angle (according to the manufacturer) was used to collect eye movement data. The system allows for a high degree of movement; consequently, a chin rest was not necessary. More information on the eye tracker used in this study can be found in Tobii [58]. Fixations were detected using an I–VT algorithm [59].

Eye-tracking metrics were calculated based on specific stimulus regions that were identified as areas of interest (AOIs) [60]. We chose the areas around the text, the graph, and the answers as AOIs (Fig. 2). We also extracted the total viewing time (also known as "dwell time," see Ref. [6]) for these AOIs for comparison with the previous research [6–8].

E. Classification via machine learning

This study aimed to replicate the findings of Klein *et al.* [7]; i.e., it aimed to distinguish correct and incorrect solvers based on their eye movements—using machine-learning methods. In line with the hybrid modeling approach combining statistical and machine-learning methods [41], we used a two-step process: First, we conducted a statistical linear regression analysis, allowing us to account for the statistical influence of independent variables, such as study discipline and the type of AOI, on visual behavior (dependent variable). Additionally, possibly nonrelevant variables can be excluded to improve the efficiency of the machine-learning model [12,61]. All statistical analyses were conducted in R. The package "lm.beta" was used for regression analyses.

Second, this statistical analysis was extended using an analysis based on machine-learning methods to answer the second research question: Can differences in students' performance be accurately identified via machine-learning models trained on students' visual behavior? We aimed to replicate the results of Klein *et al.* [7] using machine-learning techniques and trained various machine-learning models to predict the participants' performance (independent variable) based on their total viewing time on the AOIs (dependent variable). We (a) compare several models (SVM, KNN, and RF; Sec. II C) to assess the models' performances and (b) compare a conventionally trained SVM with an optimized model [46]. We adopted a machine-learning method suited to small datasets that improves reliability and reduces bias [46]. We used fivefold cross-validation in conjunction with permutation tests and hyperparameter tuning to train the algorithm and evaluate its performance using the Matthews correlation coefficient (MCC). This method is particularly suited to small datasets [46]. We compared this optimized method with a simple

TABLE II. An overview of the results of the multiple linear regression used to assess the importance of the independent variables on viewing time, including study discipline (step 1), AOI (step 2), test date (pretest vs post-test; step 3), as well as the subject (mathematics vs physics vs medicine; step 4) and concept tested by each task (area vs slope; step 5), Note: [a]The adjusted R^2 indicates how well the model explains the observed outcome. [b] SE B is the standard error of the unstandardized beta, which indicates whether the b-value differs significantly from 0. [c] β indicates how important a predicts is. For further information regarding these metrics, please refer to Ref. [47].

	Adjusted R^{2}[a]	SE B[b]	β[c]	p[d]
Step 1	0.0002			0.15
Constant		4.06 (0.11)		<0.001***
Physics		0.23 (0.16)	0.02	0.15
Step 2	0.2			<0.0001***
Constant		2.71 (0.19)		<0.001***
Physics		0.22 (0.14)	0.02	0.11
AOI axes		0.84 (0.24)	0.05	0.01**
AOI graph		5.79 (0.24)	0.33	<0.001***
AOI text		4,62 (0.24)	0.27	<0.001***
AOI x-label		−1.32 (0.24)	−0.08	<0.001***
AOI y-label		−1.82 (0.24)	−0.10	<0.001***
Step 3	0.2			<0.001***
Constant		2.17 (0.2)		<0.001***
Physics		0.18 (0.14)	0.01	0.2
AOI axes		0.84 (0.24)	0.05	0.01**
AOI graph		4.62 (0.24)	0.33	<0.001***
AOI text		5.79 (0.24)	0.27	<0.001***
AOI x-label		−1.32 (0.24)	−0.08	<0.001***
AOI y-label		−1.82 (0.24)	−0.10	<0.001***
Test pre		1.09 (0.14)	0.08	<0.001***
Step 4	0.23			<0.001***
Constant		0.8 (0.2)		<0.001***
Physics		0.18 (0.14)	0.14	0.2
AOI axes		0.84 (0.24)	0.04	<0.001***
AOI graph		4.62 (0.24)	0.33	<0.001***
AOI text		5.79 (0.24)	0.27	<0.001***
AOI x-label		−1.32 (0.24)	−0.08	<0.001***
AOI y-label		−1.82 (0.14)	−0.10	<0.001***
Test pre		1.09 (0.14)	0.08	<0.001***
Subject medicine		2.61 (0.17)	0.19	<0.001***
Subject physics		1.48 (0.17)	0.11	<0.001***
Step 5	0.23			<0.001***
Constant		1.04 (0.23)		<0.001***
Physics		0.18 (0.14)	0.14	0.19
AOI axes		0.84 (0.24)	0.05	<0.001***
AOI graph		4.62 (0.24)	0.33	<0.001***
AOI text		5.79 (0.24)	0.27	<0.001***
AOI x-label		−1.32 (0.24)	−0.08	<0.001***
AOI y-label		−1.82 (0.24)	−0.10	<0.001***
Test pre		1.09 (0.14)	0.08	<0.001***
Subject medicine		2.61 (0.17)	0.19	<0.001***
Subject physics		1.49 (0.17)	0.11	<0.001***
Concept slope		−0.5 (0.14)	−0.04	<0.001***

SVM trained on 60% of the dataset and tested on the remaining 40% of the data. This split was used to ensure that there was enough data to calculate feature importance. We used an SVM to predict performance based on the participants' total viewing time on specific AOIs because previous studies have shown that this algorithm can be successfully applied to similar tasks [52,53]. In the context of binary classification, SVMs attempt to find the ideal hyperplane that separates groups of data points by maximizing the margin between the data and the hyperplane [45]; consequently, SVMs can be considered to be a type of regression analysis. Due to its ability to transferability to unknown data and identify patterns in complex data (Table II), we believe that implementing an SVM would be an ideal way of comparing the results of machine-learning regression techniques to statistical regression. This will allow us to compare the variable (statistics) and feature importance (machine learning) of both types of analyses.

IV. RESULTS

Here, we present the results of the participants' answers and the analysis of their visual behavior. A p value of less than 0.05 was used as the threshold for a statistically significant test statistic [47]. The assumptions of all statistical tests, such as normal distribution, were checked and nonparametric tests were used if these assumptions were violated.

A. Graph comprehension gain

The participants answered the same items in both the pretest and the post-test. A Wilcoxon signed-rank test revealed that students took significantly longer to solve the pretest ($\mu = 16.8$ min, $\sigma = 5.1$) than the post-test ($\mu = 13.2$ min, $\sigma = 3.5$); $V = 40$, $p = 0.001$. 916 items were answered correctly, while 236 items were answered incorrectly. The graph comprehension gain was assessed by subtracting the percentage of correct answers in the pretest from the percentage of correct answers in the post-test. Across the entire sample, a dependent t test showed no significant difference in the graph comprehension gain of students on the area under the curve and the slope tasks; $t(45.98) = 0.42$, $p = 0.68$. A Wilcoxon signed-rank test revealed that physics majors exhibited improved graph comprehension gain with respect to the tasks associated with the area under the curve ($\mu = 0.11$, $\sigma = 0.18$) compared to tasks associated with the slope of the graph ($\mu = -0.02$, $\sigma = 0.19$); $W = 166663020$, $p < 0.001$. In contrast, the opposite was observed for (veterinary) medicine students, who exhibited improved performance on the slope tasks ($\mu = 0.05$, $\sigma = 0.2$) compared to the tasks associated with the area under the curve ($\mu = -0.01$, $\sigma = 0.24$); $W = 81328200$, $p < 0.001$. As there were no overall differences across the sample, and because the improvements in the performance of the students on different types of graph comprehension tasks were in opposition, we used the mean graph comprehension gain between the slope and area tasks for the remaining analyses. The mean graph comprehension gain for physics and (veterinary) medicine students across the three subjects tested is presented in Fig. 3.

The graph comprehension gain of physics and medical students across the three different subjects assessed by the test material (mathematics, physics, and medicine) was compared via a 2 (discipline) × 3 (subject) ANOVA. Graph comprehension gain did not differ between physics ($\mu = 0.05$, $\sigma = 0.14$) and medical students ($\mu = 0.05$, $\sigma = 0.19$; $F(1, 48) = 0.01$, $p = 0.9$, $\omega^2 = -.02$). However, graph comprehension gain did differ significantly between subjects; $F(2, 48) = 4.11$, $p = 0.02^*$, $\omega^2 = 0.11$; this can be considered a medium effect [62]. Bonferroni *post hoc* tests showed that the graph comprehension gain in medical items ($\mu = -0.02$, $\sigma = 0.19$) was significantly smaller than the graph comprehension gain in physics items ($\mu = 0.13$, $\sigma = 0.16$).

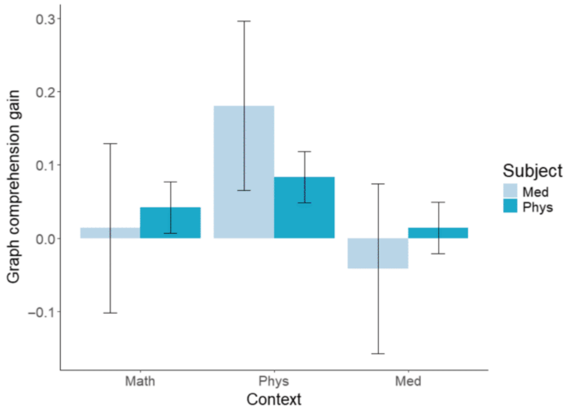


FIG. 3. Mean graph comprehension gain for students from each discipline across the different subjects. Error bars indicate the standard deviation of the results.

B. Total viewing time

This section presents the results of the viewing times analyses conducted on the eye-tracking data. The effects of the independent variables on the viewing time were first assessed using linear regression. This revealed several independent variables that should be considered when designing the machine-learning model, specifically viewing times on AOIs and the test date. Based on these results, an SVM was trained and the results were compared with optimized and non-optimized machine-learning algorithms as described in Section III E [46]. We also investigated the feature importance of the final SVM model.

1. Statistical analysis

We used multiple linear regression to determine the effects of the following independent variables on the total viewing time (dependent variable): the students' discipline, parameters associated with specific AOIs (Fig. 2), the test date, the subject of each task (Fig. 1), and the concept tested. In this way, the effects of all possible influences on visual behavior were evaluated. The results of linear regression models are presented in Table II. An ANOVA revealed that each step significantly improved the fit; $p < 0.001$. The first model—which only included the participants' discipline to predict their viewing time—did not perform well, adjusted $R^2 = 0.0002$. The model's performance was significantly increased by adding the AOI as a predictor variable, adjusted $R^2 = 0.02$. Adding the test date as a variable did not increase performance, adjusted $R^2 = 0.02$. However, adding the subject assessed by each item improved the model, adjusted $R^2 = 0.023$. Adding the specific graph comprehension concept tested did not improve the model. These results suggest that 23% of the variance in viewing time can be explained by this linear regression model, which includes the viewing time associated with each AOI and the subject assessed by the task.

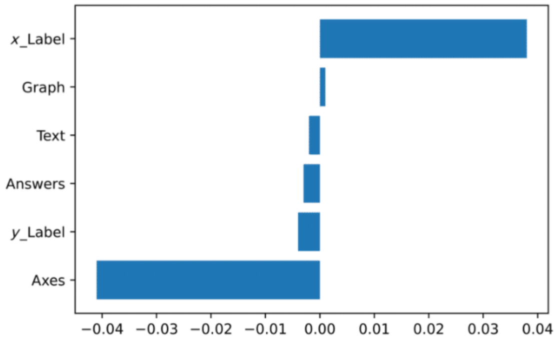


FIG. 4. Feature importance coefficients derived from the simple SVM.

2. Machine-learning analysis

The insights gained from the multiple linear regression were used to guide the design of the machine-learning algorithm by looking at the influence of the individual variables. Participants' performance was predicted using the viewing time on each AOI for every item because this seemed to be the most relevant variable based on the linear regression results. Reducing the number of features can also increase the efficiency of the machine-learning model [12,61]. Separate models were built for the pretest and post-test datasets, as this was a significant variable in the regression analysis. This study aimed to replicate the results of Klein *et al.* [7] via machine-learning methods. Specifically, we used a machine-learning method tailored to small datasets to train an SVM; the final model was evaluated using MCC to judge its performance on unbalanced datasets and to reduce bias [46]. The results revealed that correct and incorrect solvers could be distinguished based on the AOIs, although the MCC values of the final model were low (pretest results: MCC $= 0.23$, $p = 0.001$; post-test results: MCC $= 0.22$, $p = 0.001$). The performance of the SVM was compared to a KNN and an RF model; these models had slightly lower MCC scores (KNN pretest results: MCC $= 0.08$, $p = 0.02$; KNN post-test results: MCC $= 0.13$, $p = 0.01$; rf pretest results: MCC $= 0.14$, $p = 0.01$; rf post-test results: MCC $= 0.18$, $p = 0.01$).

It is important to note that the method used to generate these models (Sec. III E) was designed specifically for small datasets and is not a conventional practice. Analyses using classical train-test-split methods led to different results. Using a simple SVM without cross validation with a 60–40 train-test split, produced a model that scored well on a variety of performance metrics (pretest metrics: accuracy $= 0.78$, F1 $= 0.87$.; post-test metrics: accuracy $= 0.82$, F1 $= 0.90$). However, the MCC of the model was 0.0 in both cases, which suggests that the method developed for small datasets led to more reliable results. The feature importance of the simple SVM instantiated with a linear kernel supported these results. Feature importance coefficients suggested that the x-label and the axes were the most important features when attempting to distinguish correct from incorrect solvers based on their visual behavior (Fig. 4). It is important to note that these coefficients were calculated on the training data.

