Eye-tracking methodology in mathematics education research: A systematic literature review

Anselm R. Strohmaier1 · Kelsey J. MacKay2 · Andreas Obersteiner3 · Kristina M. Reiss1

Published online: 9 June 2020
© The Author(s) 2020

Abstract
Eye tracking is an increasingly popular method in mathematics education. While the technology has greatly evolved in recent years, there is a debate about the specific benefits that eye tracking offers and about the kinds of insights it may allow. The aim of this review is to contribute to this discussion by providing a comprehensive overview of the use of eye tracking in mathematics education research. We reviewed 161 eye-tracking studies published between 1921 and 2018 to assess what domains and topics were addressed, how the method was used, and how eye movements were related to mathematical thinking and learning. The results show that most studies were in the domain of numbers and arithmetic, but that a large variety of other areas of mathematics education research was investigated as well. We identify a need to report more methodological details in eye-tracking studies and to be more critical about how to gather, analyze, and interpret eye-tracking data. In conclusion, eye tracking seemed particularly beneficial for studying processes rather than outcomes, for revealing mental representations, and for assessing subconscious aspects of mathematical thinking.

Keywords  Cognitive processes · Eye movements · Eye tracking · Mathematics education · Numerical cognition

Anselm R. Strohmaier
anselm.strohmaier@tum.de

1 Heinz Nixdorf Chair of Mathematics Education, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany
2 Centre for Instructional Psychology and Technology, Katholieke Universiteit Leuven, Dekenstraat 2, Box 3773, 3000 Leuven, Belgium
3 Institute for Mathematics Education, Freiburg University of Education, Kunzenweg 21, 79117 Freiburg, Germany

Educational Studies in Mathematics (2020) 104:147–200
https://doi.org/10.1007/s10649-020-09948-1
1 Introduction

The eyes play a pivotal role in decoding visual information. Thus, the observation of eye movements can offer insights into cognitive processing (Holmqvist et al., 2011; Just & Carpenter, 1976). This is probably the reason why the method of eye tracking has been used in numerous studies of different domains (see Duchowski, 2007, for an overview), including research on learning (see Lai et al., 2013, for a review) and mathematics education (Barmby, Andrà, Gomez, Obersteiner, & Shvarts, 2014; Hartmann, 2015; Lilienthal & Schindler, 2019; Mock, Huber, Klein, & Moeller, 2016; Perttula, 2017; Schindler, Haataja, Lilienthal, Moreno-Esteva, & Shvarts, 2018).

The increasing emergence of all studies using eye tracking comes with challenges for mathematics education researchers. First, the number of studies has arguably reached a point at which it has become difficult for researchers to monitor. Eye tracking seems to be widespread across many domains of mathematics education research, making it difficult to keep track of its usage, potential, limitations, and specific challenges. Particularly, researchers who use eye tracking should monitor studies not only in their specific domain, but also outside of it, as there could be studies that use a similar methodological approach to a different end.

Second, with the development of new eye-tracking hardware and data analysis software, the possibilities of data collection and interpretation have become increasingly manifold. For instance, using eye-tracking glasses as opposed to a static eye tracker not only introduces the chance to address new research questions but also requires new methodological considerations. Overall, it has become more and more difficult to summarize, compare, and relate results from eye-tracking research. In mathematics education, this trend has been reported in recent years (Lilienthal & Schindler, 2019; Schindler et al., 2018).

Third, among educational research domains, the domain of mathematics is unique in the way it uses text, mathematical symbols, and visualizations and in how these forms of representation are integrated (Andrà et al., 2009; Ott, Brunken, Vogel, & Malone, 2018). Accordingly, studying people’s eye movements during mathematical activities (e.g., reading of mathematical proofs) may require different or modified approaches than those applied successfully in other domains (e.g., text reading or science education). However, these specific affordances of the eye-tracking method in studying mathematics are not well understood.

Lastly, manufacturers have greatly improved the usability of eye-tracking devices and made them more affordable so that an increasing number of researchers has gained access to eye-tracking devices. However, using eye tracking in research still requires complex knowledge, which many researchers in mathematics education had to acquire by themselves (Lilienthal & Schindler, 2019). Accordingly, the mathematics education community has seen an increasing call for guidance and clarity about the possibilities and limitations of the method, as evidenced by working groups at international conferences such as the Annual Conference of the International Group for the Psychology of Mathematics Education (PME; Barmby et al., 2014; Schindler et al., 2018).

Four previous review studies summarized eye-tracking research relevant for mathematics education. Lai et al.’s (2013) review described the state-of-the-art of eye-tracking research in the field of education but was not specific to mathematics education. Moreover, in the 7 years since its publication, the number of studies using eye-tracking technology may have greatly increased, and thus a critical examination of the current potential of eye tracking and its challenges seems timely. The other three studies only covered specific aspects of mathematics.
education and thus lack comprehensiveness. A conference paper published by Perttula (2017) provided a summary of eye-tracking studies in mathematics education. This review included a limited selection of 28 studies that focused on the use of eye tracking in studies on mathematical representations. Mock et al. (2016) reviewed 45 studies in their systematic review; however, it was solely focused on the subdomain of numerical cognition. Most recently, Lilienthal and Schindler (2019) reviewed 34 contributions related to eye tracking in the proceedings of the PME. This review illustrated the continued relevance and the potential of the eye-tracking method but focused only on the proceeding papers of a specific conference, rather than the mathematics education literature as a whole. A systematic, comprehensive review on the use of eye tracking in the mathematics education literature does not yet exist. The current paper seeks to fill this gap.

2 Goals and research questions

We present a systematic review, which critically investigates the use of eye tracking in mathematics education, defining three aims and associated research questions.

Our first aim was to provide an overview of all studies using eye tracking in mathematics education. This included domains and topics that were addressed as well as the date and type of publication. Articulating the range of topics can help researchers to become aware of gaps in the literature, distinguish possible benefits of the method for specific areas of mathematics education, and support them in finding relevant eye-tracking research in fields that are similar to their own. Therefore, our first research questions were:

RQ1a: How many studies used eye tracking in mathematics education and when and in which journals were they published?

RQ1b: In which domains of mathematics education is eye tracking used and what overarching topics have been addressed?

Our second aim was to critically review the methodology of eye-tracking research in mathematics education, including both technical and statistical aspects. To obtain meaningful and reliable data in an eye-tracking experiment, the implementation of the technology must be carefully considered. This issue regards, for example, the calibration and setup of the apparatus, the design of the stimuli, and the analysis of the raw data. We analyzed how these issues were addressed in the reviewed studies and what implications should be drawn for future research. Accordingly, our second research question was:

RQ2: How was the eye-tracking methodology implemented and what details of this implementation were reported in the studies?

The third aim of the current review was to assess the way in which eye-tracking data are typically interpreted in mathematics education research. The interpretation of eye-tracking data is challenging because the same data may be linked to various cognitive processes (Schindler & Lilienthal, 2019). Interpretations vary depending on the research questions and the particular type of tasks involved. Thus, our third research question was:
RQ3: How were eye-tracking data interpreted in mathematics education research?

3 Method

3.1 Paper selection

We selected papers both through a systematic database search and by carefully checking cross-references from all relevant results. First, we conducted a database search in Scopus, PsycARTICLES, Education Source, ERIC, Science Direct, Web of Science, and MathEduc, which arguably reflect the most common sources for studies in the field of mathematics education. We considered studies that were published until and including 2018. The first step was the database search, for which we used the search string: (eye OR gaze) AND (movement* OR track* OR record*) AND (math* OR “numerical cognition”), referring to titles and abstracts. Duplicates were automatically discarded. This resulted in a total of 1491 studies. Step two involved screening titles and abstracts using the following criteria: (a) The study was published in a journal article, a book chapter, or in conference proceedings; (b) the study was published in English; (c) the topic of the study was broadly related to mathematics education, meaning that the study reported using a mathematical task or investigating mathematical learning in any way. After this screening, 188 studies remained.

In a third step, these studies were coded by the first two authors of this paper. During this step, studies that did not meet the following criteria were excluded: (a) Eye-tracking data were directly related to mathematics education. This applied if a study used eye-tracking data to analyze mathematical abilities, the solution of a mathematical task, or the acquisition of mathematical content.2 Mathematical was considered anything that is part of mathematics curricula; (b) the length of the article was at least three pages. This was necessary since shorter papers did not report enough information about the method to meaningfully review their use of eye tracking; (c) if both a conference paper and a journal article or book chapter reported the same data and analyses, we excluded the conference paper. This third step led to the exclusion of 79 studies.