Permutation feature importance uses the test data to calculate feature importance; this technique randomly shuffles the features, which generally causes the model to perform worse. A feature's relevance is determined by how much worse the model performs on the test set when the feature is randomly permutated. Permutation feature importance for the simple SVM revealed that the viewing time on the axes was the most important feature in distinguishing between correct and incorrect solvers, followed by the viewing time spent on the graph (Fig. 5). Notably, the feature importance coefficient regarding viewing time spent on the AOIs associated with the axes of the graph changed from negative to positive when permutation feature importance was considered; in addition, the high x-label coefficient observed in the standard feature importance was not observed when permutation feature importance analysis was conducted. These results suggest that the simple SVM was not optimally trained, although the prediction's accuracy appeared to be high.

V. DISCUSSION

This section discusses the first-semester physics majors and (veterinary) medicine students. We investigated graph comprehension gain and differences in visual behavior for students faced with math, physics, or medical graph comprehension tasks. We first compare the task performance of the physics and (veterinary) medicine students for each of the three subjects (math, physics, and medicine). We then evaluate the participants' visual behavior based on their total viewing time on each AOI (Fig. 2).

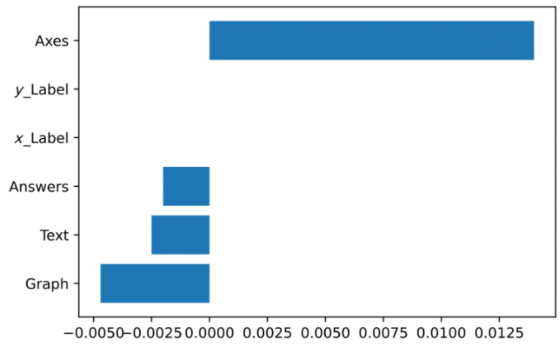


FIG. 5. Permutation feature importance of the simple SVM.

A. Comparison of graph comprehension gain between physics and medical students

Previous research found that physics students performed better at graph comprehension tasks than psychology [8] and economics students [6,7]. In this study, we compared physics with (veterinary) medicine students and found no significant differences in graph comprehension gain over the first semester. These results suggest that first-semester physics and medicine students have similar graph comprehension skills when faced with problems associated with mathematics, physics, and medicine.

The participants were tasked with solving isomorphic graph comprehension problems across three subjects and used the same visual behavior to solve each question. Participants' equal performance across different subjects suggests that all participants—i.e., both physics majors and (veterinary) medicine students—were able to apply the solution strategy regardless of subject. This also suggests that students were able to identify the relevant variables in all graphs regardless of subject and were able to relate the data presented in each graph to the subject of the question being asked. One explanation for this observation is that both physics majors and (veterinary) medicine students possessed a similar level of representational competence: Both physics and medicine students appeared to exhibit a comparable ability in terms of the use of problem-solving strategies in isomorphic representations across different subjects (i.e., familiarity with graphing conventions and connecting the representations to the concepts).

Another reason for the comparable performance between the physics majors and (veterinary) medicine students in this study compared to the social science students assessed in previous studies [6–8] could be the access to STEM courses: social sciences students typically only have one STEM course during the first semester (statistics). In contrast, all students participating in this study attended more than one STEM course. STEM courses provide students with a real-world context for the abstract mathematical concepts taught in their respective disciplines; this may occur concurrently with learning about their application through experimentation themselves, such as in physics labs. Thus, the description of phenomena in abstract mathematical terms could potentially enhance a student's ability to transfer these problem-solving skills to other subjects.

Each item was relatively simple (see examples in Fig. 1); some tasks were intended for 9th-grade students [39]—i.e., first-semester university students would be expected to solve them without much difficulty. This was reflected in the high degree of accuracy across all subjects and disciplines: 80% of the items were solved correctly. Another potential reason for this similar performance among the participants could be their relatively equal skill level as shown by comparable A-level grades of students from both disciplines. Furthermore, both physics majors and (veterinary) medicine students voluntarily chose disciplines that require a high degree of graph comprehension skills: the study of physics generally involves graphs containing experimental data, while graphs that depict patient data are common in medicine. Physics and (veterinary) medicine students would presumably also be familiar with the STEM courses that are taken as part of their degree and can be assumed to be proficient STEM learners. For example, physics majors are believed to have an aptitude for science [63]. Finally, students did not appear to improve between the pretest and post-test, which could indicate that they possessed an overall high level of graph comprehension skills even before starting their university studies.

In contrast to previous studies [6–8], we did not find any significant differences in graph comprehension between physics majors and medical students. This suggests that students who take STEM courses have a similar level of graph comprehension skills. Although we have assumed that STEM courses play a role in facilitating graph comprehension, we cannot conclude that STEM courses are the sole *reason* for the participants' comparable graph comprehension skills: Other factors, such as aptitude and self-selection of the university discipline, may also play a role. However, we did find that correct solvers used different visual processes to solve the graph comprehension tasks compared to incorrect solvers. This suggests that some students are more successful in identifying the relevant variables (correct solvers) compared to others (incorrect solvers), suggesting different levels of representational competence.

B. Viewing times of correct and incorrect solvers

The second research question posed in this study concerns the total viewing time on specific AOIs for correct and incorrect solvers: Can differences in students' performance be accurately identified via machine-learning models trained on students' visual behavior? We addressed this question in two ways (Sec. III E): We first conducted a statistical analysis to determine the relative influence of each independent variable on the dependent variable viewing time. We then used the total viewing time on each AOI to predict performance. Finally, we compared several machine-learning models, which included an SVM trained using a method developed specifically for small datasets [46]. An SVM, a type of machine-learning-based regression, ensures that the models' feature importance is comparable to the variable importance obtained from a statistical regression.

Multiple linear regression revealed that the total viewing time did not depend on the participants' discipline, instead, it was most strongly influenced by the different AOIs (Table II). Specifically, longer viewing times were associated with increased time spent looking at the graph and the text and reduced time spent looking at the axes labels. The model's accuracy increased when the subject tested by the item was added as a variable (from adjusted $R^2 = 0.2$ to adjusted $R^2 = 0.23$; Table II). In particular, tasks that posed problems related to medicine and physics appeared to influence the viewing time; medical items required slightly longer viewing compared to physics items. These results were consistent with our overall observations on graph comprehension gain: medical tasks appeared to be more difficult than physics tasks (Fig. 3). Consequently, participants are likely to spend more time on tasks with greater difficulty. Participants' graph comprehension gain did not differ between physics, medicine, and mathematics (Sec. IV A).

This study also aimed to replicate previous findings regarding the differences in the visual behavior of correct and incorrect solvers [7] with machine-learning methods. The SVM model performed better than the KNN and rf models. We also found significant differences between correct and incorrect solvers in the pretest and the post-test. This suggests that machine-learning algorithms can distinguish between participants who either correctly or incorrectly approached a graphical comprehension task. These results are consistent with previous research, suggesting that the total viewing time spent on concept-specific areas can differ between correct and incorrect solvers [7,30]. Although the performance of the optimized machine-learning algorithm was not as good as the models presented in previous research, the results remained satisfactory. For example, Dzsotjan *et al.* [49] reported an F1 score of 0.66 when using a combination of the best features, which included features other than the viewing time on AOIs. It should be noted that an F1 score is also more likely to support the null hypothesis than an MCC score [46]. This is supported by the high accuracy (pretest: 0.78, posttest: 0.82) and F1 scores (pretest: 0.87, posttest: 0.90) obtained when using a simple SVM to predict participant performance. These results would suggest that a simple SVM might be a good model for the data; however, the MCC of the SVM model was 0.00 for both pretest and post-test datasets, indicating that this was not the case. In contrast, the MCC for the optimized model, which used cross-validation techniques and permutation tests, was much higher, suggesting that it is better suited to predict performance based on eye-tracking data.

It is important to note that the feature importance of the simple SVM model changed depending on how it was calculated. The coefficients used by the SVM for the training data suggested that the viewing time on the x-label and the axes were the most important features (Fig. 4). However, when feature importance was calculated using permutation tests on the test data, the second most relevant feature was the viewing time spent on the graph (though the axes were the most relevant features; Fig. 5). These observations were more consistent with the results of the multiple linear regression, which suggested that the viewing time spent on the graph was the most important variable. Therefore, we confirmed that machine learning seems to be a valid method to predict students' performance based on their visual behavior [12,49,50]; particularly, with an optimized machine-learning method developed for small datasets [46], which yields results comparable to classical statistical methods. The accuracy of the feature importance tests depends on the size of the test set; this study used a test set size of 40% to ensure there was enough data to calculate feature importance. As the feature importances derived from the simple SVM were similar but not identical to the results of multiple linear regression analysis, the size of the dataset appears to be too small for simple machine-learning algorithms. This highlights the importance of using an appropriate method to analyze small datasets.

C. Implications for practice

This study investigated the differences in graph comprehension skills between physics majors and (veterinary) medicine students and found that students from the two disciplines had relatively similar graph comprehension skills. This was in contrast to prior research that reported that STEM majors tended to outperform non-STEM social science students who traditionally do not take STEM courses as part of their curriculum [6–8]. The commonalities of the disciplines in our study and the differences in the results between our study and previous research suggest that students might benefit from taking courses that involve the graphical representation of data associated with specific real-world concepts [21] as is often the case in STEM courses. Nevertheless, there are likely to be other factors that influence students' graph comprehension skills, such as personal interest.

This study replicated previous studies [7] distinguishing correct and incorrect solvers based on their eye movements using machine-learning methods. This suggests that machine learning has the potential to be successfully applied as a tool in STEM education research, especially in the context of eye movement analyses. A comparison with simple algorithms suggests that the method proposed by Steinert *et al.* [46] is appropriate for analyzing small datasets, such as those used in this study. These results suggest that eye-tracking measurements can be used as inputs for adaptive learning systems [9]. Applications that use eye-tracking metrics will likely become increasingly relevant for teachers, especially in conjunction with personalized systems that can support individual learners. For example, machine-learning applications could help teachers identify students encountering problems with the course content.

D. Limitations and future research

There are several limitations to this research. The most relevant limitation is the low number of participants. Another limitation is that the medical items used in the pretest and post-test had not been previously tested, which may affect the comparability with the isomorphic items associated with mathematics and physics. Furthermore, other factors that may influence graph comprehension skills —such as self-selection of the study discipline or the number of STEM courses that each student took—were not addressed in our research questions. Future research should extend this research into other potential variables that could influence the development of graph comprehension skills.

Future research should also validate the medical test items used in this study to assess their comparability with test instruments used in previous research. Comparisons with other STEM disciplines, such as mathematics or computer science students, could also provide further insights as to why students of some disciplines outperform others in graph comprehension tasks.

While SVMs appear to be well suited to making predictions about correct and incorrect solvers, this study indicated a relatively low MCC score; future research may wish to improve the model presented in this study by testing other types of machine-learning algorithms, such as neural networks. The model performance may also improve by adding more features.

Finally, future research should consider the real-world application of eye-tracking data in adaptive learning systems and their ability to support students during the learning process. These systems should ideally be accessible and capable of operating in real time. Indeed, SVMs can be used in real time [55,56] and web-based eye tracking can now be easily implemented on certain websites [10]. These tools would provide students with individual support while also providing teachers with immediate information on whether students have issues with a specific type of problem or whether a particular student needs more assistance.

VI. CONCLUSION

Graph comprehension is an important skill for students of various disciplines. Physics majors appear to have better graph comprehension skills compared to students from social sciences disciplines. This study compared the graph comprehension skills of physics and medical students across three different subjects: math, physics, and medicine. Each participant attempted these tasks at the beginning and the end of the first semester. There were no significant differences in graph comprehension gain or visual behavior between physics majors and medical students, suggesting that students from these disciplines had similar graph comprehension skills. One reason for this finding could be that physics majors and medicine students often engage with graphs due to the multitude of STEM courses in their respective disciplines.

An investigation of the visual processing of correct and incorrect solvers across both disciplines using a novel machine-learning approach tailored to small datasets revealed that eye movements could be used as a predictor for performance. This demonstrates that machine-learning analyses can be a valuable tool for education research even for small datasets, as long as this is accounted for during the analysis.

V. R., Y. D., S. K., M. B., D. K., J. B., J. S., M. F., and J. K. conceptualized the work; V. R., Y. D., S. K., and J. K. determined the methodology; V. R. conducted the formal analysis; S. S. developed the machine-learning algorithm tailored to small datasets; V. R. and Y. D. conducted the investigation; V. R., Y. D., S. K., M. B., D. K., J. B., J. S., M. F., and J. K. validated the data; V. R. curated the data; V. R. prepared the original draft; V. R., Y. D., S. K., M. B., S. S., D. K., J. B., J. S., M. F., and J. K. reviewed and edited the manuscript; V. R. provided the visualizations; S. K. and J. K. supervised the project; J. K. dealt with the project administration.

[1] R. Mayer and L. Fiorella, *The Cambridge Handbook of Multimedia*, 3rd ed. (Cambridge University Press, Cambridge, England, 2012), 10.1017/9781108894333.

[2] Program for international student assessment (PISA). PISA 2022 mathematics framework, https://pisa2022-maths.oecd.org/ (2022).

[3] L. Boels, A. Bakker, W. Van Dooren, and P. Drijvers, Conceptual difficulties when interpreting histograms: A review, Educ. Res. Rev. **28**, 100291 (2019).

[4] N. Glazer, Challenges with graph interpretation: a review of the literature, Stud. Sci. Educ. **47**, 183 (2011).

[5] G. Leinhardt, O. Zaslavsky, and M. K. Stein, Functions, graphs, and graphing: Tasks, learning, and teaching, Rev. Educ. Res. **60**, 1 (1990).

[6] S. Brückner, O. Zlatkin-Troitschanskaia, S. Küchemann, P. Klein, and J. Kuhn, Changes in students' understanding of and visual attention on digitally represented graphs across two domains in higher education: A postreplication study, Front. Psychol. **11**, 1 (2020).

[7] P. Klein, S. Küchemann, S. Brückner, O. Zlatkin-Troitschanskaia, and J. Kuhn, Student understanding of graph slope and area under a curve: A replication study comparing first-year physics and economics students, Phys. Rev. Phys. Educ. Res. **15**, 020116 (2019).

[8] A. Susac, A. Bubic, E. Kazotti, M. Planinic, and M. Palmovic, Student understanding of graph slope and area under a graph: A comparison of physics and nonphysics students, Phys. Rev. Phys. Educ. Res. **14**, 020109 (2018).

[9] K. Kennel, Using eye tracking data for enhancing adaptive learning systems, in *Proceedings of the 2022 Symposium on Eye Tracking Research and Applications, ETRA 2022* (2022), Vol. 34, pp. 1–3, 10.1145/3517031.3532195.

[10] X. Yang and I. Krajbich, Webcam-based online eye-tracking for behavioral research, Judgm. Decis. Mak. **16**, 1485 (2021).

[11] M. Sailer, E. Bauer, R. Hofmann, J. Kiesewetter, J. Glas, I. Gurevych, and F. Fischer, Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning, Learn Instr. **83**, 101620 (2023).

[12] S. Küchemann, P. Klein, S. Becker, N. Kumari, and J. Kuhn, Classification of students' conceptual understanding in STEM education using their visual attention distributions: A comparison of three machine-learning approaches, in *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU)* edited by H. C. Lane, S. Zvacek, and J. Uhomoibhi (SciTePress, Portugal, 2020), Vol. 1, pp. 36–46, 10.5220/0009359400360046.

[13] L. Boels, E. G. Moreno-Esteva, A. Bakker, and P. Drijvers, Automated gaze-based identification of students' strategies in histogram tasks through an interpretable mathematical model and a machine learning algorithm, Int. J. Artif. Intell. Educ. **34**, 931 (2023).

[14] A. Susac, A. Bubic, P. Martinjak, and M. Palmovic, Graphical representations of data improve student understanding of measurement and uncertainty: An eye-tracking study, Phys. Rev. Phys. Educ. Res. **13**, 020125 (2017).

[15] P. Shah and J. Hoeffner, Review of graph comprehension research: Implications for instruction, Educ. Psychol. Rev. **14**, 47 (2002).

[16] S. Friel, F. R. Curcio, and G. W. Bright, Making sense of graphs: Critical factors influencing comprehension and instructional implications, J. Res. Math. Educ. **32**, 124 (2001).

[17] E. G. Freedman and P. Shah, Toward a model of knowledge-based graph comprehension, in *Diagrammatic Representation and Inference. Diagrams 2002*, edited by M. Hegarty, B. Meyer, and N. H. Narayanan, *Lecture Notes in Computer Science Vol. 2317* (Springer, Berlin, Heidelberg, 2002), pp. 18–30, 10.1007/3-540-46037-3_3.

[18] E. Harrison, B. McBride, and M. C. Linn, Revision analysis of students' position-time graphs, CSCL **2** (2019), https://par.nsf.gov/biblio/10181648-revision-analysis-students-position-time-graphs.

[19] C. F. J. Pols, P. J. J. M. Dekkers, and M. J. de Vries, What do they know? Investigating students' ability to analyse experimental data in secondary physics education, Int. J. Sci. Educ. **43**, 274 (2021).

[20] Y. Okan, M. Galesic, and R. Garcia-Retarmero, How people with low and high graph literacy process health graphs: Evidence from eye-tracking, J. Behav. Decis. Making **29**, 271 (2016).

[21] J. Woolnough, How do students learn to apply their mathematical knowledge to interpret graphs in physics? Res. Sci. Educ. **30**, 259 (2000).

[22] M. Rau, Conditions for the effectiveness of multiple visual representations in enhancing STEM learning, Educ. Psychol. Rev. **29**, 717 (2017).

[23] A. A. diSessa and B. L. Sherin, Meta-representation: an introduction, J. Math. Behav. **19**, 385 (2000).

[24] A. Angra and S. M. Gardner, Development of a framework for graph choice and construction, Adv. Physiol. Educ. **40**, 123 (2016).

[25] D. Carraher and A. Schliemann, The transfer dilemma, J. Learn. Sci. **11**, 1 (2002).

[26] E. Alemdag and K. Cagiltay, A systematic review of eye tracking research on multimedia learning, Comput. Educ. **125**, 413 (2018).

[27] M. L. Lai, M. J. Tsai, F. Y. Yang, C. Y. Hsu, T. C. Liu, S. W. Y. Lee, M. H. Lee, G. L. Chiou, J. C. Liang, and C. C. Tsai, A review of using eye-tracking technology in exploring learning from 2000 to 2012, Educ. Res. Rev. **10**, 90 (2013).

[28] M. A. Just and P. A. Carpenter, A theory of reading: From eye fixations to comprehension, Psychol. Rev. **87**, 329 (1980).

[29] L. Hahn and P. Klein, Eye tracking in physics education research: A systematic literature review, Phys. Rev. Phys. Educ. Res. **18**, 013102 (2022).

[30] V. Ruf, A. Horrer, M. Berndt, S. I. Hofer, F. Fischer, M. R. Fischer, J. M. Zottmann, J. Kuhn, and S. Küchemann, A literature review comparing experts' and non-experts' visual processing of graphs during problem-solving and learning, Educ. Sci. **13**, 1 (2023).

[31] K. A. Ericsson and W. Kintsch, Long-term working memory, Psychol. Rev. **102**, 211 (1995).

[32] H. L. Kundel, C. F. Nodine, E. F. Conant, and S. P. Weinstein, Holistic component of image perception in mammogram interpretation: Gaze-tracking study, Radiology **242**, 396 (2007).

[33] H. Haider and P. A. Frensch, Eye movement during skill acquisition: More evidence for the information-reduction hypothesis, J. Exp. Psychol. **25**, 172 (1999).

[34] A. Maries and C. Singh, Helping students become proficient problem solvers part I: A brief review, Educ. Sci. **13**, 156 (2023).

[35] A. Gegenfurtner, E. Lehtinen, and R. Säljö, Expertise differences in the comprehension of visualizations: a Meta-analysis of eye-tracking research in professional domains, Educ. Psychol. Rev. **23**, 523 (2011).

[36] S. Brams, G. Ziv, O. Levin, J. Spitz, J. Wagemans, A. M. Williams, and W. F. Helsen, The relationship between gaze behavior, expertise, and performance: A systematic review, Psychol. Bull. **145**, 980 (2019).

[37] J. A. Harsh, M. Campillo, C. Murray, C. Myers, J. Nguyen, and A. v. Maltese, "Seeing" data like an expert: An eye-tracking study using graphical data representations, CBE Life Sci. Educ. **18**, ar32 (2019).

[38] S. Becker, L. Knippertz, S. Ruzika, and J. Kuhn, Persistence, context, and visual strategy of graph understanding: Gaze patterns reveal student difficulties in interpreting graphs, Phys. Rev. Phys. Educ. Res. **19**, 020142 (2023).

[39] S. Ceuppens, L. Bollen, J. Deprez, W. Dehaene, and M. De Cock, 9th grade students' understanding and strategies when solving $x(t)$ problems in 1D kinematics and $y(x)$ problems in mathematics, Phys. Rev. Phys. Educ. Res. **15**, 010101 (2019).

[40] S. Becker, S. Küchemann, P. Klein, A. Lichtenberger, and J. Kuhn, Gaze patterns enhance response prediction: More than correct or incorrect, Phys. Rev. Phys. Educ. Res. **18**, 020107 (2022).

[41] A. Daoud and D. Dubhashi, Statistical modeling: the three cultures, arXiv.2012.04570.

[42] A. Namoun and A. Alshanqiti, Predicting student performance using data mining and learning analytics techniques: A systematic literature review, Appl. Sci. **11**, 237 (2020).

[43] T. Kabudi, I. Pappas, and D. H. Olsen, AI-enabled adaptive learning systems: A systematic mapping of the literature, Comput. Educ. **2**, 100017 (2021).

[44] L. Breiman, Statistical modeling: The two cultures, Stat. Sci. **16**, 199 (2001).

[45] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. (O'Reilly Media Inc., Canada, 2019).

[46] S. Steinert, V. Ruf, D. Dzsotjan, N. Großmann, A. Schmidt, J. Kuhn, and S. Küchemann, A refined approach for evaluating small datasets via binary classification using machine learning, PLoS One **19**, e0301276 (2024).

[47] A. Field, J. Miles, and Z. Field, *Discovering Statistics Using R* (SAGE Publications, 2012).

[48] W. Xu and F. Ouyang, The application of AI technologies in STEM education: A systematic review from 2011 to 2021, Int. J. STEM Educ. **9**, 59 (2022).

[49] D. Dzsotjan, K. Ludwig-Petsch, S. Mukhametov, S. Ishimaru, S. Küchemann, and J. Kuhn, The predictive power of eye-tracking data in an interactive AR learning environment, in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers* (ACM, New York, 2021), pp. 467–471, 10.1145/3460418.3479358.

[50] S. Küchemann, S. Becker, P. Klein, and J. Kuhn, Gaze-based prediction of students' understanding of physics line-graphs: An eye-tracking-data based machine-learning approach, in *Computer Supported Education*, edited by H. C. Lane, S. Zvacek, and J. Uhomoibhi (Springer International Publishing, 2021), pp. 450–467, 10.1007/978-3-030-86439-2_23.

[51] R. J. Beichner, Testing student interpretation of kinematics graphs, Am. J. Phys. **62**, 750 (1994).

[52] S. Mozaffari *et al.*, Quantifying gaze-nased strategic patterns in physics vector field divergence, in *ICAART 2020. Lecture Notes in Computer Science*, edited by A. P. Rocha, L. Steels, and J. van den Herik (Springer, 2021), pp. 465–481, 10.1007/978-3-030-71158-0_22.

[53] N. S. Rebello *et al.*, Machine learning predicts responses to conceptual tasks using eye movements, presented at PER Conf. (2018), https://www3.cs.stonybrook.edu/~minhhoai/papers/PERC2018_Rebello.pdf.

[54] V. Bachurina *et al.*, A machine learning investigation of factors that contribute to predicting cognitive performance: Difficulty level, reaction time and eye-movements, Decis. Support Syst. **155**, 113713 (2022).

[55] D. Ifenthaler and C. Widanapathirana, Development and validation of a learning analytics framework: Two case studies using support vector machines, Tech. Know. Learn. **19**, 221 (2014).

[56] C. Verma, V. Stoffová, and Z. Illés, Prediction of students' awareness level towards ICT and mobile technology in Indian and Hungarian University for the real-time: preliminary results, Heliyon **5**, e01806 (2019).

[57] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.20.020138 for the original test items in German.

[58] Tobii. Tobii Pro Nano, https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-nano#specifications (2022).

[59] D. D. Salvucci and J. H. Goldberg, Identifying fixations and saccades in eye-tracking protocols, in *Proceedings of the Symposium on Eye Tracking Research & Applications —ETRA '00*, (ACM, New York, 2000), pp. 71–78, 10.1145/355017.355028.

[60] K. Holmqvist and R. Andersson, *Eye Tracking: A Comprehensive Guide to Methods, Paradigms and Measures* (Lund Eye–Tracking Research Institute, Lund, Sweden, 2017).

[61] W. Jia, M. Sun, J. Lian, and S. Hou, Feature dimensionality reduction: A review, Complex Intell. Syst. **8**, 2663 (2022).

[62] R. E. Kirk, Practical significance: A concept whose time has come, Educ Psychol Meas **56**, 746 (1996).

[63] B. E. Woolnough, Why students choose physics, or reject it, Phys. Educ. **29**, 368 (1994).

5. General Discussion

This thesis aimed to (a) analyse graphing as a tool in learning and problem-solving, (b) investigate expertise in learning and problem-solving with graphs, and (c) explore differences between study disciplines during problem-solving with graphs. The theoretical and empirical background of this research was discussed in section 1. All studies fall under the broader scope of exploring the two aspects of graphing competence (see section 1.2.3): the aspect of creating graphs by reviewing graphing and the aspect of graph comprehension by investigating expertise differences in visual behaviour and comparing the problem-solving skills of students of two disciplines. The first study presented a synthesis of empirical studies on graphing in K-12 education as a systematic literature review (see section 2). The second study focused on the second aim by investigating expertise differences in visual processing during learning and problem-solving with graphs (see section 3). The last study addressed graph comprehension by comparing the performance and the visual behaviour of physics and non-physics students solving graph tasks (see section 4).

The results of all three studies are summarised in section 5.1. Then, the theoretical (see section 5.2) and practical (see section 5.3) implications are discussed. Limitations and future research are described afterwards (see sections 5.45.4 and 5.5, respectively).

5.1. Summary of the Results of the Articles

5.1.1. Results of Study 1

The first study (see section 2) presented a systematic review of the empirical research on graphing statistical data in STEM education. This literature review identified 44 peer-reviewed studies published between 1979 and 2021 investigating this topic. From these studies, information about the population (e.g., high-school students), the STEM discipline (e.g., physics), the graphing method (e.g., manual) and guidance (e.g., explicit), the type of graph (e.g., line graphs) and the types of data (e.g., bivariate), the study design (e.g., problem-solving), the results (e.g., positive), and the students' difficulties (e.g., scaling the axes) were extracted. All extracted codes are available under https://osf.io/4wtac/?view_only=137943ec30ee47fd98950aef2cef43a0. The studies included in the review were analysed regarding the implementation of graphing in STEM education research (RQ 1), the added value of graphing in STEM education (RQ 2), and students' difficulties during graphing (RQ 3).