A fourth and final step was to check the references within all studies that met the aforementioned criteria (i.e., steps two and three), which added another 38 studies. In addition, we included 14 studies from an additional manual search in the Conference Proceedings of the PME.3

1 Different search strings were compared and tested in preliminary analyses. Here, it turned out that studies from the field of numerical cognition often did not relate themselves to mathematics but were nevertheless considered relevant for the present review. To avoid a blind spot for those studies, we included numerical cognition as a separate term. This increased the number of hits by up to 8%, depending on the database. Because we did not find potential blind spots for other domains, we did not consider it necessary or practical to include additional terms (e.g., geometry or statistical reasoning).
2 This meant that studies were not included that primarily addressed methodological questions (e.g., Schindler & Lilienthal, 2017) or studies where gaze replays were used as stimuli, but that did not analyze the eye-tracking data itself (e.g., Gallagher-Mitchell, Simms, & Litchfield, 2018; van Marlen, van Wermeskerken, Jarodzka, & van Gog, 2016, 2018; van Wermeskerken, Litchfield, & van Gog, 2018; Wang, Pi, & Hu, 2018).
3 Note that the review by Lilienthal and Schindler (2019) included another 25 PME papers that we did not include in our review for the following reasons: 21 were short reports, 2 did not directly relate data to mathematics education, and 2 have since been published as journal articles that are included in our review.
Eventually, 161 studies were included in this review. Of these, 31 were published in conference proceedings, 5 were book chapters, and the remaining 125 studies were published in journals. It is notable that the total number of studies included in this review was substantially higher than in prior reviews (Perttula, 2017: 28 studies; Mock et al., 2016: 45 studies; Lilienthal & Schindler, 2019: 33 studies).

3.2 Coding procedure

Codes were made according to the research questions in seven overarching categories. With respect to RQ1a and RQ1b, we included the categories (1) publication (e.g., year, journal) and (2) domain and topic (e.g., domain, task type). RQ2 led to the categories (3) apparatus (e.g., manufacturer, sampling rate), (4) stimuli (e.g., task type, presentation, areas of interest), (5) sample and research design (e.g., sample size, procedure, statistical method), and (6) data treatment (e.g., event detection, statistical analysis). RQ3 led to the category (7) interpretation of eye movements (e.g., parameters, interpretation). If a categorization was considered unclear by the coder, a consensus was reached in a discussion between the first three authors of this paper. After the initial coding, all data were cross-checked for coherence, for example, in the description of the stimulus.

A table with selected information from all reviewed studies can be found in Appendix Table 3. The forthcoming results section is structured according to these seven overarching categories.

4 Results

4.1 Publication

To address RQ1a, we analyzed when and in which journals the reviewed studies were published. In line with previous reviews of eye-tracking research, we found a notable increase of studies in the period between 2006 and 2014 (see Fig. 1). Since 2014, the number of published studies was around 20 per year, meaning that 61% of the studies included in this review were published in 2014 or later. Only a small percentage (16%) of those studies that were journal articles were published in journals specialized in mathematics education. The mathematics education journal with the largest number of eye-tracking studies was the International Journal of Science and Mathematics Education (five studies). Most articles were published in journals focusing on psychology, educational psychology, or eye tracking. The most prevalent journals for the articles in the current review were Acta Psychologica (nine studies), Psychological Research (seven studies), and the Quarterly Journal of Experimental Psychology (seven studies).

4 After planning the categories and before the coding of the studies, the first two authors independently coded the same five randomly selected studies. The results from coding these studies were discussed to ensure the information extracted was identical. Throughout the coding procedure, both coders communicated frequently about issues and required refinements of the categories and always used the same set and description of categories.
4.2 Domains and topics

An overview of the domains and topics of mathematics education that were addressed by the studies reviewed here is given in Table 1. In the following, we briefly summarize these categories. For the specific tasks used in the studies, see Appendix Table 3.

Numbers and arithmetic The majority of studies included in the current review (90 studies; 56%) addressed numbers and arithmetic. Within this larger category, we grouped the studies into eight subcategories. Necessarily, these subcategories overlap, and their distinction will leave room for debate. The first subcategory included 16 studies (10%) investigating the perception, mental representation, and basic processing of single-digit numbers, number words, and non-symbolic numbers, including counting. The second subcategory included eight studies (5%) that used eye tracking to examine how participants represent and process multi-digit numbers. The third subcategory included 18 studies (11%) where eye tracking was used to investigate spatial-numerical associations (SNAs; Dehaene, Bossini, & Giraux, 1993) in the context of the mental representation of numbers and magnitude (e.g., the mental number line) and mathematical operations (e.g., the operational momentum effect; Klein et al., 2014). SNAs are assumed to emerge in a variety of numerical cognitive processes (for a critical review, see Shaki & Fischer, 2018). The fourth subcategory included nine studies (6%) that analyzed performance and strategic processes on a number line estimation task. In a fifth subcategory, 23 studies (14%) examined cognitive processes during basic arithmetic operations with Arabic numerals. For example, studies in this subcategory focused on eye movements during equation-solving (e.g., Verschaffel, De Corte, Gielen, & Struyf, 1994) or more complex arithmetic operations like multiplication (e.g., Ganor-Stern & Weiss, 2016). The sixth subcategory included five studies (3%) focusing on the development of number perception and number magnitude by using eye tracking with infants. Finally, the seventh subcategory involved the processing of rational numbers and proportionality, which was composed of 12 studies (7%). Most of these studies analyzed how people process the numerical values of rational numbers, predominantly fractions, although Plummer et al. (2017) included decimals and visual (rather than symbolic) representations of rational numbers.

Geometry, shape, and form The close connection between geometrical thinking and visual perception makes eye tracking a suitable and pertinent method for research that analyzes the perception and processing of geometrical objects. In total, 22 studies in this review (14%) analyzed tasks or abilities in this area. This included geometric proofs, analyzing mental rotation tasks, construction of geometric objects, the perception of objects in a Cartesian plane, geometric calculations in dynamic problems, and the processing of vector fields and geometric shapes.
### Table 1: Domains and topics of mathematics education addressed by the studies included in the review