Results indicated that most studies were conducted during more than one lesson and that graphing was often analysed in the context of problem-solving (RQ 1 of study 1, see section 1.4.1). There seemed to be no preference for either manual or tool-based graphing. The most common types of graphs were line graphs. A typical study design included instruction over multiple lessons to investigate the effectiveness of such instruction on graphing skills for line graphs. Several studies reported positive effects of various kinds of graphing instruction that indicated the benefit of graphing. Additionally, instructing graphing was beneficial for graph comprehension (RQ 2 of study 1, see section 1.4.1). The included studies reported various student difficulties (RQ 3 of study 1, see section 1.4.1). Student difficulties with graph construction can be sorted into three categories: (1) difficulties during graph construction, (2) difficulties during variable ordering, or (3) difficulties with translating data between types of representations. Theoretical difficulties during graphing, for example, with interpretation or connecting the data to the underlying concept, were also reported frequently. However, difficulties during graph construction were the most common. Many studies reported both conventional and theoretical student difficulties. The prevalence of both types of difficulties in multiple studies suggests a possible connection between them.

5.1.2. Results of Study 2

The second study (see section 3) reviewed literature comparing the visual behaviour of experts and non-experts during learning and problem-solving with graphs. Thirty-two articles exploring this topic were included in the review. From these studies, the STEM discipline (e.g., mathematics), the type of graph (e.g., line graphs), the type of eye-tracking metric (e.g., fixations), and key findings were extracted. The findings of the included studies were analysed regarding the types of eye-tracking metrics used (research aim 1) and differences in visual behaviour between experts and non-experts were synthesised (research aim 2). Eye-tracking metrics were distinguished not only by their types but also based on the size of the areas of interest (AOIs) used to calculate them. The outcomes reported in the studies were investigated for micro-level as well as meso- and macro-level AOIs (Andrá et al., 2015). An example of these AOIs can be seen in Figure 1 (see section 1.4.2). Furthermore, the method of expertise determination in the included studies was extracted and analysed.

Dwell time, fixation duration, and fixation count were typical eye-tracking metrics, independent of AOI size (research aim 1 of Study 2, see section 1.4.2). The results indicated that experts paid more attention to relevant parts of a graph than non-experts (research aim 2 of study 2, see section 1.4.2). This is in line with the information-reduction hypothesis (Haider & Frensch, 1999). Experts also seemed to make more gaze switches between the areas of a graph. Participants who looked at the relevant variables early on and afterwards identified them in the graphs had the most efficient strategy to extract such data.

5.1.3. Results of Study 3

The third study (see section 4) investigated the effect of context and discipline on learning gain and visual behaviour during graph comprehension tasks. Differences in learning gain (RQ 1) and visual behaviour (RQ 2) were compared between participants of varying disciplines. Twelve medical/veterinary and twelve physics students participated in the data collection at both the beginning and the end of the semester. The participants answered the same 24 isomorphic questions in the context of mathematics, physics, and medicine at both times.

There were no differences in learning gain between physics and non-physics first-semester students (RQ 1 of Study 3, see section 1.4.3). However, there was a statistically significant difference between contexts: participants improved more between the pretest and posttest in items in the context of physics compared to items in a medical context. Visual behaviour between correct and incorrect solvers differed in both the pretest and the posttest as indicated by a significant *p*-value (RQ 2 of study 3, see section 1.4.3). In a comparison of various machine-learning algorithms and an examination of their performance metrics, an SVM optimised for small datasets (Steinert et al., 2024) seemed to be best suited for analysing eye-tracking results. This replicated previous results calculated via statistical methods (Klein, Küchemann, et al., 2019) using machine-learning techniques. In line with the information-reduction hypothesis (Haider & Frensch, 1999), Klein, Küchemann, et al. (2019) found that correct solvers paid more attention than incorrect solvers to relevant areas. An investigation of the feature importance of the machine-learning algorithm identified that the dwell time on the axes and the graph were the most important AOIs for predicting the participants' performance. Additional statistical analyses using multiple linear regression suggested that the area of a graph that participants paid attention to was related to the total dwell time on a stimulus. In contrast, looking at the axes labels seemed to indicate less time spent on the task.

5.2. Theoretical Implications

Graphical representations are crucial educational tools (see section 1.2). Being able to use numerical information and interpret data are important skills for students as well as communicating with data (Program for International Student Assessment, 2022). Therefore, students should learn how to comprehend graphs as well as how to create them (Glazer, 2011). This is called *graphing competence*. A high graphing competence enables students to interpret data correctly as well as to efficiently convey information in the form of graphs. The focus of the three studies of this thesis is on graphing (Study 1) and graph comprehension (Studies 2 and 3). Based on the results of these studies (see section 5.1), theoretical implications in the context of graphing and graph comprehension can be drawn.

5.2.1. Student Difficulties With Graphing

One of the aspects of graphing competence is creating graphs (Glazer, 2011). Graphing is a constructive learning activity (Chi & Wylie, 2014). Constructing graphical representations has multiple benefits for learners (see section 1.2.2): It is a generative activity during which learners can visualise and externalise information (Schmidgall et al., 2019). During generation activities, learners have the chance for detailed self-explanations (Fiorella & Kuhlmann, 2020) or to represent information only implicitly mentioned in the text (Scheiter et al., 2017). However, generation should be taught carefully (Fiorella & Zhang, 2018; Scheiter et al., 2017) because learners need enough cognitive resources to generate representations (Schwamborn et al., 2011). As a generative activity, these advantages should also apply to the generation of graphs. The findings of the first study indicated that graphing instruction can improve not only learners' graphing skills but also their graph comprehension. Students seem to pay close attention to a graph’s details during graphing (Gerard et al., 2012), which could lead to an improvement in their graph comprehension. An examination of the data (points) making up a graph could facilitate graph interpretation on a local level (Leinhardt et al., 1990). This is supported by the assumption that visualising can help organise knowledge (Fiorella, 2023). Furthermore, the findings of the review on graphing indicate that graphing instruction could also facilitate scientific skills, such as generating hypotheses (Gultepe & Kilic, 2015). These results are in line with the benefits of construction activities described in the ICAP framework (Chi & Wylie, 2014).

Study 1 also synthesised results regarding students' difficulties during graphing (RQ 3). This has been a focus of research on graphing competence (see section 1.2.3), as both graphing and graph comprehension are not easy for students (Glazer, 2011; Leinhardt et al., 1990). For example, Leinhardt et al. (1990) investigated students' misconceptions about graphs. These misconceptions can be based on misunderstanding previous instructions (Leinhardt et al., 1990). Using the wrong strategies can cause students to misinterpret the data, such as interpreting a graph as a picture of a situation (Clement, 1985). Difficulties have also been found during graphing (e.g., Wavering, 1985). Therefore, good instruction regarding graphing competence is crucial for dealing with students' difficulties. Although difficulties during graph comprehension have been reviewed previously (Boels et al., 2019; Clement, 1985; Glazer, 2011; Leinhardt et al., 1990), so far, there have been no extensive reviews regarding students' difficulties specifically during graphing.

The difficulties during graphing reported in the studies included in the systematic review (see section 2) varied. For example, students have trouble with scaling the axes (Åberg-Bengtsson, 2006; von Kotzebue et al., 2015) or graphing the data points, for example, because they forget to consider possible deviations (Dewi et al., 2018). These difficulties can be considered in the framework of the graph construction process (Lachmayer et al., 2007): problems during scaling are construction difficulties because they concern the structure of the graph, and difficulties with data points are variable ordering difficulties due to the affiliation with charting the points. Another difficulty is translating data from one type of representation to another, such as from a table to a graph (Oslington et al., 2020). Difficulties during the construction process were the most frequently reported type of student difficulties in Study 1. Students can also have difficulties relating to theoretical aspects of graphing: For example, connecting the data to the underlying concept, as is the case for graph-as-picture errors (Clement, 1985; Gerard et al., 2012). Other theoretical difficulties include interpreting the data, such as misinterpreting the relationship between the variables indicated by the x-axis and y-axis (Dewi et al., 2018) and finding the correct type of graph for plotting the data (Ozmen et al., 2020). Such theoretical difficulties should be addressed by teachers in order to solve them (Boels et al., 2019).

Eighteen of the studies included in the literature review of graphing reported that students had both conventional and theoretical difficulties. This indicated a possible relation between these two types of difficulties. For example, one study, that also investigated the conceptual understanding of students, found that high conceptual understanding was related to high graphing skills (Gültepe, 2016). Students' graphing difficulties might therefore not only be due to not understanding graphing conventions but also due to not being able to to comprehend the data. In turn, this might influence how students choose to display data during graphing. As both aspects, graphing and graph comprehension, are part of graphing competence (Glazer, 2011), it seems reasonable to assume that both conventional and theoretical difficulties play a role in graphing. On the other hand, students who can correctly graph data might have a better grasp on its interpretation as well. How graphing skill and graph comprehension influence each other and whether this relationship is directional, has not yet been investigated.

5.2.2. Visual Behaviour During Graph Comprehension

One method to investigate learning and problem-solving processes during graph comprehension is eye tracking (see section 1.3.1). Eye tracking is a method for investigating visual behaviour. This can be useful, e.g., to investigate expertise differences (see section 1.3.2). Experts process information more efficiently than non-experts (Ericsson & Kintsch, 1995). There are three main theories about expertise that lead to distinct assumptions regarding the visual processing of experts: (1) The information-reduction hypothesis assumes that experts focus more on relevant information by ignoring irrelevant information on a perceptual level (Haider & Frensch, 1999). Based on this hypothesis, experts should focus more and longer on relevant information (Gegenfurtner et al., 2011). (2) The holistic model of image perception states that experts process images globally (Kundel et al., 2007) and therefore fixate on relevant information more quickly than non-experts (Gegenfurtner et al., 2011). (3) According to the assumption of long-term working memory, experts store information more efficiently than non-experts (Ericsson & Kintsch, 1995; Guida et al., 2012). Consequently, experts should spend less time on relevant information (Gegenfurtner et al., 2011). Previous research found support for all theories (Brams et al., 2019; Gegenfurtner et al., 2011; Sheridan & Reingold, 2017). However, the holistic model of image perception seems to be most prevalent in the medical field (Brams et al., 2019; Sheridan & Reingold, 2017), although para-foveal processing necessary for global image is also one of the processes mentioned in the CTVE (Gegenfurtner et al., 2023).

This thesis analyses expertise in the context of graph comprehension by investigating the visual processing of experts and non-experts during learning and problem-solving with graphs (research aim b). The second study presents a systematic review of studies comparing the visual behaviour of participants with various expertise levels during learning and problem-solving with graphs (see section 3). It should be noted, that, unlike previous reviews of graph interpretation (Boels et al., 2019), this study does not focus on a specific type of graph, but includes all types of graphs. When investigating visual processing based on eye movements, the size of the AOIs should be considered because eye-tracking metrics are calculated for AOIs (Holmqvist & Andersson, 2017). Therefore, the studies' findings were analysed based on the size of the AOIs used by the studies' authors. The review analysed two categories of AOI sizes: AOIs based on larger areas, such as the entire graph or large parts of it (macro- and meso-level AOIs), as well as AOIs separating very small areas, such as individual ticks on the axes (micro-level AOIs).

Overall, the findings at the macro- and meso-level were similar to those at the micro-level. At all levels, experts seem to fixate longer on relevant information. This result is in line with the information-reduction hypothesis (Gegenfurtner et al., 2011; Haider & Frensch, 1999). Therefore, experts seem to be able to ignore the irrelevant information in graph comprehension tasks to better focus on the relevant information. The analysis of macro- and meso-level AOIs also indicated that experts made more dynamic eye movements, such as revisits and saccades, related to integrating information. Building connections is also an indicator of expertise (Gegenfurtner et al., 2023). This process is, for example, important during the organisation of image chunks (Gegenfurtner et al., 2023). Consequently, eye movements seem to be valid indicators for expertise determination in the context of graph comprehension.

Expertise, as indicated by solution corrects, has been extensively researched in various studies. For example, physics students outperformed both psychology (Susac et al., 2018) and economics (Brückner et al., 2020; Klein, Küchemann, et al., 2019) students in graph comprehension tasks in both their familiar (physics) and unfamiliar (finance) disciplines. Klein et al. (2019) could distinguish participants based on their performance in solving a task correctly and incorrectly via their dwell time on concept-specific AOIs. Other studies also used statistical methods to predict performance based on dwell time in relevant areas (Becker et al., 2022) or average fixation duration (Chen et al., 2014). Besides statistical methods, machine-learning algorithms can be used to predict performance (see section 1.3.3). Supervised machine-learning methods are common in educational research (Namoun & Alshanqiti, 2021). Supervised methods use labelled training data to train an algorithm to label unknown test data (Géron, 2019), such as predicting a correct or incorrect solution based on eye movements. For example, Küchemann et al. (2021) could predict high-school students' performance on graph comprehension tasks based on their transition frequency and their dwell time on the graph using an SVM. Dzsotjan et al. (2021) also used an SVM to predict participants' learning gain based on their eye movements for participants walking the shape of a position-time graph. These results indicate that machine-learning methods are suitable methods to analyse eye movements. Unlike statistical models, machine-learning methods can, for example, be used to analyse complex relationships between the predicted variable and its predictors which can be transferred to unknown datasets due to the split between training and test data. Furthermore, feature relevance can indicate how important input parameters were for the alogrithms' performance which can add a level of interpretability to machine-learning algorithms.

The third study included in this thesis investigated differences between physics and non-physics students during problem-solving with graphs (research aim c). The setup was analogous to previous studies (Brückner et al., 2020; Klein, Küchemann, et al., 2019; Susac et al., 2018). In this study, physics and non-physics students were asked to solve graph tasks at the beginning and end of their first semester. Unlike previous research on performance differences (Brückner et al., 2020; Klein, Küchemann, et al., 2019; Susac et al., 2018), there were no differences in learning gain between students of different disciplines. However, non-physics students were not economics or psychology students but medical and veterinary students who also took STEM courses. These courses might teach problem-solving routines which could be transferable to other problem-solving tasks (see section 4). However, there were differences in visual behaviour between correct and incorrect solvers that could be predicted using machine-learning methods. This supports the assumption that visual processing differs between correct and incorrect solvers, indicating that expertise theories (see section 1.3.2) can be applied to comparing the visual behaviour of correct and incorrect solvers during graph comprehension tasks. Previous results by Klein, Küchemann et al. (2019) supporting the information-reduction hypothesis (Haider & Frensch, 1999) could be replicated using machine-learning methods. This is in line with the second study of this thesis, which also found support for the information-reduction hypothesis during the visual processing of graph tasks. Correct solvers seem to pay more attention to task-relevant areas (Klein, Küchemann, et al., 2019). In the context of machine learning, the participants' dwell time on the graph seems to be a particularly important predictor of their performance. The algorithm used was an SVM optimised for small datasets that included repeated-nested cross-validation in combination with permutation tests (Steinert et al., 2024). The optimised methods performed better on the data compared to regular applications of an SVM without cross-validation or permutation tests. This indicates that such an optimised method is a suitable tool for investigating eye movements, even in the context of small datasets. Such an analysis could be useful for future research, for example, for developing adaptive learning systems based on eye movements that could give students real-time feedback about their performance.

5.3. Implications for Practice

The studies presented in this thesis reported empirical results, interpretations, and theoretical implications that have implications for teaching practice. All studies consider graphing competence. Practical implications for teaching concern dealing with students' difficulties during graphing as well as increasing students' level of expertise during graph comprehension.

Graphing instruction has previously been recommended (Glazer, 2011). Based on the literature review presented in Study 1, we found that various types of instruction were beneficial for students, such as instruction to improve graphing skills, for example, during lab-based activities (Gerard et al., 2012), as well as instruction to improve scientific argumentation (Gultepe & Kilic, 2015). Although most studies specifically focused on improving the students' graphing skills, not only those skills benefitted from instruction. For example, graphing instruction also had a positive effect on graph comprehension (Gerard et al., 2012). These two skills are both considered in graphing competence (Glazer, 2011). Due to the relevance of graphing competence in education (see section 1.2.3), graphing instruction seems a valuable aspect of education.

Teachers should consider students' difficulties during graphing instruction. This is especially relevant for STEM education because graphing plays an important role there (Leinhardt et al., 1990). The findings of the first study of this thesis indicated that students' difficulties with graphing conventions and their difficulties with theoretical aspects might be related. Therefore, both aspects should be considered in graphing instruction. Students benefit from exploring their scientific ideas before conventional instruction on a topic as well as from a deep analysis of provided data (Vitale et al., 2019). A combination of those activities might address both theoretical difficulties, such as graph-as-picture errors (Clement, 1985), by explicitly demonstrating which aspects of a concept students might have difficulties understanding, and students' difficulties with graphing conventions, such as scaling (Lachmayer et al., 2007), as a deep analysis of the data might give students the chance to revise their graphs.

Students' level of expertise regarding graphing competence can be facilitated in other ways as well. As eye movements can be indicators of expertise for graph tasks (see section 3), instruction based on experts' eye movements might also be beneficial for students. For example, eye-movement modelling examples can help students find relevant areas more quickly (Xie et al., 2021) by drawing their attention to them (Tunga & Cagiltay, 2023). This is in line with the information--reduction hypothesis (Haider & Frensch, 1999), the expertise theory supported by the second study of this thesis. Apart from learning the visual strategy, students should know how to implement it (Harsh et al., 2019). They should not merely learn the teacher's interpretation instead of learning the process of coming to the correct conclusion themselves (Bowen et al., 1999). Teachers should instruct students in the correct strategies for interpreting a graph while being careful that students learn how to determine the correct solution themselves. Eye-movement modelling examples could be a good way to teach students the correct strategy without necessarily giving students the exact answer.

Students could also be supported using machine learning. Eye movements can be useful indicators of expertise (see section 3) and performance (see section 4) in graph comprehension tasks. Furthermore, eye movements can be used to improve adaptive learning environments (Kennel, 2022). A good performance prediction based on students' eye movements, ideally in real-time, could be used to diagnose students' difficulties and offer personalised learning support.

5.4. Limitations

This thesis presented three studies, whose methodologies might lead to limitations regarding the theoretical and practical implications presented above.

First, the focus of all three studies was on graphs depicting bivariate data. This limits the overall generalizability of the results, as other types of representations, such as 3D representations might have other difficulties or necessitate divergent visual strategies. For example, constructing graphs in a 3D environment might necessitate using construction software and visual processing in a 3D environment, such as virtual reality, which might require different strategies for finding relevant information as the graph would be more complex. Study 1 presented a literature review of graphing in K-12 STEM education (see section 2). The inclusion criteria did not exclude multivariate data and graphs of multivariate data were considered in the analysis. For example, students were asked to graph objects of varying buoyancy (sink, float) in volume-mass graphs (Vitale et al., 2019). However, graphs depicting bivariate data were the most common type of graphs. The second study reviewed existing literature that analysed eye-tracking results. There were no 3D representations of graphs used as stimuli in the studies included in this review although they were not explicitly excluded. One reason for the lack of 3D representations might be the study prerequisites: Eye-tracking is often conducted stationary in front of a screen and metrics are calculated based on AOIs (see section 1.3.1). For analysing graphs of multivariate data, these AOIs would have to be adapted accordingly. As AOIs should encompass relevant information with enough space to resolve eye movements as recorded by the eye tracker (Holmqvist & Andersson, 2017), a detailed analysis of graphs depicting multivariate data might be difficult. Additionally, previously used test items, such as the items used in Study 3 (Ceuppens et al., 2019; Susac et al., 2018), are often designed with graphs based on bivariate data. The results of the reviews presented in this thesis nevertheless are robust across various types of graphs as no graph type was excluded in either review and neither review was limited to a specific type of graph. Additionally, the theoretical difficulties reported in this thesis are in line with the results of previous reviews of misconceptions (Clement, 1985) and interpretation of histograms (Boels et al., 2019). This is also the case for two of the three conventional difficulties of graph construction, that are based on a structural model describing the graph construction process (Lachmayer et al., 2007). Moreover, all difficulties have been inductively determined based on the difficulties reported in the studies included in the review on graphing. The visual strategies of experts and non-experts identified in the second review were also reported across various types of graphs. The results confirmed the information-reduction hypothesis (Haider & Frensch, 1999). The results of this thesis therefore seem stable across various types of graphs. However, generalisability to, for example, 3D graphs depicting multivariate data cannot be guaranteed. This should be considered in future research.

Other limitations are based on the methodology of the specific studies reported in this thesis. This limitation concerns specifically Study 1. The study presented a review of graphing in K-12 STEM education. A systematic search identified 44 studies meeting the inclusion criteria. The underlying theories used in all included studies were analysed. However, this proved difficult as very few studies provided a theoretical basis for their study design and most justified their design based on the practical applications of graphing. Therefore, the study design varied considerably between the included studies. This made it difficult to evaluate the added value of graphing (RQ 2 of Study 1) because there were no consistent group comparisons and analysis methods varied. Although the benefits of graphing were found across various types of instructions, there might be distinct boundary conditions for different kinds of instruction that could not be investigated in this review: None of the studies included in the literature review presented in Study 1 analysed boundary conditions for effective graphing. For example, previous research on drawing found that this generative activity was more effective for older students (Brod, 2021; Y. Zhang et al., 2021) and for students with low prior knowledge (Lin et al., 2017). Similar boundary conditions might apply for graphing. For a stringent investigation, a comparison of pretest – posttest studies including effect sizes would be ideal because these could also be used for a meta-analysis. However, only one study included in this review reported an effect size (Adams & Shrum, 1990) and such an analysis was therefore not possible in Study 1.

A similar limitation applies to Study 2. This review compared the visual processing of experts and non-experts. However, the categorisation of expertise depended solely on descriptions of the authors of the studies included in this review, such as the comparison of correct and incorrect solvers (Klein, Küchemann, et al., 2019). Another limitation therefore concerns the definition of expertise: There does not seem to be a comprehensive, overarching definition of expertise. Sternberg (2003) described an expert student as a student who can intelligently use their knowledge and solve tasks creatively with the ability to successfully transfer learned information to practice. However, this definition only refers to students and many studies included professionals, such as university faculty (Harsh et al., 2019). Various groups of participants were therefore considered as experts and non-experts. However, four important factors when determining expertise in graph comprehension were identified based on previous literature: (1) graphical literacy (Shah & Hoeffner, 2002), (2) knowledge about the domain (Brückner et al., 2020), (3) prior mathematical knowledge (Curcio, 1987), and (4) task knowledge (Friel et al., 2001). However, these factors have not been empirically tested.

Further limitations concern the methodology of the third study presented in this thesis. This study aimed to explore differences between study disciplines during problem-solving with graphs (research aim c). To achieve this, physics and non-physics students were asked to solve previously employed graph tasks in the context of physics and math (Ceuppens et al., 2019; Susac et al., 2018) as well as in the context of medicine at the beginning and the end of their first semester. This approach was analogous to that of Brückner et al. (2020). Therefore, analysing the expertise of first-semester students who were the participants in Study 3 restricts the generalisability of the results. It would be very difficult for students to achieve expertise during their first semester in their respective disciplines. This could lead to difficulties interpreting the learning gain investigated in Study 3 because all participants had the same educational level at the beginning of the semester. However, differences in performance have been found in comparisons of first-semester students of different disciplines earlier (Klein, Küchemann, et al., 2019; Susac et al., 2018). Study 3 is therefore comparable to previous research regarding the choice of participants. For the analysis of the participants' visual behaviour, performance was chosen and based on this the visual behaviour of correct and incorrect solvers was analysed. This criterion has also been used previously (Klein, Küchemann, et al., 2019). However, a replication of Study 3 under consideration of the four factors that might be relevant for determining expertise in graph comprehension might provide further insights into the visual processing of experts during graph comprehension tasks.

An additional limitation also involves the methodology of Study 3, specifically the graph comprehension tasks used as test items and the number of participants. None of the task items of any context were validated test instruments. However, the physics and math items have been used in previous studies (Ceuppens et al., 2019; Susac et al., 2018). The medical items were developed for this study based on these examples. Although the test should be validated in future research, a comparability of the results is ensured by the tasks' use in previous studies, specifically in the mathematical and physics contexts (Brückner et al., 2020; Klein, Küchemann, et al., 2019; Susac et al., 2018).

5.5. Directions for Future Research

There are various directions for future research on graphing competence, both for graphing and graph comprehension. Research aspects include the definition of expertise, for either both graphing and graph comprehension or individually for the two aspects of graphing competence. Such a definition of expertise might influence how graphing competence is instructed and investigated as it would make study results more comparable. Graphing competence might also vary depending on the visualisation of the graph, for example, whether a graph is depicted in 2D or 3D. Visual behaviour might also differ between these types of visualisation. Furthermore, an analysis of participants' eye movements during graphing could provide valuable insight into the graphing process. In the following section, each of these aspects is elaborated.

One of the most relevant aspects for future research is the definition of expertise. There are four important aspects of graph comprehension: graphical literacy (Shah & Hoeffner, 2002) and knowledge about (2) the domain (Brückner et al., 2020), (3) the underlying math (Curcio, 1987), and (4) the task (Friel et al., 2001). However, these factors have not yet been systematically investigated. The impact of these factors on graph comprehension skills could vary, for example, knowledge about the domain of the graph comprehension task might be more important than knowledge about the underlying math. These could also vary depending on the kind of graph comprehension task, for example, there might be differences in determining the relation between variables and for extracting a value. Possible correlations between factors could be relevant as well. Furthermore, these factors could vary based on the type of the graph, for example, knowledge about the underlying math might be less relevant for comprehending histograms but very relevant for comprehending complex graphs with multiple variables. Additionally, the transferability of the four factors from graph comprehension to graphing should be investigated. Although all factors seem relevant in the context of graphing competence and theoretical and conventional difficulties seem to be connected, there could be variations between graphing and graph comprehension, for example, they might have differing boundary conditions. It would also be interesting to analyse how expertise in graphing competence develops. Starting with the four proposed factors, relevant aspects for future research could be the knowledge about the underlying math, the task, and the domain and how they should be taught, for example, whether they should be instructed in the same step or whether some of these factors might build on one another. Again, these factors could be investigated separately for graphing and graph comprehension as there might be differences in the development of these two skills.

These directions for future research highlight another open question: How are graphing and graph comprehension related and how should they be taught? They both should be instructed (Glazer, 2011) because they are both relevant in education (Leinhardt et al., 1990). Based on the research presented in this thesis, it seems that a certain level of graph comprehension is necessary for successful graphing; however, graphing instruction also seems to facilitate graph comprehension. This indicates that it might be beneficial for students to instruct these two skills of graphing competence together.

These findings were based on the review of graphing in K-12 STEM education. However, graphing competence is also relevant in university education (Nixon et al., 2016) and in professional praxis, such as engineering (Ahmed et al., 2021). For example, Nixon et al. (2016) investigated university students constructing best-fit lines in a physics lab course. In this study, university students had similar difficulties as K-12 students, such as connecting the data to the underlying concept. A systematic investigation of graphing in university education could provide further insights into how graphing skills develop after receiving formal instruction in school. Additionally, there is research where participants create graphs with their body movements. For example, Dzsotjan et al. (2021) developed a virtual reality environment where participants are supposed to recreate a position-time graph with their body motion. This study was based on embodiment because a “user physically experiences the mapping between real-world movement and kinematic graphs” (Dzsotjan et al., 2021, p. 468). Enacting movements like this can also be considered a generative learning activity (Fiorella, 2023).