| Domain and topic | Total |
|------------------|-------|
| **Numbers and arithmetic** | 90 |
| Perception and mental representation of single digit and non-symbolic numbers, number words, counting | 16 |
| Abreu-Mendoza & Arias-Trejo, 2015 | Reike & Schwarz, 2016 |
| Gandini et al., 2008 | Rinaldi et al., 2015 |
| Gielen et al., 1991 | Schiefer & Landef, 2011 |
| Li et al., 2010 | Sophian & Crosby, 2008 |
| Merkley & Ansari, 2010 | van Oeifelen & Vos, 1984a |
| Milosavjevic et al., 2011 | van Oeifelen & Vos, 1984b |
| Moeller, Neuburger, et al., 2009 | Watson et al., 2005 |
| Price et al., 2017 | Watson et al., 2007 |
| Perception and mental representation of multi-digit numbers | 9 |
| Bahnmueller et al., 2016 | Meyerhoff et al., 2012 |
| Brysbaert, 1995 | Moeller, Fischer, et al., 2009a |
| Hintz & Meyer, 2015 | Moeller, Fischer, et al., 2009b |
| Huber et al., 2015 | Schindler & Lilienthal, 2018 |
| Huber, Mann, et al., 2014 | |
| **Spatial-numerical associations** | 18 |
| Bulf et al., 2014 | Masson, Letesson, & Pesenti, 2018 |
| Bulf et al., 2016 | Myachykova et al., 2015 |
| Fischer et al., 2004 | Myachykova et al., 2016 |
| Hartmann, Mast, & Fischer, 2016 | Pressigout et al., 2018 |
| Holmes et al., 2016 | Risko et al., 2013 |
| Klein et al., 2014 | Ruiz Fernandez et al., 2011 |
| Loetscher et al., 2008 | Salvaggio et al., 2018 |
| Loetscher et al., 2010 | Schwarz & Keus, 2004 |
| Macchi Cassia et al., 2016 | Zhu et al., 2018 |
| Number line estimation task | 9 |
| Gomez et al., 2017 | Sullivan et al., 2011 |
| Heine et al., 2010 | van der Weijden et al., 2018 |
| Klein et al., 2014 | van Viersen et al., 2013 |
| Reinert et al., 2015 | van’t Noordende et al., 2016 |
| Schneider et al., 2008 | |
| Domain and topic                                      | References                                                                 |
|------------------------------------------------------|-----------------------------------------------------------------------------|
| Basic arithmetic                                     | Bolden et al., 2015                                                        |
|                                                      | Chesney et al., 2013                                                        |
|                                                      | Curtis et al., 2016                                                         |
|                                                      | Evans et al., 2011                                                          |
|                                                      | Ganor-Stem & Weiss, 2016                                                    |
|                                                      | Godau, Haider et al., 2014                                                  |
|                                                      | Godau, Wirth et al., 2014                                                   |
|                                                      | Green et al., 2007                                                          |
|                                                      | Hamada & Iwaki, 2012                                                        |
|                                                      | Hartmann et al., 2015                                                       |
|                                                      | Hartmann, Laubrock, & Fischer, 2016                                        |
|                                                      | Huebner & LeFevre, 2018                                                    |
|                                                      | Moeller et al., 2011a                                                       |
|                                                      | Moeller et al., 2011b                                                       |
|                                                      | Okamoto & Kuroda, 2014                                                     |
|                                                      | Potgieter & Blignaut, 2018                                                  |
|                                                      | Schneider et al., 2012                                                      |
|                                                      | Suppes et al., 1982                                                        |
|                                                      | Suppes et al., 1983                                                        |
|                                                      | Susac et al., 2014                                                          |
|                                                      | Verschaffel et al., 1994                                                    |
|                                                      | Winoto et al., 2017                                                         |
|                                                      | Zhou et al., 2012                                                           |
| Number perception and magnitude in infants           | Bremner et al., 2017                                                        |
|                                                      | Bulf et al., 2016                                                           |
|                                                      | Canfield & Smith, 1996                                                     |
|                                                      | Ceulemans et al., 2014                                                     |
|                                                      | Lécuyer et al., 2004                                                        |
| Rational numbers and proportionality                 | Abrahamson et al., 2016                                                     |
|                                                      | Atagi et al., 2016                                                           |
|                                                      | Huber, Klein et al., 2014                                                   |
|                                                      | Huber, Moeller, & Nuerk, 2014                                               |
|                                                      | Hurst & Cordes, 2016                                                        |
|                                                      | Ischebeck et al., 2016                                                      |
|                                                      | Miller Singley & Bunge, 2018                                                |
|                                                      | Obersteiner & Staudinger, 2018                                              |
|                                                      | Obersteiner & Tumpek, 2016                                                  |
|                                                      | Obersteiner et al., 2014                                                    |
|                                                      | Plummer et al., 2017                                                        |
|                                                      | Wall et al., 2015                                                           |
| Geometry, shape, and form                            | Alqassab et al., 2018                                                       |
|                                                      | Lin et al., 2012                                                            |
|                                                      | Muldner & Burleson, 2015                                                    |
|                                                      | Schindler et al., 2016                                                      |
| Mental rotation                                      | Chen & Yang, 2014                                                           |
|                                                      | Fry, 1988                                                                   |
|                                                      | Merschmeyer-Brüwer, 2001                                                   |
|                                                      | Roach et al., 2016                                                          |
|                                                      | Wang et al., 2014                                                           |
| Domain and topic                        | Authors                                      | Total |
|----------------------------------------|----------------------------------------------|-------|
| Construction                           | Epelboim & Suppes, 2001                      | 5     |
|                                        | Hannula & Williams, 2016                     |       |
|                                        | Lee & Wu, 2018                               |       |
|                                        | Schimpf & Spannagel, 2011                    |       |
|                                        | Shvarts, 2018a                               |       |
| Cartesian plane                         | Krichevets et al., 2014                      | 2     |
| Calculation                             | Lin & Lin, 2014a                             | 2     |
| Vector fields                           | Klein et al., 2018                           | 2     |
| Shapes                                  | Haataja et al., 2018                         | 2     |
| Reading and word problem solving        | De Corte & Verschaffel, 1986a                | 9     |
| Prototype word problems                 | Hegarty et al., 1992                         |       |
|                                        | Hegarty et al., 1995                         |       |
|                                        | van der Schoot et al., 2009                  |       |
|                                        | Verschaffel et al., 1992                     |       |
| Complex mathematical text or word       | Cimen & Campbell, 2012                       | 7     |
| problems                                | Lee & Wu, 2018                               |       |
|                                        | Sajka & Rosiek, 2015                         |       |
|                                        | Stolinska et al., 2014                       |       |
| Mathematical parsing                    | Moutsios-Rentzos & Stamatis, 2013            | 5     |
|                                        | Kohlhase et al., 2018                        |       |
| Reasoning and Proof                     | Hods et al., 2014                            | 14    |
| Reading and comprehension of proofs     | Inglis & Alcock, 2012                        |       |
|                                        | Roy et al., 2017                             |       |
|                                        | Shayan et al., 2017                          |       |
| Proportional reasoning                  | Duijzer et al., 2017                         | 2     |
| Domain and topic                  | Cohen & Staub, 2015 | Fleig et al., 2017 | Total |
|----------------------------------|---------------------|-------------------|-------|
| Probabilistic reasoning          |                     |                   | 2     |
| Logic                            | Espino et al., 2005 | Ott et al., 2018  | 3     |
| Kim et al., 2018                 |                     |                   |       |
| Functional thinking              | Cohen-Fresenborg et al., 2010 | Crisp et al., 2011 | 2     |
| Use of representations           | Andrà et al., 2009  | Clinton et al., 2016 | 14    |
| Andrà et al., 2015               |                     | Dewolf et al., 2015 |       |
| Atagi et al., 2016               |                     | Lee & Wu, 2018    |       |
| Beitlich et al., 2014            |                     | Michal et al., 2016 |       |
| Beitlich et al., 2015            |                     | Ott et al., 2018  |       |
| Bolden et al., 2015              |                     | Rozek et al., 2014 |       |
| Boels et al., 2018               |                     | Schindler & Lilienthal, 2018 |       |
| Learning difficulties            | Abreu-Mendoza & Arias-Trejo, 2015 | van der Weijden et al., 2018 | 8     |
| Gomez et al., 2017               |                     | van Versen et al., 2013 |       |
| Moeller, Neuburger, et al., 2009 |                     | van’t Noordende et al., 2016 |       |
| Schindler & Lilienthal, 2018    |                     | Winoto et al., 2017 |       |
| Computer-supported learning      | Hernandez-Sabate et al., 2016 | Roy et al., 2017  | 6     |
| Kiili et al., 2014               |                     | Schimpf & Spannagel, 2011 |       |
| Olsen et al., 2017               |                     | Yağmur & Çakır, 2016 |       |
| Mathematical problem solving     | Andrzejewska & Stolinska, 2016 | Knoblich et al., 2001 | 4     |
| Haataja et al., 2018             |                     | Werner & Raab, 2014 |       |
| Statistics                       | Boels et al., 2018  | Fleig et al., 2017 | 3     |
| Cohen & Staub, 2015              |                     |                   |       |
| Affective variables              | Hunt et al., 2015   | Strohmaier et al., 2017 | 2     |

Studies may be assigned to more than one category. Categories are set in **bold**; subcategories are set regular.
Reading and word problem solving Studies on mathematical reading and word problem solving comprised 21 (13%) of the studies reviewed here. One of the oldest and most common uses of eye tracking is in research on reading (Rayner, 1998; Rayner, Pollatsek, Ashby, & Clifton, 2012). Consequently, the first studies using eye tracking in mathematics education focused on reading of mathematical texts or word problems. Terry (1921) analyzed fundamental characteristics of reading in mathematics, which he related to the first models of eye movements during reading. This association was also investigated by nine studies that analyzed eye movements during prototype word problem solving. Apart from Terry (1921), six studies considered more complex word problems and longer passages of mathematical text, for example, the integration of illustrations (Dewolf et al., 2015), formulae (Kohlhase et al., 2018), or figures (Lee & Wu, 2018). In addition, basic processes in decoding and parsing of mathematical language were analyzed in five studies.

Reasoning and proof Fourteen studies (9%) addressed mathematical reasoning and proof. This included studies investigating eye-movement patterns during the reading and validation of proofs or during proportional reasoning. Furthermore, studies in this category addressed probabilistic reasoning, mathematical logic, and functional thinking. Note that studies on reasoning and proof in geometry were included in the category geometry.

Use of mathematical representations In general, the role of representations and multimedia in learning is subject to numerous studies in research using eye tracking (see Hyönä, 2010; Lai et al., 2013; Mayer, 2010; van Gog & Scheiter, 2010, for overviews). With respect to the present review, 14 (9%) of the studies investigated the role of mathematical representations. In mathematics, information is commonly represented in three different ways: formulae, graphics, and texts. Using eye tracking, Andrà et al. (2009) and Andrà et al. (2015) showed that these representations differ fundamentally in the way that information is retrieved. Several studies analyzed differences in presentation format in various domains: fractions (Atagi et al., 2016), argumentation and proof (Beitlich et al., 2014; Beitlich et al., 2015), propositional logic (Ott et al., 2018), problem solving (Rozek et al., 2014), multiplication (Bolden et al., 2015), word problem solving (Dewolf et al., 2015), and geometry (Lee & Wu, 2018).

Learning difficulties Eye tracking has been used to characterize and analyze learning disabilities in mathematics. These eight studies (5%) compared typically developing students and students with dyscalculia (Moeller, Neuburger, et al., 2009; van der Weijden et al., 2018; van Viersen et al., 2013), Down syndrome (Abreu-Mendoza & Arias-Trejo, 2015), autism (Winoto et al., 2017), developmental coordination disorder (Gomez et al., 2017), and general mathematical learning difficulties (Schindler & Lilienthal, 2018; van’t Noordende et al., 2016).

Computer-supported learning Six studies (4%) used eye tracking to analyze learning processes in a computer-supported learning environment. Kiili et al. (2014) and Hernandez-Sabate et al. (2016) used eye tracking to analyze students’ attention to different features of educational mathematics games. Other studies evaluated the usability and effectiveness of learning software, namely, Cinderella (Schimpf & Spannagel, 2011), e-Proof (Roy et al., 2017), and GeoGebra (Yağmur & Çakır, 2016). Olsen et al. (2017) evaluated a collaborative tutoring system using eye tracking by examining joint visual attention.
Mathematical problem solving  Mathematical problem solving was examined in four studies (2%), investigating relations between eye movements and objective/subjective task difficulty (Andrzejewska & Stolinska, 2016), insight problem solving (Knoblich et al., 2001), and the association of body movements and problem solving processes, assuming effects of embodied cognition (Werner & Raab, 2014). Haataja et al. (2018) used a collaborative problem-solving task to investigate a teacher’s attention during scaffolding.

Statistics  Three studies (2%) focused on statistics. Each study analyzed the interpretation of statistical data, addressing misconceptions in contingency (Fleig et al., 2017) and Bayesian reasoning (Cohen & Staub, 2015), and difficulties in interpreting histograms (Boels et al., 2018).

Affective variables  Affect reflects an important aspect of learning in general and of mathematics education in particular (e.g., Goldin et al., 2016). Two studies (1%) related eye movements to mathematics-specific affective variables, namely mathematical self-concept (Strohmaier et al., 2017) and mathematics anxiety (Hunt et al., 2015).

4.3 Apparatus

Most of the recent studies (since 2014) in our review used equipment from one of three manufacturers (see Holmqvist et al., 2011 for an introduction of the manufacturers): SR Research (Ottawa, Canada; 33%), Tobii (Danderyd, Sweden; 29%), or SMI (Potsdam, Germany; 25%). SR Research models have typically provided a higher sampling rate of 500 to 1000 Hz, which might be necessary to compare temporal measures and reaction times.

Eye trackers are commonly distinguished in two categories (Holmqvist et al., 2011). Static systems are attached to the stimulus, which is typically presented on a computer monitor. The second category are head-mounted devices, where the eye tracker is attached to a head mount or integrated in eye-tracking glasses. Head mounted systems can record data on portable devices like mobile phones. When they do not require a physical connection to a static recording device, we refer to these systems as mobile. In this review, 132 studies used a static eye tracker, 19 used head-mounted systems, 8 of which were mobile.

Depending on the set-up, the accuracy of the apparatus varies notably (Holmqvist, Nyström, & Mulvey, 2012). Using current technology, a set-up that allows for free head movements typically provides accuracies of 0.5° to 1° of visual angle. Accuracy can be increased to about 0.1° by fixating participants’ heads, for example, with a chin rest. However, depending on the design of the stimulus, such precision might not be necessary as the area of good visual acuity extends to about 2° of visual angle (Rayner, 1998). Only 24% of the studies in this review explicitly reported that participants’ heads were free-moving. In 37% of the studies, a head rest, head frame, chin rest, or bite bar was reported or could be inferred from the apparatus type. However, in many studies (39%), no information regarding movement restrictions was reported or could be inferred.

There are a number of different techniques to record eye movements (see Duchowski, 2007, for an overview). The oldest study reported here (Terry, 1921) used the reflection of an ordinary light beam onto a spooling film roll to record eye movements in one dimension (details of this apparatus are reported in Gray, 1917).
the present review, 151 (94%) of the studies used a video-based pupil/corneal reflection technique. For this technique, one or more cameras record an image of the eye with an infrared light source located next to the camera. Two reference points from the recorded image are compared to determine the orientation of the eye. The first reference point is the reflection of the infrared light source on the cornea, and the second reference point is the center of the pupil. The positional difference between these reference points changes with the rotation of the eye and can be translated to the gaze position (see Duchowski, 2007, for a detailed description).

In four studies, eye movements were manually coded by a human observer during the experiment (Moutsios-Rentzos & Stamatis, 2013, 2015), from video recordings (Canfield & Smith, 1996) or both (Macchi Cassia et al., 2016). Three studies used a search coil technique, which determines the gaze position by a contact lens containing a thin wire, which is then detected by a magnetic field surrounding the participant (Duchowski, 2007). A further two studies used electrooculography (EOG), which makes use of a natural, small electric potential between the front and the back of the human eye to detect its movements (Hamada & Iwaki, 2012; Zhou et al., 2012). EOG and search coils are usually used to increase accuracy, while qualitative observation of eye movements is especially advantageous with infants and in authentic learning environments. For one study, it was not possible to identify the eye-tracking technique (Fry, 1988).

Most studies reported details about the stimulus presentation (e.g., monitor size, resolution, refresh rate, or viewing distance). However, information about the eye-tracking apparatus itself was often incomplete: 24% of the studies did not report the sampling rate and 29% did not mention a calibration procedure and if they did, only 12% of these studies reported the accuracy required for accepting a calibration. This information is crucial for interpreting the data since the accuracy of the apparatus and a typical calibration error can easily amount to errors of up to 2° of vision, which corresponds to about 2.5 cm on the screen in a regular, static set-up. Depending on the stimulus design and the required precision, this can negatively affect data quality.

4.4 Stimuli

The design of the stimuli was usually characterized by a trade-off between methodological considerations and authenticity. Some studies used simplified stimuli containing few elements and took technical details, such as background color, font, or the synchronization of sampling rates of the stimulus monitor and the eye tracker into careful consideration. Other studies, in contrast, focused on educational authenticity and were more concerned with creating stimuli that were as close to real-world or school mathematics as possible. This included the implementation of dynamic stimuli (e.g., Canfield & Smith, 1996; Schimpf & Spannagel, 2011; Shayan et al., 2017), tablet computers (Abrahamson et al., 2016; Shayan et al., 2017; yağmur & Çakır, 2016), material from textbooks (Beitlich et al., 2014; Kohlhase et al., 2018; Lee & Wu, 2018), or pictures (e.g., Dewolf et al., 2015). Studies using head-mounted eye tracking sometimes used realistic, paper-based stimuli (Schindler et al., 2016) or authentic classroom interaction (Haataja et al., 2018; Hannula & Williams, 2016). However, these studies needed to take into account that head-mounted eye tracking typically provides a lower resolution and sampling rate than static eye trackers. Moreover, because the eye-tracking device is not attached to the screen with the stimulus, the gaze coordinates have to be mapped onto the stimulus using a second camera which records the participant’s field of view. This requires a significant amount of computation and adds additional sources of inaccuracy.
For analyses in 101 papers, *Areas of Interest* (AOIs) were defined. These are predefined areas of the stimulus that are used for analyzing eye movements on specific elements of the stimulus, such as representations, target words, or keywords (Holmqvist et al., 2011). Importantly, the size of the AOIs influences many eye-tracking measures. For example, the number of fixations will typically increase in larger AOIs (Holmqvist et al., 2011). Many of the studies that made use of AOIs (49%) avoided this issue by using AOIs that had the same size. However, AOIs with the same size do not necessarily contain the same amount of relevant information. In other cases, AOI data were not compared with each other but between trials or participants (e.g., Bolden et al., 2015; Roy et al., 2017). If a standardization was necessary, for example, to compare eye movements on pictures and text that did not have the same size, some studies standardized the AOI sizes by their area (e.g., Alqassab et al., 2018; Beitlich et al., 2015) or used measures that were less strongly affected by AOI dimensions (e.g., revisits, Hegarty et al., 1995; time to first fixation, Bulf et al., 2014).

### 4.5 Sample and research design

All studies included in this review, except for three (Olsen et al., 2017; Shvarts, 2018a, 2018b), tested participants individually. The 161 studies included a total of 189 experiments. On average, each experiment included a sample of $M = 28.56$ ($SD = 21.70$) participants. The majority of studies (59%) included participants from tertiary education, while only 28% included participants form primary or secondary schools.