Using the body is a new way of graphing that should be considered in future research; for example, it could be useful to teach students how data is connected to concepts via direct experience.

Moreover, the reviews presented the first two studies of this thesis concentrate on graphs in 2D. There was no preference for either manual or tool-based graphing in the studies included in the review. Furthermore, the data used for graphing was mostly bivariate. 3D graphs are probably more common in computer-based learning environments working with more than two variables. Additionally, there might be differences during the graphing process between 2D and 3D graphs. The review of literature analysing differences in visual behaviour between experts and non-experts during learning and problem-solving with graphs also focused on 2D representations. Eye movements are analysed for specific regions of a stimulus, which could make it complicated to design 3D representations with fitting relevant areas that can be constructively interpreted for analysing the visual strategy of participants. Furthermore, 3D graphs are not common in educational contexts included in the review, which might make it difficult to find participants with sufficient expertise.

Finally, this thesis can be extended via research investigating the visual strategies of students during the graphing process. Such research could provide insights into the strategies students use to create a graph. It would also be interesting to compare students' graphing strategies with those of experts. Knowing students' graphing strategies could be useful for improving graphing instruction. For example, one could give students support before they make mistakes due to misconceptions. An idea would be to show students help statements in an artificial reality environment which is also capable of recording and analysing eye movements in real time.

6. Conclusion

Dealing with data is an important aspect of everyday life as well as education, especially in a STEM context. Data is often presented in the form of graphs. To proficiently utilise graphs, one needs graphing competence; this describes the ability to comprehend and create graphs (Glazer, 2011). This thesis therefore analyses these aspects of graphing competence in the context of education.

Graphing is a constructive activity and requires learners to actively engage in the activity (Chi & Wylie, 2014). Graphing has multiple benefits and is frequently researched, especially in the context of STEM education. The findings of the systematic review on graphing presented in this thesis indicate, that graphing instruction is often investigated throughout more than one lesson and conducted with line graphs. The review's findings indicate that instruction can facilitate graphing skills. Graphing instruction can also improve graph comprehension. However, students frequently seem to have trouble constructing graphs. Students' difficulties can be categorised into two types: The first type of difficulties can be based on graphing conventions, such as constructing a graph. Additionally, students can have theoretical difficulties during graphing, such as interpreting the data. Conventional and theoretical difficulties are often jointly reported. This indicates that both types of difficulties influence students' graphing and instruction should therefore include both graphing and graph comprehension.

Eye movements can be an indicator of expertise during learning and problem-solving with graphs. During these activities, the findings of the second study indicate that experts pay more attention to relevant areas than non-experts. This supports the information-reduction hypothesis (Haider & Frensch, 1999). Additionally, the results indicate that eye movements based on fixations, such as fixation duration, are suited to investigate expertise in graph comprehension. Physics students are assumed to have a high level of expertise in graph comprehension as they performed better in graph comprehension tasks than students of other disciplines (Brückner et al., 2020; Klein et al., 2019; Susac et al., 2018). Correct and incorrect solvers can be distinguished based on their eye movements (Becker et al., 2022; Chen et al., 2014; Klein, Küchemann, et al., 2019). These results could be replicated with a machine-learning method optimised for small datasets (Steinert et al., 2024) in a study comparing physics and medical students' eye movements in graph comprehension tasks. The results suggest that machine-learning algorithms for small datasets are well suited for assessing expertise by analysing eye-tracking data. This could be useful for future research, for example, in developing adaptive learning systems.

References

Åberg-Bengtsson, L. (2006). “Then you can take half … almost”—Elementary students learning bar graphs and pie charts in a computer-based context. *The Journal of Mathematical Behavior*, *25*(2), 116–135. https://doi.org/10.1016/j.jmathb.2006.02.007

Adams, D. D., & Shrum, J. W. (1990). The effects of microcomputer-based laboratory exercises on the acquisition of line graph construction and interpretation skills by high school biology students. *Journal of Research in Science Teaching*, *27*(8), 777–787. https://doi.org/10.1002/tea.3660270807

Ahmed, A., Hurwitz, D., Gestson, S., & Brown, S. (2021). Differences between Professionals and Students in Their Visual Attention on Multiple Representation Types While Solving an Open-Ended Engineering Design Problem. *Journal of Civil Engineering Education*, *147*(3), 04021005. https://doi.org/10.1061/(ASCE)EI.2643-9115.0000044

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, *33*(2), 131–152. https://doi.org/10.1016/S0360-1315(99)00029-9

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to Learn in Science. *Science*, *333*(6046), 1096–1097. https://doi.org/10.1126/science.1204153

Ainsworth, S., & Scheiter, K. (2021). Learning by Drawing Visual Representations: Potential, Purposes, and Practical Implications. *Current Directions in Psychological Science*, *30*(1), 61–67. https://doi.org/10.1177/0963721420979582

Albelihi, H. H. M. (2022). Intermediate English as a Foreign Language learners’ formulaic language speaking proficiency: Where does the teaching of lexical chunks figure? *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.949675

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development*, *68*(5), 2095–2119. https://doi.org/10.1007/s11423-020-09748-7

Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., & Sabena, C. (2015). Reading mathematics representations: An eye-tracking study. *International Journal of Science and Mathematics Education*, *13*(2), 237–259. https://doi.org/10.1007/s10763-013-9484-y

Atkins, R. M., & McNeal, K. S. (2018). Exploring Differences Among Student Populations During Climate Graph Reading Tasks: An Eye Tracking Study. *Journal of Astronomy & Earth Sciences Education (JAESE)*, *5*(2), Article 2. https://doi.org/10.19030/jaese.v5i2.10219

Ayres, P. (2015). State-of-the-Art Research into Multimedia Learning: A Commentary on Mayer's Handbook of Multimedia Learning. *Applied Cognitive Psychology*, *29*(4), 631–636. https://doi.org/10.1002/acp.3142

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Bayri, N. G., & Kurnaz, M. A. (2015). Investigation Of The Secondary School Students Transition Situations Between Different Representation Types. *The Turkish Online Journal of Educational Technology*, *2*, 336–340.

Becker, S., Küchemann, S., Klein, P., Lichtenberger, A., & Kuhn, J. (2022). Gaze patterns enhance response prediction: More than correct or incorrect. *Physical Review Physics Education Research*, *18*(2), 020107. https://doi.org/10.1103/PhysRevPhysEducRes.18.020107

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs | American Journal of Physics | AIP Publishing. *American Journal of Physics*, *62*(8), 750–762. https://doi.org/10.1119/1.17449

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Boels, L., Bakker, A., Van Dooren, W., & Drijvers, P. (2019). Conceptual difficulties when interpreting histograms: A review. *Educational Research Review*, *28*, 100291. https://doi.org/10.1016/j.edurev.2019.100291

Bowen, G. M., Roth, W.-M., & McGinn, M. K. (1999). Interpretations of graphs by university biology students and practicing scientists: Toward a social practice view of scientific representation practices. *Journal of Research in Science Teaching*, *36*(9), 1020–1043. https://doi.org/10.1002/(SICI)1098-2736(199911)36:9<1020::AID-TEA4>3.0.CO;2-#

Brams, S., Ziv, G., Levin, O., Spitz, J., Wagemans, J., Williams, A. M., & Helsen, W. F. (2019). The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological Bulletin*, *145*(10), 980–1027. https://doi.org/10.1037/bul0000207

Brod, G. (2021). Generative Learning: Which Strategies for What Age? *Educational Psychology Review*, *33*(4), 1295–1318. https://doi.org/10.1007/s10648-020-09571-9

Brückner, S., Zlatkin-Troitschanskaia, O., Küchemann, S., Klein, P., & Kuhn, J. (2020). Changes in Students' Understanding of and Visual Attention on Digitally Represented Graphs Across Two Domains in Higher Education: A Postreplication Study. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.02090

Butcher, K. R. (2014). The Multimedia Principle. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 174–205). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.010

Camp, G., Surma, T., & Kirschner, P. A. (2021). Foundations of Multimedia Learning. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 17–24). Cambridge University Press. https://doi.org/10.1017/9781108894333.004

Caruso, M., Peacock, C. E., Southwell, R., Zhou, G., & D'Mello, S. K. (2022). Going Deep and Far: Gaze-Based Models Predict Multiple Depths of Comprehension during and One Week Following Reading. *Proceedings of the 15th International Conference on Educational Data Mining*, 145–157. https://eric.ed.gov/?id=ED624054

Ceuppens, S., Bollen, L., Deprez, J., Dehaene, W., & De Cock, M. (2019). 9th grade students' understanding and strategies when solving \$x(t)\$ problems in 1D kinematics and \$y(x)\$ problems in mathematics. *Physical Review Physics Education Research*, *15*(1), 010101. https://doi.org/10.1103/PhysRevPhysEducRes.15.010101

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, *8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

Chang, H.-Y. (2018). Students' representational competence with drawing technology across two domains of science. *Science Education*, *102*(5), 1129–1149. https://doi.org/10.1002/sce.21457

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chen, S.-C., She, H.-C., Chuang, M.-H., Wu, J.-Y., Tsai, J.-L., & Jung, T.-P. (2014). Eye movements predict students' computer-based assessment performance of physics concepts in different presentation modalities. *Computers & Education*, *74*, 61–72. https://doi.org/10.1016/j.compedu.2013.12.012

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Clement, J. (1985). Misconceptions in graphing. *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education*, *1*, 369–375. http://people.umass.edu/clement/pdf/Misconceptions%20in%20Graphing.pdf

Clement, J. (1989). The concept of variation and misconceptions in Cartesian graphing. *Focus on Learning Problems in Mathematics*, *11*, 77–87.

Cojean, S., & Jamet, E. (2022). Does an interactive table of contents promote learning from videos? A study of consultation strategies and learning outcomes. *British Journal of Educational Technology*, *53*(2), 269–285. https://doi.org/10.1111/bjet.13164

Coskun, A., & Cagiltay, K. (2022). A systematic review of eye-tracking-based research on animated multimedia learning. *Journal of Computer Assisted Learning*, *38*(2), 581–598. https://doi.org/10.1111/jcal.12629

Cromley, J. G., Du, Y., & Dane, A. P. (2020). Drawing-to-Learn: Does Meta-Analysis Show Differences Between Technology-Based Drawing and Paper-and-Pencil Drawing? *Journal of Science Education and Technology*, *29*(2), 216–229. https://doi.org/10.1007/s10956-019-09807-6

Curcio, F. R. (1987). Comprehension of Mathematical Relationships Expressed in Graphs. *Journal for Research in Mathematics Education*, *18*(5), 382–393. https://doi.org/10.5951/jresematheduc.18.5.0382

Dewi, A. R. C., Putra, N. M. D., & Susilo. (2018). Analysis of graphic representation ability in oscillation phenomena. *Journal of Physics: Conference Series*, *983*(1), 012024. https://doi.org/10.1088/1742-6596/983/1/012024

Díaz-Levicoy, D., Osorio, M., & Ruz, F. (2018). *Activities and types of statistical graphs in textbooks of primary education in Perú* [Contribución a Actas de Congreso]. International Statistical Institute. http://funes.uniandes.edu.co/12567/

Dirkx, K. J. H., Skuballa, I., Manastirean-Zijlstra, C. S., & Jarodzka, H. (2021). Designing computer-based tests: Design guidelines from multimedia learning studied with eye tracking. *Instructional Science*, *49*(5), 589–605. https://doi.org/10.1007/s11251-021-09542-9

Donnelly-Hermosillo, D. F., Gerard, L. F., & Linn, M. C. (2020). Impact of graph technologies in K-12 science and mathematics education. *Computers & Education*, *146*, 103748. https://doi.org/10.1016/j.compedu.2019.103748

Dzsotjan, D., Ludwig-Petsch, K., Mukhametov, S., Ishimaru, S., Kuechemann, S., & Kuhn, J. (2021). The Predictive Power of Eye-Tracking Data in an Interactive AR Learning Environment. *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 467–471. https://doi.org/10.1145/3460418.3479358

Engledowl, C., & Weiland, T. (2021). Data (Mis)representation and COVID-19: Leveraging Misleading Data Visualizations For Developing Statistical Literacy Across Grades 6–16. *Journal of Statistics and Data Science Education*, *29*(2), 160–164. https://doi.org/10.1080/26939169.2021.1915215

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Fiorella, L. (2023). Making Sense of Generative Learning. *Educational Psychology Review*, *35*(2), 50. https://doi.org/10.1007/s10648-023-09769-7

Fiorella, L., & Kuhlmann, S. (2020). Creating drawings enhances learning by teaching. *Journal of Educational Psychology*, *112*(4), 811–822. https://doi.org/10.1037/edu0000392

Fiorella, L., & Mayer, R. E. (2016). Eight Ways to Promote Generative Learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fiorella, L., & Mayer, R. E. (2021). The Generative Activity Principle in Multimedia Learning. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 339–350). Cambridge University Press. https://doi.org/10.1017/9781108894333.036

Fiorella, L., & Zhang, Q. (2018). Drawing Boundary Conditions for Learning by Drawing. *Educational Psychology Review*, *30*(3), 1115–1137. https://doi.org/10.1007/s10648-018-9444-8

Flegr, S., Kuhn, J., & Scheiter, K. (2023). When the whole is greater than the sum of its parts: Combining real and virtual experiments in science education. *Computers & Education*, *197*, 104745. https://doi.org/10.1016/j.compedu.2023.104745