Of the studies reviewed here, 54% used a within-subject design or a mixed design. Studies that used between-subject designs (22%) usually compared specific populations, for example, a particular age, achievement, or expertise.

### 4.6 Data treatment

Eye-tracking instruments provide raw data about eye movements, usually in the form of coordinates. Since perception depends on the specific nature of the eye movements (Matin, 1974), it is theoretically important for many research questions to manually (by inspecting visualizations of raw data) or automatically (through an automated algorithm) categorize the raw data into events, typically fixations, saccades, and blinks. During saccades and blinks, almost no information is processed, but these events can account for 5 to 15% of the raw data (for typical event durations, see Holmqvist et al., 2011). There are several established automatic event detection algorithms used in eye-tracking research. In most eye-tracking analysis software, these algorithms are applied automatically, but thresholds can be modified. The selection of both the algorithm and the thresholds influences the filtered data and thus potentially the results of a study (Blignaut, 2009). Therefore, they should be reported in publications. However, 60% of the studies included in this review did not report an event detection algorithm, and only 21% reported thresholds.

Data loss is a critical issue in studies using eye tracking (Holmqvist et al., 2011) and was substantial in many studies included in this review. When studies reported participant

---

5 The two most popular algorithms use a dispersion threshold to cluster raw data points to fixations (I-DT) or a velocity or acceleration threshold to detect saccades (I-VT; Salvucci & Goldberg, 2000). Once fixations are detected, saccades can then be defined as movements between fixations, and vice versa.
exclusion, this affected an average of 15% of the total sample size. Data loss seems to be especially common in studies with very young or old participants. Lécuyer et al. (2004) used data only from 12 of their 50 four-month-old infants (a data loss of 76%). Ischebeck et al. (2016) had to exclude 21% of their 6-year-old participants, but only 3% of their 8-year-olds in the same experiment. Similarly, Watson et al. (2005) lost about twice as much data in their group of 57- to 79-year-old adults compared with their group of 18- to 25-year-old students.

In a majority of studies reviewed here (76%), eye tracking was not the only source of data in the experiments. In half of these studies, other data were analyzed in relation to eye movements (50%), while the other half analyzed the additional data independently. Frequently, measures like reaction time and accuracy were analyzed in parallel. In general, if data types were analyzed in relation to each other, it was often the relation between eye movements and accuracy that was analyzed. Richer triangulations were done with gestures and communication (Hannula & Williams, 2016; Shvarts, 2018a, 2018b), interviews and stimulated recall (Klein et al., 2018; Shayan et al., 2017; see also Schindler & Lilienthal, 2019), think-aloud protocols and self-reports (Cimen & Campbell, 2012; Green et al., 2007; Ögren et al., 2017; Schindler & Lilienthal, 2018), cognitive load (Lin & Lin, 2014b), affective variables (Hunt et al., 2015; Strohmaier et al., 2017), or skin conductance and EEG (Muldner & Burleson, 2015).

4.7 Interpretation of eye movements

Eye tracking offers possibilities for qualitative and quantitative research. Of the reviewed studies, 66% used a quantitative approach, 22% used a qualitative approach, and 11% used a combination of qualitative and quantitative approaches.

Interpreting eye-tracking data is not straightforward. Lai et al. (2013) and Holmqvist et al. (2011) provide an overview of the most common measures of eye movements in educational research as well as the theoretical interpretations of these measures. Crucially, eye-tracking

| Table 2 | Interpretations of eye movements used by the studies included in the review |
|---------|--------------------------------------------------------------------------------|
| Interpretation of eye movements | Number of studies |
| Visual focus and overt attention, eye mind assumption | 96 |
| Attention on single object | 6 |
| Compare attention on 2 or more objects | 28 |
| Order of attention (quantitative) | 40 |
| Order of attention (qualitative) | 23 |
| Mental representation and covert attention | 39 |
| Visual mental representation | 13 |
| Motoric mental representation | 4 |
| Brain hemisphericity | 2 |
| Areas of covert attention | 3 |
| Attentional anchors | 4 |
| Reaction time | 9 |
| Anticipation | 4 |
| Cognitive effort and resources | 28 |
| Information extraction | 25 |
| Memory capacity | 3 |
| Joint attention | 3 |
| Metacognitive control | 2 |
| Explorative | 2 |

Studies may be assigned to more than one category
measures can be interpreted in various ways. In many cases, the same measure can indicate different cognitive processes (Holmqvist et al., 2011). Therefore, this section is structured by the interpretations of eye movements suggested in the reviewed studies, which are listed in Table 2. In this section, we give the most common measures of eye movements that were used for each interpretation and exemplary studies for each measure. For a list of the measures used in each study and their associated interpretations, see Appendix Table 3.

**Visual focus and overt attention, eye mind assumption** Although the studies presented here are diverse in their research interests, the link between eye movements and cognitive processes in mathematics education often builds on similar theoretical considerations. The most common interpretation of eye movements is based on the eye mind assumption (EMA) formulated by Just and Carpenter (1980). It initially stated that “the eye remains fixated on a word as long as the word is being processed” (p. 330) and has since been interpreted as a more general rule of visual focus equals cognitive focus. Recent research showed that this assumption is a strong simplification and indicated that it does not hold rigorously (for a detailed review of research investigating the association between vision and attention, see Carrasco, 2011; see also Duchowski, 2007; Schindler & Lilienthal, 2019). In general, only so-called overt attention can be directly observed through the position of the visual focus. However, attention can also be shifted without moving the eyes, which is referred to as covert attention (Carrasco, 2011). Even though overt and covert attention do overlap in the majority of cases (Carrasco, 2011), Schindler and Lilienthal (2019) showed that eye movements and self-reports of the attentional focus often diverge, which indicates that the EMA might not hold under certain circumstances. While ambiguity may not be a reason to discard the assumption entirely, it calls for researchers to be aware of the limits of its interpretation. In the studies reviewed here, the majority (60%) interpreted eye movements in accordance with the EMA.

A fundamental question that can be addressed by analyzing eye movements is when, whether, and how much a single aspect of a visual stimulus is attended to by a participant (4% of the studies interpret eye movements this way). The onset of attention can be measured, for example, by the time to first fixation (Schimpf & Spannagel, 2011) and the first fixation position (Ruiz Fernández et al., 2011). It should be noted that these measures can also indicate covert attention indirectly, since it is usually the driving force behind the initiation of fixations (Rayner et al., 2012). The amount of attention to objects or areas of the stimulus is often measured by the number of fixations on the object (Dewolf et al., 2015) or by inspecting visualizations such as heat maps or scan paths (also referred to as gaze paths, Winoto et al., 2017; scan patterns, or gaze sequences, Holmqvist et al., 2011). For example, Dewolf et al. (2015) evaluated whether students attended to illustrations in word problem solving by measuring the number of fixations on the illustrations. This approach is also popular in evaluating computer-supported learning, since eye tracking can be used to assess whether students notice certain elements of the learning environment (Kiili et al., 2014; Schimpf & Spannagel, 2011). Critically, attention to a certain element can have two fundamentally different causes: Lin & Lin (2014b) argue that high-performing students spend less time and fixations on important areas because they extract the information faster (see also Gegenfurtner, Lehtinen, & Säljö, 2011). At the same time, other researchers found that experts can better differentiate between relevant and irrelevant information and therefore spend relatively more time on relevant than irrelevant information (e.g., Fleig et al., 2017; Kim et al., 2018).

Similar to examining single aspects in a visual stimulus, and based on the EMA, the attention allocated to different parts of a visual stimulus can be compared. Out of the studies included in this review, 17% use this interpretation. The most popular measures for this
comparison are the relative number of fixations and the relative fixation duration. For example, Ott et al. (2018) compared the number of fixations and total fixation duration between text and formulae and text and graphics, respectively. Furthermore, measures can be compared between different periods of time: De Corte et al. (1990) compared fixation durations on numbers and words between a participant reading the word problem for the first time and the consecutive reading process.

In addition, 25% of the reviewed studies analyzed attentional patterns of eye movements, which are often associated with solution strategies. For example, the number and order of transitions between certain aspects of a stimulus was used to assess parallel compared with sequential strategies in number processing (Merkley & Ansari, 2010; Meyerhoff et al., 2012), fraction comparison strategies (Miller Singley & Bunge, 2018; Obersteiner et al., 2014; Obersteiner & Tumpek, 2016), or information integration processes (Alqassab et al., 2018; Crisp et al., 2011; Ögren et al., 2017). As another measure, the position of the first fixation was considered an indicator for a preferred order of information processing (Michal et al., 2016). Moreover, saccade length was used as an indicator for local (short saccades) compared to global (long saccades) strategies in information retrieval (Inglis & Alcock, 2012; Klein et al., 2018; Stolinska et al., 2014) and information integration (Godau, Wirth, et al., 2014).