Fleischer, T., Deibl, I., Moser, S., Strahl, A., Maier, S., & Zumbach, J. (2023). Mobile Eye Tracking during Experimenting with Digital Scaffolding—Gaze Shifts between Augmented Reality and Experiment during Zinc Iodide Electrolysis Set-Up. *Education Sciences*, *13*(2), Article 2. https://doi.org/10.3390/educsci13020170

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158. https://doi.org/10.2307/749671

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness Fading in Mathematics and Science Instruction: A Systematic Review. *Educational Psychology Review*, *26*(1), 9–25. https://doi.org/10.1007/s10648-014-9249-3

Galesic, M., & Garcia-Retamero, R. (2011). Graph Literacy: A Cross-Cultural Comparison. *Medical Decision Making*, *31*(3), 444–457. https://doi.org/10.1177/0272989X10373805

García-García, J., & Dolores-Flores, C. (2021). Pre-university students' mathematical connections when sketching the graph of derivative and antiderivative functions. *Mathematics Education Research Journal*, *33*(1), 1–22. https://doi.org/10.1007/s13394-019-00286-x

Gardner, S. M., Suazo-Flores, E., Maruca, S., Abraham, J. K., Karippadath, A., & Meir, E. (2021). Biology Undergraduate Students' Graphing Practice in Digital Versus

Pen and Paper Graphing Environments. *Journal of Science Education and Technology*, *30*(3), 431–446. https://doi.org/10.1007/s10956-020-09886-w

Gegenfurtner, A., Gruber, H., Holzberger, D., Keskin, Ö., Lehtinen, E., Seidel, T., Stürmer, K., & Säljö, R. (2023). Towards a cognitive theory of visual expertise: Methods of inquiry. In *Re-theorising Learning and Research Methods in Learning Research* (1st edition). Routledge.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Gerard, L., Zertuche, A., & Linn, M. (2012). *Learning to Graph: A Comparison Study of Using Probe or Draw Tools in a Web-Based Learning Environment*. https://repository.isls.org//handle/1/2380

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media Inc. https://books.google.com.tr/books?hl=de&lr=&id=HnetDwAAQBAJ&oi=fnd&pg=PT9&ots=kPZtBDxHAa&sig=ld9Y1wenAwsD3KdxCwgIKNJXSrg&redir_esc=y#v=onepage&q&f=false

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, *47*(2), 183–210. https://doi.org/10.1080/03057267.2011.605307

Guida, A., Gobet, F., Tardieu, H., & Nicolas, S. (2012). How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: A two-stage framework. *Brain and Cognition*, *79*(3), 221–244. https://doi.org/10.1016/j.bandc.2012.01.010

Gültepe, N. (2016). Reflections on high school students' graphing skills and their conceptual understanding of drawing chemistry graphs. *Educational Sciences: Theory & Practice*, *16*(1), 53–81. https://doi.org/10.12738/estp.2016.1.2837

Gultepe, N., & Kilic, Z. (2015). Effect of Scientific Argumentation on the Development of Scientific Process Skills in the Context of Teaching Chemistry. *International Journal of Environmental and Science Education*, *10*(1), 111–132.

Hahn, L., & Klein, P. (2022). Eye tracking in physics education research: A systematic literature review. *Physical Review Physics Education Research*, *18*(1), 013102. https://doi.org/10.1103/PhysRevPhysEducRes.18.013102

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 172–190. https://doi.org/10.1037/0278-7393.25.1.172

Harrison, E., McBride, B., & Linn, M. C. (2019). Revision Analysis of Students' Position-Time Graphs. *Computersupported Collaborative Learning*, *2*. https://par.nsf.gov/biblio/10181648-revision-analysis-students-position-time-graphs

Harsh, J. A., Campillo, M., Murray, C., Myers, C., Nguyen, J., & Maltese, A. V. (2019). "Seeing" Data Like an Expert: An Eye-Tracking Study Using Graphical Data Representations. *CBE—Life Sciences Education*, *18*(3), ar32. https://doi.org/10.1187/cbe.18-06-0102

Harsh, J. A., & Schmitt-Harsh, M. (2016). Instructional Strategies to Develop Graphing Skills in the College Science Classroom. *The American Biology Teacher*, *78*(1), 49–56. https://doi.org/10.1525/abt.2016.78.1.49

Herrlinger, S., Höffler, T. N., Opfermann, M., & Leutner, D. (2017). When Do Pictures Help Learning from Expository Text? Multimedia and Modality Effects in Primary Schools. *Research in Science Education*, *47*(3), 685–704. https://doi.org/10.1007/s11165-016-9525-y

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., & Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education*, *9*(3), e3310. https://doi.org/10.1002/rev3.3310

Holmqvist, K., & Andersson, R. (2017). *Eye Tracking: A Comprehensive Guide to Methods, Paradigms, and Measures*. Lund Eye-Tracking Research Institute.

Hoyer, C., & Girwidz, R. (2020). Animation and interactivity in computer-based physics experiments to support the documentation of measured vector quantities in diagrams: An eye tracking study. *Physical Review Physics Education Research*, *16*(2), 020124. https://doi.org/10.1103/PhysRevPhysEducRes.16.020124

Huang, Y., Miller, K. F., Cortina, K. S., & Richter, D. (2023). Teachers' professional vision in action. *Zeitschrift Für Pädagogische Psychologie*, *37*(1–2), 122–139. https://doi.org/10.1024/1010-0652/a000313

Javora, O., Hannemann, T., Volná, K., Děchtěrenko, F., Tetourová, T., Stárková, T., & Brom, C. (2021). Is contextual animation needed in multimedia learning games for children? An eye tracker study. *Journal of Computer Assisted Learning*, *37*(2), 305–318. https://doi.org/10.1111/jcal.12489

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kalyuga, S. (2011). Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Kanzaki, N., & Miwa, K. (2011). Experimental investigation of effects of representations and contexts on comprehension and generation of line graphs. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*, 2196–2201. https://escholarship.org/content/qt0949g34v/qt0949g34v.pdf

Kastberg, S., & Leatham, K. (2005). Research On Graphing Calculators at the Secondary Level: Implications for Mathematics Teacher Education. *Contemporary Issues in Technology and Teacher Education*, *5*(1), 25–37.

Kennel, K. (2022). Using Eye Tracking Data for Enhancing Adaptive Learning Systems. *2022 Symposium on Eye Tracking Research and Applications*, 1–3. https://doi.org/10.1145/3517031.3532195

Klein, P., Küchemann, S., Brückner, S., Zlatkin-Troitschanskaia, O., & Kuhn, J. (2019). Student understanding of graph slope and area under a curve: A replication study comparing first-year physics and economics students. *Physical Review Physics Education Research*, *15*(2), 020116. https://doi.org/10.1103/PhysRevPhysEducRes.15.020116

Klein, P., Lichtenberger, A., Küchemann, S., Becker, S., Kekule, M., Viiri, J., Baadte, C., Vaterlaus, A., & Kuhn, J. (2020). Visual attention while solving the test of understanding graphs in kinematics: An eye-tracking analysis. *European Journal of Physics*, *41*(2), 025701. https://doi.org/10.1088/1361-6404/ab5f51

Klein, P., Viiri, J., & Kuhn, J. (2019). Visual cues improve students' understanding of divergence and curl: Evidence from eye movements during reading and problem solving. *Physical Review Physics Education Research*, *15*(1), 010126. https://doi.org/10.1103/PhysRevPhysEducRes.15.010126

Kohnle, A., Ainsworth, S., & Passante, G. (2020). Sketching to support visual learning with interactive tutorials. *Physical Review Physics Education Research*, *16*(2), 020139. https://doi.org/10.1103/PhysRevPhysEducRes.16.020139

Kokkonen, T., & Schalk, L. (2021). One Instructional Sequence Fits all? A Conceptual Analysis of the Applicability of Concreteness Fading in Mathematics, Physics, Chemistry, and Biology Education. *Educational Psychology Review*, *33*(3), 797–821. https://doi.org/10.1007/s10648-020-09581-7

Kollmer, J., Schleinschok, K., Scheiter, K., & Eitel, A. (2020). Is drawing after learning effective for metacognitive monitoring only when supported by spatial scaffolds? *Instructional Science*, *48*(5), 569–589. https://doi.org/10.1007/s11251-020-09521-6

Küchemann, S., Becker, S., Klein, P., & Kuhn, J. (2021). Gaze-Based Prediction of Students' Understanding of Physics Line-Graphs: An Eye-Tracking-Data Based Machine-Learning Approach. In H. C. Lane, S. Zvacek, & J. Uhomoibhi (Eds.), *Computer Supported Education* (pp. 450–467). Springer International Publishing. https://doi.org/10.1007/978-3-030-86439-2_23

Küchemann, S., Klein, P., Becker, S., Kumari, N., & Kuhn, J. (2020). Classification of Students' Conceptual Understanding in STEM Education using Their Visual Attention Distributions: A Comparison of Three Machine-Learning Approaches. *Proceedings of the 12th International Conference on Computer Supported Education - Volume 1: CSEDU*, *1*, 36–46. https://doi.org/10.5220/0009359400360046

Kumari, N., Ruf, V., Mukhametov, S., Schmidt, A., Kuhn, J., & Küchemann, S. (2021). Mobile Eye-Tracking Data Analysis Using Object Detection via YOLO v4. *Sensors*, *21*(22), Article 22. https://doi.org/10.3390/s21227668

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic Component of Image Perception in Mammogram Interpretation: Gaze-tracking Study. *Radiology*, *242*(2), 396–402. https://doi.org/10.1148/radiol.2422051997

Lachmayer, S., Nerdel, C., & Prechtl, H. (2007). Modellierung kognitiver Fähigkeiten beim Umgang mit Diagrammen im naturwissenschaftlichen Unterricht [Modelling of cognitive abilities regarding the handling of graphs in science education]. *Zeitschrift für Didaktik der Naturwissenschaften*, *13*, 145–160.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90–115. https://doi.org/10.1016/j.edurev.2013.10.001

Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, *11*(1), 65–100. https://doi.org/10.1016/S0364-0213(87)80026-5

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, Graphs, and Graphing: Tasks, Learning, and Teaching. *Review of Educational Research*, *60*(1), 1–64. https://doi.org/10.3102/00346543060001001

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). The heuristic interpretation of box plots. *Learning and Instruction*, *26*, 22–35. https://doi.org/10.1016/j.learninstruc.2013.01.001

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Lin, L., Lee, C. H., Kalyuga, S., Wang, Y., Guan, S., & Wu, H. (2017). The Effect of Learner-Generated Drawing and Imagination in Comprehending a Science Text. *The Journal of Experimental Education*, *85*(1), 142–154. https://doi.org/10.1080/00220973.2016.1143796

Lindner, M. A., Eitel, A., Barenthien, J., & Köller, O. (2021). An integrative study on learning and testing with multimedia: Effects on students' performance and metacognition. *Learning and Instruction*, *71*, 101100. https://doi.org/10.1016/j.learninstruc.2018.01.002

Lindner, M. A., Eitel, A., Strobel, B., & Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An eye-movement analysis. *Learning and Instruction*, *47*, 91–102. https://doi.org/10.1016/j.learninstruc.2016.10.007

Lounis, C., Peysakhovich, V., & Causse, M. (2021). Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study. *PLOS ONE*, *16*(2), e0247061. https://doi.org/10.1371/journal.pone.0247061

Ludewig, U., Lambert, K., Dackermann, T., Scheiter, K., & Möller, K. (2020). Influences of basic numerical abilities on graph reading performance. *Psychological Research*, *84*(5), 1198–1210. https://doi.org/10.1007/s00426-019-01144-y

Mahmoud, R., & Zoghaib, S. Z. (2023). The Effects of Different Data Visualisation Formats on News Recall and Comprehension. *Media Watch*, *14*(2), 155–176. https://doi.org/10.1177/09760911231158746

Malone, S., Altmeyer, K., Vogel, M., & Brünken, R. (2020). Homogeneous and heterogeneous multiple representations in equation-solving problems: An eye-tracking study. *Journal of Computer Assisted Learning*, *36*(6), 781–798. https://doi.org/10.1111/jcal.12426

Maries, A., & Singh, C. (2023). Helping Students Become Proficient Problem Solvers Part I: A Brief Review. *Education Sciences*, *13*(2), Article 2. https://doi.org/10.3390/educsci13020156

Marisda, D. H., Handayani, Y., & Rahmawati. (2020). The combination of interactive conceptual learning models and multimedia interactive to minimize misconceptions on the science content. *Journal of Physics: Conference Series*, *1572*(1), 012069. https://doi.org/10.1088/1742-6596/1572/1/012069

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*(1), 95–109. https://doi.org/10.1016/j.compedu.2012.07.011

Mayer, R. E. (2014a). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Mayer, R. E. (2014b). Introduction to Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 1–24). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.002

Mayer, R. E. (Ed.). (2014c). *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Mayer, R. E. (2021). Cognitive Theory of Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 57–72). Cambridge University Press. https://doi.org/10.1017/9781108894333.008

McDermott, L. C., Rosenquist, M. L., & van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics*, *55*(6), 503–513. https://doi.org/10.1119/1.15104

Midway, S. R. (2020). Principles of Effective Data Visualization. *Patterns*, *1*(9). https://doi.org/10.1016/j.patter.2020.100141

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, *141*, 103618. https://doi.org/10.1016/j.compedu.2019.103618

Namoun, A., & Alshanqiti, A. (2021). Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. *Applied Sciences*, *11*(1), Article 1. https://doi.org/10.3390/app11010237

Nixon, R. S., Godfrey, T. J., Mayhew, N. T., & Wiegert, C. C. (2016). Undergraduate student construction and interpretation of graphs in physics lab activities. *Physical Review Physics Education Research*, *12*(1), 010104. https://doi.org/10.1103/PhysRevPhysEducRes.12.010104

Okan, Y., Galesic, M., & Garcia-Retamero, R. (2016). How People with Low and High Graph Literacy Process Health Graphs: Evidence from Eye-tracking. *Journal of Behavioral Decision Making*, *29*(2–3), 271–294. https://doi.org/10.1002/bdm.1891

Orru, G., & Longo, L. (2019). The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads: A Review. In L. Longo & M. C. Leva (Eds.), *Human Mental Workload: Models and Applications* (pp. 23–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-14273-5_3

Oslington, G., Mulligan, J., & Van Bergen, P. (2020). Third-graders' predictive reasoning strategies. *Educational Studies in Mathematics*, *104*(1), 5–24. https://doi.org/10.1007/s10649-020-09949-0

Ozmen, Z. M., Guven, B., & Kurak, Y. (2020). Determining the Graphical Literacy Levels of the 8th Grade Students. *Eurasian Journal of Educational Research*, *20*(86), Article 86.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*(3), 241–263. https://doi.org/10.1037/h0027272

Penglase, M., & Arnold, S. (1996). The graphics calculator in mathematics education: A critical review of recent research. *Mathematics Education Research Journal*, *8*(1), 58–90. https://doi.org/10.1007/BF03355481

Pospiech, G., Geyer, M. A., Ceuppens, S., Cock, M. D., Deprez, J., Dehaene, W., Michelini, M., & Stefanel, A. (2019). Role of graphs in the mathematization process in physics education. *Journal of Physics: Conference Series*, *1287*(1), 012014. https://doi.org/10.1088/1742-6596/1287/1/012014

Program for International Student Assessment. (2022). *PISA 2022 Mathematics Framework*. https://pisa2022-maths.oecd.org/#Twenty-First-Century-Skills

Rincón-Flores, E. G., López-Camacho, E., Mena, J., & Olmos, O. (2022). *Teaching through Learning Analytics: Predicting Student Learning Profiles in a Physics Course at a Higher Education Institution*. https://doi.org/10.9781/ijimai.2022.01.005

Rodemer, M., Eckhard, J., Graulich, N., & Bernholt, S. (2021). Connecting explanations to representations: Benefits of highlighting techniques in tutorial videos on students' learning in organic chemistry. *International Journal of Science Education*, *43*(17), 2707–2728. https://doi.org/10.1080/09500693.2021.1985743

Roth, W.-M., & Bowen, G. M. (2001). Professionals Read Graphs: A Semiotic Analysis. *Journal for Research in Mathematics Education*, *32*(2), 159–194. https://doi.org/10.2307/749672

Rubel, L. H., Nicol, C., & Chronaki, A. (2021). A critical mathematics perspective on reading data visualizations: Reimagining through reformatting, reframing, and renarrating. *Educational Studies in Mathematics*, *108*(1), 249–268. https://doi.org/10.1007/s10649-021-10087-4

Ruf, V., Küchemann, S., Kuhn, J., & Klein, P. (2022). Comparison of Written and Spoken Instruction to Foster Coordination between Diagram and Equation in Undergraduate Physics Education. *Human Behavior and Emerging Technologies*, *2022*, e5860936. https://doi.org/10.1155/2022/5860936

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71–78. https://doi.org/10.1145/355017.355028

Scheiter, K., Schleinschok, K., & Ainsworth, S. (2017). Why Sketching May Aid Learning From Science Texts: Contrasting Sketching With Written Explanations. *Topics in Cognitive Science*, *9*(4), 866–882. https://doi.org/10.1111/tops.12261

Schindler, M., & Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: Towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics*, *101*(1), 123–139. https://doi.org/10.1007/s10649-019-9878-z

Schmidgall, S. P., Eitel, A., & Scheiter, K. (2019). Why do learners who draw perform well? Investigating the role of visualization, generation and externalization in learner-generated drawing. *Learning and Instruction*, *60*, 138–153. https://doi.org/10.1016/j.learninstruc.2018.01.006

Schnotz, W. (2014). Integrated Model of Text and Picture Comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 72–103). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.006

Schüler, A., & Merkt, M. (2021). Investigating text-picture integration in videos with the multimedia contradiction paradigm. *Journal of Computer Assisted Learning*, *37*(3), 718–734. https://doi.org/10.1111/jcal.12518

Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behavior*, *27*(1), 89–93. https://doi.org/10.1016/j.chb.2010.05.028

Shah, P. (1997). A model of the cognitive and perceptual processes in graphical display comprehension. *Reasoning with Diagrammatic Representations*, 94–101.

Shah, P., & Hoeffner, J. (2002). Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review*, *14*(1), 47–69. https://doi.org/10.1023/A:1013180410169

Sheridan, H., & Reingold, E. M. (2017). The Holistic Processing Account of Visual Expertise in Medical Image Perception: A Review. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01620

Simon, H. A. (1974). How Big Is a Chunk? By combining data from several experiments, a basic human memory unit can be identified and measured. *Science*, *183*(4124), 482–488.

Skulmowski, A., & Xu, K. M. (2022). Understanding Cognitive Load in Digital and Online Learning: A New Perspective on Extraneous Cognitive Load. *Educational Psychology Review*, *34*(1), 171–196. https://doi.org/10.1007/s10648-021-09624-7

Slamucka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 153–163. https://doi.org/10.1016/S0022-5371(83)90112-3

Smit, J., Bakker, A., van Eerde, D., & Kuijpers, M. (2016). Using genre pedagogy to promote student proficiency in the language required for interpreting line graphs. *Mathematics Education Research Journal*, *28*(3), 457–478. https://doi.org/10.1007/s13394-016-0174-2

Smith, G. C. S., Seaman, S. R., Wood, A. M., Royston, P., & White, I. R. (2014). Correcting for Optimistic Prediction in Small Data Sets. *American Journal of Epidemiology*, *180*(3), 318–324. https://doi.org/10.1093/aje/kwu140

Steinert, S., Ruf, V., Dzsotjan, D., Großmann, N., Schmidt, A., Kuhn, J., & Küchemann, S. (2024). A refined approach for evaluating small datasets via binary classification using machine learning. *PLOS ONE, 19*(5), 1-21. https://doi.org/10.1371/journal.pone.0301276

Sternberg, R. J. (2003). What Is an "Expert Student?" *Educational Researcher*, *32*(8), 5–9. https://doi.org/10.3102/0013189X032008005

Strobel, B., Grund, S., & Lindner, M. A. (2019). Do seductive details do their damage in the context of graph comprehension? Insights from eye movements. *Applied Cognitive Psychology*, *33*(1), 95–108. https://doi.org/10.1002/acp.3491

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, *104*(2), 147–200. https://doi.org/10.1007/s10649-020-09948-1

Susac, A., Bubic, A., Kazotti, E., Planinic, M., & Palmovic, M. (2018). Student understanding of graph slope and area under a graph: A comparison of physics and nonphysics students. *Physical Review Physics Education Research*, *14*(2), 020109. https://doi.org/10.1103/PhysRevPhysEducRes.14.020109

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Theodoridis, S. (2020). Chapter 7 - Classification: A Tour of the Classics. In S. Theodoridis (Ed.), *Machine Learning (Second Edition)* (2nd ed., pp. 301–350). Academic Press. https://doi.org/10.1016/B978-0-12-818803-3.00016-7

Topczewski, J. J., Topczewski, A. M., Tang, H., Kendhammer, L. K., & Pienta, N. J. (2017). NMR Spectra through the Eyes of a Student: Eye Tracking Applied to NMR Items. *Journal of Chemical Education*, *94*(1), 29–37. https://doi.org/10.1021/acs.jchemed.6b00528

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, *58*(1), 375–385. https://doi.org/10.1016/j.compedu.2011.07.012

Tunga, Y., & Cagiltay, K. (2023). Looking through the model's eye: A systematic review of eye movement modeling example studies. *Education and Information Technologies*, *28*(8), 9607–9633. https://doi.org/10.1007/s10639-022-11569-5

Utrecht University. (2021). *ASReview Lab* (0.18) [Computer software]. https://asreview.nl

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, *3*(2), Article 2. https://doi.org/10.1038/s42256-020-00287-7

van Gog, T. (2014). The Signaling (or Cueing) Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 263–278). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.014

van Gog, T. (2021). The Signaling (or Cueing) Principle in Multimedia Learning. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 221–230). Cambridge University Press. https://doi.org/10.1017/9781108894333.022

van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., de Bock, J. J. P. R., & van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, *32*, 10–21. https://doi.org/10.1016/j.learninstruc.2014.01.004

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, *31*(2), 142–166. https://doi.org/10.1016/j.cedpsych.2005.04.001

Van Meter, P., & Garner, J. (2005). The Promise and Practice of Learner-Generated Drawing: Literature Review and Synthesis. *Educational Psychology Review*, *17*(4), 285–325. https://doi.org/10.1007/s10648-005-8136-3

Vitale, J. M., Applebaum, L., & Linn, M. C. (2019). Coordinating between Graphs and Science Concepts: Density and Buoyancy. *Cognition and Instruction*, *37*(1), 38–72. https://doi.org/10.1080/07370008.2018.1539736

von Kotzebue, L., Gerstl, M., & Nerdel, C. (2015). Common Mistakes in the Construction of Diagrams in Biological Contexts. *Research in Science Education*, *45*(2), 193–213. https://doi.org/10.1007/s11165-014-9419-9

Wang, Q., Zhu, Y., Wei, L., & Deng, H. (2022). Visual Attention Pattern of Middle School Students During Problem-Solving in Physics. *Mind, Brain, and Education*, *16*(2), 99–111. https://doi.org/10.1111/mbe.12313

Wavering, M. J. (1985). *Logical reasoning necessary to make line graphs*. 1–29. https://files.eric.ed.gov/fulltext/ED254409.pdf

Wetzels, S. A. J., Kester, L., & van Mariënboer, J. (2010). Use of external representations in science: Prompting and reinforcing prior knowledge activation. In *Use of Representations in Reasoning and Problem Solving* (1st ed., pp. 225–241). Routledge.

Wittrock, M. C. (1974). A Generative Model of Mathematics Learning. *Journal for Research in Mathematics Education*, *5*(4), 181–196. https://doi.org/10.2307/748845

Wu, S. P. W., & Rau, M. A. (2018). Effectiveness and efficiency of adding drawing prompts to an interactive educational technology when learning with visual representations. *Learning and Instruction*, *55*, 93–104. https://doi.org/10.1016/j.learninstruc.2017.09.010

Wu, S. P. W., & Rau, M. A. (2019). How Students Learn Content in Science, Technology, Engineering, and Mathematics (STEM) Through Drawing Activities.

Educational Psychology Review, *31*(1), 87–120. https://doi.org/10.1007/s10648-019-09467-3

Xie, H., Zhao, T., Deng, S., Peng, J., Wang, F., & Zhou, Z. (2021). Using eye movement modelling examples to guide visual attention and foster cognitive performance: A meta-analysis. *Journal of Computer Assisted Learning*, *37*(4), 1194–1206. https://doi.org/10.1111/jcal.12568

Yen, M.-H., Lee, C.-N., & Yang, Y.-C. (2012). Eye Movement Patterns in Solving Scientific Graph Problems. In P. Cox, B. Plimmer, & P. Rodgers (Eds.), *Diagrammatic Representation and Inference* (pp. 343–345). Springer. https://doi.org/10.1007/978-3-642-31223-6_46

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition*, *27*(6), 1073–1079. https://doi.org/10.3758/BF03201236

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, Q., & Fiorella, L. (2021). Learning by drawing: When is it worth the time and effort? *Contemporary Educational Psychology*, *66*, 101990. https://doi.org/10.1016/j.cedpsych.2021.101990

Zhang, Y., Guo, X., Pi, Z., & Yang, J. (2021). Learning by Drawing in STEM: A Meta-Analysis. *2021 Tenth International Conference of Educational Innovation through Technology (EITT)*, 83–87. https://doi.org/10.1109/EITT53287.2021.00025

Zheng, X., Ma, Y., Yue, T., & Yang, X. (2023). Effects of different types of cues and self-explanation prompts in instructional videos on deep learning: Evidence from multiple data analysis. *Educational Technology Research and Development*, *71*(3), 807–831. https://doi.org/10.1007/s11423-023-10188-2

Eidesstattliche Versicherung
Statement of Scientific Integrity

Ruf, Verena
Name, Vorname
Last name, first name

Ich versichere, dass ich die an der Fakultät für Psychologie und Pädagogik der Ludwig-Maximilians-Universität München zur Dissertation eingereichte Arbeit mit dem Titel:
I assert that the thesis I submitted to the Faculty of Psychology and Pedagogy of the Ludwig-Maximilian-Universität München under the title:

Graphical Representations of Data in STEM Education: Investigation of Graphing and Graph Comprehension

selbst verfasst, alle Teile eigenständig formuliert und keine fremden Textteile übernommen habe, die nicht als solche gekennzeichnet sind. Kein Abschnitt der Doktorarbeit wurde von einer anderen Person formuliert, und bei der Abfassung wurden keine anderen als die in der Abhandlung aufgeführten Hilfsmittel benutzt.
is written by myself, I have formulated all parts independently and I have not taken any texts components of others without indicating them. No formulation has been made by someone else and I have not used any sources other than indicated in the thesis.

Ich erkläre, das ich habe an keiner anderen Stelle einen Antrag auf Zulassung zur Promotion gestellt oder bereits einen Doktortitel auf der Grundlage des vorgelegten Studienabschlusses erworben und mich auch nicht einer Doktorprüfung erfolglos unterzogen.
I assert I have not applied anywhere else for a doctoral degree nor have I obtained a doctor title on the basis of my present studies or failed a doctoral examination.

München, 06.11.2023
Ort, Datum
Place, Date

Verena Ruf
Unterschrift Doktorandin/Doktorand
Signature of the doctoral candidate