When a quantitative description of strategies was not possible, a number of studies (14%) used analyses based on visualizations like scatterplots (Inglis & Alcock, 2018), heat maps, scan paths, or replays of the eye-tracking recording to manually detect patterns of eye movements and associated strategies (e.g., Lee & Wu, 2018). This approach was often limited to smaller sample sizes as it requires more time than computerized analyses.

**Mental representation and covert attention** Other than eye-tracking measures based on the EMA, 24% of the studies used eye movements as more direct indicators of mental representations (e.g., Klein et al., 2014) and processes (e.g., Hamada & Iwaki, 2012). This approach was found especially in studies on SNAs and number line tasks, where the position of fixations was assumed to map onto a mental, spatial representation of numbers (8%). Additionally, nine studies used the time to the first fixation on a target as a measure of reaction time. By placing the targets in different areas of the visual field, SNAs could be observed (e.g., Schwarz & Keus, 2004). Moreover, some studies associated imperceptible eye movements like microsaccades as a motoric indicator of cognitive representations (four studies; e.g., Myachykov et al., 2016).

Another interpretation was the association between eye movements and brain hemisphericity during mathematical tasks, proposed in two studies (Moutsios-Rentzos & Stamatis, 2013, 2015).

Other studies investigated areas of covert attention, i.e., when the focus of attention is not equivalent to the visual focus, through eye movements. For example, the position of the first fixation on a stimulus was interpreted to map on the previous focus of covert attention (three studies; e.g., Risko et al., 2013). *Attentional anchors*, which are goal-oriented perceptual structures in the sensory field that enable better coupling with the environment, were located through gaze patterns (four studies; e.g., Duijzer et al., 2017). For infants, the gaze time and saccade latency were interpreted as indicators for the mental anticipation of objects (four studies; e.g., Canfield & Smith, 1996).

**Cognitive effort and resources** A final set of interpretations of eye movements was based on the assumption that eye movements are direct indicators of the cognitive effort involved in decoding visual information. This more fundamental interpretation was used by 17% of the reviewed studies. For instance, the mean duration of fixations was used as an indicator for mental workload or cognitive effort and depth in processing information (e.g., Hodds et al., 2014). Moreover, the
number of revisits on specific elements, the total fixation duration, or the time to the first fixation were interpreted as an indication of memory capacity (three studies; e.g., Watson et al., 2005).

**Joint attention, metacognitive control** The three studies making use of dual eye tracking (i.e., two participants’ eye movements being tracked in parallel) used the proximity of learners’ fixations as a measure of joint attention (e.g., Shvarts, 2018a). Two studies associated eye movements with metacognition, using the blink rate (Cimen & Campbell, 2012) as well as scan paths and the total fixation duration (Cohors-Fresenborg et al., 2010).

### 5 Discussion

This review investigated the use of eye tracking in the field of mathematics education research, addressing three research questions. We discuss the results in the order of these questions.

#### 5.1 Overview of the use of eye tracking in mathematics education

In accordance with our first research question, we provided an overview of the domains and topics addressed in mathematics education research using eye tracking, and when and how these studies were published. Our findings illustrated that the number of studies in mathematics education that made use of eye tracking has increased rapidly in the last decade and continues at a rate of around 20 studies published per year. This illustrates the ongoing popularity and importance of eye tracking within mathematics education. Our results showed that eye tracking was used in a wide range of fields within mathematics education, although a majority of studies focused on numbers and arithmetic and fundamental processes of mathematical thinking like number perception, counting, and basic arithmetic operations. These studies were often conducted in controlled and systematically designed laboratory experiments and aimed to precisely assess specific cognitive processes that were often impossible to investigate through other methods. However, it is noteworthy that our review also included a variety of studies that went beyond strictly controlled laboratory settings to include authentic learning situations (e.g., Hannula & Williams, 2016; Kiili et al., 2014; Lin & Lin, 2014b). Here, eye tracking benefited twofold; it allowed for relatively authentic learning environments, and it was an unobtrusive method to gather data about learning processes. In sum, it seems clear that the possibilities offered by the method of eye tracking are diverse in the domain of mathematics education, with this method being adaptable to the various subdomains. Recent developments like mobile or dual eye tracking indicate that the method will continue to evolve in the future, particularly in the direction of more authentic experimental settings. However, the domains found in our review did not cover all mathematical topics in the same depth. Thus, a wealth of opportunity remains for future studies where eye movements could be informative, for example, in the field of statistics, where only three studies were found (see Boels, Bakker, & Drijvers, 2019, for recent developments).

It should be noted that the scope of this first research question was not to summarize the specific research goals of each study but rather to provide an overview of the topics addressed within these goals. Accordingly, our review analyzes the research from a methodological perspective and does not scrutinize the specific results and implications that each study offers in its respective field. We acknowledge that this is a narrow focus that omits findings and consequences of these studies. However, we hope that the overview of domains and topics will guide the reader to a more detailed investigation of the consequences of the specific studies.
5.2 Eye-tracking methodology

To answer our second research question, we reviewed technical and methodological aspects of eye-tracking studies in mathematics education. In terms of research design, most studies used within-subject or mixed designs, which allowed for small sample sizes with large interindividual variance. At the same time, many studies revealed difficulties due to a substantial loss of data, which might be problematic if sample sizes are already small. As for the age of participants, most studies included university students who arguably often represent a convenience sample. Because many research questions in mathematics education are related to school-aged children, the predominance of university students in eye-tracking studies in mathematics education seems problematic. Age is an important factor affecting eye movements (Holmqvist et al., 2011; Rayner et al., 2012) meaning that the generalizability of findings from adults might be limited. Moreover, university students typically reflect a high-achieving selection of young adults. For these reasons, the issue of generalizability should be acknowledged and systematically addressed in future research.

In addition to research design and participant characteristics, we examined the use of the eye-tracking method in the reviewed studies. Although studies necessarily vary in the specific eye-tracking method they use, we found large inconsistencies in the reporting of these methods. Because eye-tracking research offers a variety of options in terms of the apparatus itself, the settings of the apparatus, and the specific data analysis methods, it seems especially important to report all necessary details with regard to eye tracking (see also Holmqvist et al., 2011). Full reporting is also crucial for other researchers to understand or replicate the specific analysis of eye movements and to evaluate the implications of the findings. We propose that this should include (but not be limited to) a precise description of the apparatus including sampling rate and average accuracy; the existence or nonexistence of movement restrictions and information regarding the setup; the size of the stimuli; the distance between the stimuli and the participant; the monitor’s refresh rate; the calibration procedure and calibration accuracy threshold; the event detection algorithm and event detection thresholds; the position and size of any AOIs; the correlation between all used measures; and the amount of and reasons for data loss.

5.3 Interpretation of eye movements

Finally, our third research question aimed to assess how eye-tracking data were interpreted in the reviewed studies. It became clear that eye tracking not only offers a wide range of possibilities for qualitative and explorative analyses but also provides data suitable for various quantitative analyses. In general, most eye-tracking studies in mathematics education claimed that the method allows for the assessment of cognitive processes that would otherwise not be observable, for example, because they are subconscious. One of the most crucial challenges in eye-tracking research is to properly link eye movements to these assumed underlying cognitive processes. Although reflecting about this link seems obvious, it is by no means ubiquitous in the studies reviewed in this paper. Even when studies analyzed similar cognitive processes, they often made use of numerous or redundant measures of eye movements. Eye-tracking measures can be highly correlated with each other for theoretical or computational reasons, but correlations are scarcely reported. As an example, Merkley and Ansari (2010) reported the correlation between the number of fixations and the number of saccades, which was \( r = .997 \). This is not surprising as saccades and fixations alternate in regular reading and event detection algorithms typically infer one from the other (Salvucci & Goldberg, 2000). In such a case, the two measures arguably cannot reflect
different cognitive processes and researchers should consider using only one of the two measures for analyzing the same research question (e.g., Hurst & Cordes, 2016).

Eye movements alone are seldom informative when the reader is not properly guided on their meaning and relation to the research questions. Thus, eye-tracking studies—as all other experiments—arguably require a plan on data interpretation before they are conducted, and possible implications gained from the data should be anticipated in advance, even in explorative studies. Apart from avoiding the risk of type I errors and post hoc hypotheses, limiting and specifying the to-be-used measures further helps to clarify, compare, and interpret the results (for a discussion of some of these considerations, see Banks et al., 2019).

We found that a majority of studies referred to the EMA as a theoretical foundation for the interpretation of eye movements. The EMA itself is, however, not universally accepted in eye-tracking literature, and there are limitations within the domain of mathematics education as well (Anderson, Bothell, & Douglass, 2004; Rayner, 1998; Schindler & Lilienthal, 2019). Schindler and Lilienthal (2019) addressed the case of a student solving a geometry problem. Using stimulated recall, the authors identified incidents in which the EMA seemed to hold—and others in which the EMA was seemingly violated. Even in incidents in which the EMA did hold, the same eye-movement pattern could have been linked to a variety of cognitive processes. The EMA is arguably a feasible tool for the interpretation of a majority of eye movement data, especially in settings in which participants have to work on visual problems within a limited time. But its suitability should be evaluated from case to case. Triangulation with other data sources could help to interpret eye movements meaningfully.

Our review shows that eye tracking was not limited to detecting the visual focus of attention following the EMA. Rather, eye tracking has also been used to assess, among other things, mental representations, cognitive workload, or joint attention in collaborative learning settings. For these processes, eye tracking provides specific benefits: For example, the relationship between mental representations and eye movements is often very strong and can be best observed during the work on the task and not after the task (e.g., during counting; Hartmann, Mast, & Fischer, 2016).

The amount of different measures that were used in the studies reviewed here illustrates not only the many possibilities that eye tracking provides but also the possible difficulties in comparing studies with regard to the measures used and their interpretation. It would be helpful if future research would make stronger use of previous studies and comprehensive guides when choosing eye-tracking measures. For example, Holmqvist et al. (2011) and Lai et al. (2013) offer overviews of measures of eye movements and their interpretations in general, and we provide an overview of the measures used in the mathematics education literature (Appendix Table 3).

The majority of studies interpreted eye-tracking measures in isolation, even if other behavioral or self-reported variables were assessed in parallel. If such other data are collected, researchers could analyze and report their relationship to eye-tracking measures. However, this was not common practice in the studies reviewed here. The use of think-aloud protocols, interviews, reaction times, or accuracy in relation to eye-tracking data can help to verify and specify the interpretations of eye movements. For example, including stimulated recall or interview data can provide additional indications if assumptions like the EMA are valid in specific experiments and help to decide how the collected data can or cannot be interpreted (e.g., Schindler & Lilienthal, 2019).

5.4 Three benefits of using eye tracking in mathematics education

Eye tracking has become a prominent method in mathematics education research. However, many, if not all, of the topics covered in the papers in this review had previously been examined
without eye tracking. Thus, what added value can the method provide? Based on the studies reviewed here, we argue that eye tracking offers unique ways to understand cognitive processes in mathematics education. In many studies reviewed here, eye-tracking measures provided information that could otherwise not be collected. This was usually the case for one of three reasons:

a) The research referred to a time-critical process rather than an outcome. Mathematical tasks often provide a variety of solution approaches and strategies. These strategies are often not visible in the final solution of a task. Moreover, given enough time, students might use strategies and approaches that are sufficient, but not optimal. Thus, observing solution processes without interrupting students is a major challenge that can be tackled through eye tracking (e.g., Inglis & Alcock, 2012; Obersteiner & Tumpek, 2016; van der Weijden et al., 2018).

b) The research included aspects of visualization and mental representations. These questions are typical for mathematics education since mathematics makes use of visualizations in many forms, but at the same time, mathematical objects are often abstract. Making mental representations of these objects visible is a general challenge that can be approached through eye tracking (e.g., Hartmann, Mast, & Fischer, 2016; Myachykov et al., 2015; Risko et al., 2013).

c) The research referred to cognitive processes that cannot be consciously reported. In mathematical thinking, cognitive processes are often complex and therefore hard or impossible to communicate, particularly for younger students. Other representations or cognitive biases might not even be consciously accessible but are nevertheless reflected in eye movements (e.g., Moeller, Neuburger, et al., 2009; Ott et al., 2018; Watson et al., 2005)

When studies further provide a sound and precise theoretical association between eye movements and cognitive processes (e.g., Alqassab et al., 2018; Curtis et al., 2016; Plummer et al., 2017), they usually offer an immediate and unique insight into mathematical thinking. Although a decisive judgment on the specific value of these findings within all areas of mathematics education was beyond the scope of this review, many authors were convinced that eye tracking did offer unique insights that would not have been possible with traditional methods (e.g., written tests or think-aloud protocols) and that brought forward their specific field of research.

6 Conclusion

Eye tracking has the potential to allow novel insights into mathematical thinking and learning. In order to make effective use of this potential, future research should strive for more clarity regarding the theoretical foundations underlying the research questions being addressed and the methodological choices being made. Moreover, the interpretation of eye movements should be based on a reasonable assumption of what eye movements measure and what cognitive processes these measures reflect. Considering the large body of studies that already use eye tracking in mathematics education, it is our hope that this review can guide future researchers in this field and support them in using eye tracking in an efficient and reflective way.

Funding Information Open Access funding provided by Projekt DEAL.
## Appendix A

### Table 3 Overview of all studies included in this review and selected information. The table is displayed in two parts

| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|--------------------|-------------|-----------------------------------------|-------------------|---------------------------|---------------|---------------------|
| Abrahamson et al., 2016 | Proportional equivalence | C, D | 11.3, 13.4 | 30, 46 | one group and cases | 2 | Tobii, X2 (static) | n/a | free |
| Abreu-Mendoza & Arias-Trejo, 2015 | Area comparison and number comparison | D | 12.3 | 48 | two groups (Down syndrome, control) | 1 | Tobii, X2–30 (static) | 30 | n/a |
| Alqassab et al., 2018 | Providing feedback on a peer geometry proof | F | 24.4 | 53 | two groups (experimental) | 1 | Ergoneers, Dikablis (head-mounted) | 25 | n/a |
| Andrà et al., 2009 | Multiple choice, different representations | F | n/a | 46 | two groups (expertise) | 4 | SMI, Hi-Speed 1250 (static) | 1250 | head rest |
| Andrà et al., 2015 | Multiple choice, different representations | F | n/a | 46 | one group | 2 | not reported (static) | n/a | n/a |
| Andrzejewska & Stolinska, 2016 | Multiple choice from different subject areas | D | n/a | 48 | two groups (expertise) | 1 | SMI, Hi-Speed 1250 (static) | 500 | head rest |
| Atagi et al., 2016 | Multiple choice, different representations | F | 21.1 | 32, 32 | one group | 1 | SR Research, Eyelink 1000 (static) | 500 | n/a |
| Bahnmueller et al., 2016 | Number comparison | F | 21.6 | 48 | two groups (language) | 2 | SR Research, Eyelink 1000 (static) | n/a | n/a |
| Beitlich et al., 2014 | Reading undergraduate illustrated textbook | G | 26 | 5 | one group | 1 | SMI, RED500 (static) | 500 | free |
| Beitlich et al., 2015 | Reading heuristic worked examples | E | 16 | 26 | one group | 1 | SMI, RED500 (static) | n/a | n/a |
| Boels et al., 2018 | Histogram and case-value plot interpretation | F | 19–27 | 6 | one group | 4 | Tobii, Pro X2–60 (static) | 60 | n/a |
| Bolden et al., 2015 | Matching representations for multiplication | C | 9–10 | 9 | one group | 2 | Eye Tech, VT2 (static) | 14 | free |
| Bremner et al., 2017 | Keeping track of toy figures | A | 0.41 | 34 | one group | 1 | ASL, Model 5000 (static) | n/a | free |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|------------------|-----------|-------------------|-------------|----------------------------------------|--------------------|--------------------------|---------------|---------------------|
| Brysbaert, 1995  | Reading two-digit arabic numbers | G         | n/a               | 15, 15      | one group                              | 2                  | Generation V (static)    | 1000          | bite bar, head rest  |
| Bulf et al., 2014| Target detection  | G         | 24.4, 24.1        | 16, 17      | one group                              | 1                  | ASL, Model 6 (static)    | 120           | car seat            |
| Bulf et al., 2016| Target detection  | A         | 0.70              | 36          | two groups (experimental)              | 1                  | ASL, Model 6 (static)    | 120           | car seat            |
| Canfield & Smith, 1996| Target detection | A         | 0.42              | 44, 32      | two groups (experimental)              | 1                  | Cohu (manual video coding) | n/a          | sitting on parent’s lap |
| Ceulemans et al., 2014 | Numerosity comparison | A         | 0.70, 0.69        | 25, 26      | two groups (experimental)              | 1                  | Tobii, T60 (static)      | 60            | n/a                |
| Chen & Yang, 2014 | Mental rotation   | F         | 21–24             | 20          | one group                              | 2                  | FaceLAB 4.6 (static)     | 60            | n/a                |
| Chesney et al., 2013 | Addition equation | F         | n/a               | 64          | one group                              | 9                  | SR Research, EyeLink 2000 (static) | 1000        | head and chin rest |
| Cimen & Campbell, 2012 | Studying material on the subject of number theory | F         | 22                | 1           | one group                              | 1                  | Tobii, 1750 (static)     | n/a          | n/a                |
| Clinton et al., 2016 | Illustrated mathematical task | C         | 11.1              | 57          | two groups (prior knowledge)           | 2                  | SR Research, EyeLink 1000 (static) | 1000        | head and chin rest |
| Cohen & Staub, 2015 | Bayes’ theorem task | F         | n/a               | 30          | one group                              | 2                  | SR Research, EyeLink 1000 (static) | 500          | n/a                |
| Cohors-Fresenborg et al., 2010 | Progressive matrices test | E         | n/a               | 57          | two groups (achievement)               | 2                  | not reported, refers to other study (static) | n/a          | n/a                |
| Crisp et al., 2011 | Deriving a function from a sequential table | F         | n/a               | 16          | two groups (education)                 | 1                  | Tobii, T120 (static)     | n/a          | n/a                |
| Curtis et al., 2016 | Simple arithmetic | F         | 20.5              | 109         | four groups (experimental)             | 1                  | SR Research, EyeLink 1000 (static) | 500          | chin rest          |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|-------------------|-------------|------------------------------------------|-------------------|---------------------------|---------------|---------------------|
| De Corte & Verschaffel, 1986a | Word problem solving | C | n/a | 20 | two groups (achievement) | 2 | Debic 80 (static) | 50 | head rest |
| De Corte & Verschaffel, 1986b | Word problem solving | C | n/a | 6 | two groups (achievement) | 4 | Debic 80 (static) | 50 | head rest |
| De Corte et al., 1990 | One-step compare problems | C | n/a | 20 | two groups (achievement) | 1 | Debic 80 (static) | 50 | head rest |
| Dewolf et al., 2015 | Word problems with illustrations | F | 20.8 | 30 | three groups (experimental) | 2 | SR Research, EyeLink II (head-mounted) | 250 | chin rest |
| Duijzer et al., 2017 | Proportional equivalence | C | 11.28 | 38 | one group | 4 | Tobii, X2–30 (static) | n/a | n/a |
| Epelboim & Suppes, 2001 | Geometry word problems | G | 24.4 | 3 | one group | 4 | Maryland Revolving-Field monitor (sensor coil) | 488 | head rest |
| Espino et al., 2005 | Syllogistic reasoning | F | n/a | 32 | one group | 1 | Forward Technologies (static) | 1000 | chin rest |
| Evans et al., 2011 | Dual-task of syntactic and arithmetic operations | F | n/a | 40 | one group | 1 | SR Research, EyeLink II (head-mounted) | n/a | n/a |
| Fischer et al., 2004 | Parity | F | 17–24 | 15 | one group | 1 | Dr. Bouis, Oculometer (static) | 130 | bite bar, head and chin rest |
| Fleig et al., 2017 | Pseudo-contingency | F | 21.8 | 56 | one group | 1 | SMI, RED500 (static) | 500 | n/a |
| Fry, 1988 | Two-step word problems, mental rotation test | F | 23.1 | 37 | one group | 3 | Micrometries, System 1200 (not reported) | 60 | n/a |
| Gandini et al., 2008 | Approximating dots | G | n/a | 30 | two groups (age) | 1 | SMI, iView X (static) | 50 | free |
| Ganor-Stern & Weiss, 2016 | Estimating two-digit multiplication | F | 20–26 | 20, 20 | one group | 1 | SR Research, EyeLink 1000 (static) | n/a | chin rest |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|--------------------|-----------|-------------------|-------------|----------------------------------------|-----------------|--------------------------|---------------|---------------------|
| Gielen et al., 1991 | Number comparison | F | n/a | 11, 11 | one group | 2 | Debic 80 (static) | 50 | head rest |
| Godau et al., 2014 | Addition | C | 8.6 | 20 | one group | 1 | SMI, RED250 (static) | 250 | n/a |
| Godau et al., 2014 | Estimation, arithmetic problem | C | 8.6 | 33 | two groups (experimental) | 2 | SMI, RED (static) | 250 | free |
| Gomez et al., 2017 | Number line | C | 8.4 | 40 | two groups (DCD, control) | 1 | Tobii, TX300 (static) | 300 | n/a |
| Green et al., 2007 | Addition of three-digit numbers | H, F | 20–24, 68.0 | 48 | two groups (age) | 1 | SMI, iView X (static) | 50 | free |
| Hannula & Williams, 2016 | Collaborative geometry problem solving, constructing a specific triangle | D | n/a | 1 | case | 4 | Own development (mobile) | 25 | free |
| Hartmann et al., 2015 | Simple addition or subtraction | F | 23 | 25 | one group | 1 | SMI, RED (static) | 50 | n/a |
| Hartmann, Laubrock, et al., 2016 | Verbal addition/subtraction with visual presentation of numbers | F | 23.2 | 29 | one group | 1 | SR Research, EyeLink 1000 (static) | n/a | chin rest |
| Hartmann, Mast, et al., 2016 | Counting upwards and downwards | F | 23.0 | 18 | one group | 1 | SMI, RED (static) | 50 | free |
| Hegarty et al., 1992 | One-step and two-step arithmetic word problems | F | n/a | 32 | one group | 1 | Iscan, RK 426 (static) | 60 | head rest |
| Hegarty et al., 1995 | Arithmetic word problems | F | n/a | 38 | one group | 1 | Iscan, RK 426 (static) | 60 | head rest |
| Heine et al., 2010 | Number line | C | 8 | 66 | three groups (age) | 2 | SR Research, EyeLink 1000 (static) | 1000 | n/a |
| Hernandez-Sabate et al., 2016 | Tower defense game | D | 14 | 2 | case | 4 | SMI, RED500 (static) | n/a | free |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|-------------------|-------------|----------------------------------------|-------------------|--------------------------|---------------|---------------------|
| Hintz & Meyer, 2015 02 | Arithmetic equation: producing a solution vs. predict a word | F | 22, 25 | 24, 24 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | chin rest |
| Hodds et al., 2014 31 | Reading of a mathematical proof | F | n/a | 28 | experimental vs. control | 1 | Tobii, T120 (static) | 60 | n/a |
| Holmes et al., 2016 03 | Blackjack card game decision (simple addition) | F | n/a | 30, 28 | one group | 1 | Tobii, ET-17 (static) | 30 | n/a |
| Huber et al., 2015 02 | Two-digit whole number comparison | F | 25.7 | 24 | one group | 1 | SR Research, EyeLink 1000 (static) | n/a | chin rest |
| Huber et al., 2014 07 | Decimal fraction comparison | F | 24.8 | 22 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | n/a |
| Huber et al., 2014 02 | Two-digit natural number comparison | F | 26.9 | 24 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | chin rest |
| Huber et al., 2014 07 | Fraction comparison | F | 24.6 | 33 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | n/a |
| Huebner & LeFevre, 2018 05 | Simple addition | F | 20 | 40 | one group | 1 | SR Research, EyeLink 1000 (static) | 500 | chin rest |
| Hunt et al., 2015 90 | Two-digit addition problems | F | 23.8 | 78 | one group | 3 | SR Research, EyeLink II (head-mounted) | 500 | n/a |
| Hurst & Cordes, 2016 07 | Number comparison, notation whole/fraction/decimal | F | 19.4 | 49 | one group | 1 | SMI, RED (static) | 120 | free |
| Inglis & Alcock, 2012 31 | Reading a mathematical proof | F, G | n/a | 30 | two groups (expertise) | 2 | Tobii, T120 (static) | 60 | n/a |
| Inglis & Alcock, 2018 31 | Reading a mathematical proof | G | n/a | 10 | one group | 4 | Tobii, T120 (static) | 60 | free |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|-------------------|-------------|----------------------------------------|-------------------|---------------------------|--------------|---------------------|
| Ischebeck et al., 2016 | Fraction comparison | F | 24.5 | 20 | one group | 1 | SR Research, EyeLink II (head-mounted) | 500 | chin rest |
| Kiili et al., 2014 | Mathematical game | C | n/a | 8 | one group | 4 | Tobii, T60 (static) | n/a | free |
| Kim et al., 2018 | Finding the error in false proof | F | n/a | 35 | two groups (expertise) | 2 | Gazepoint, G3 (static) | n/a | n/a |
| Klein et al., 2014 | Number line task after addition or subtraction | F | 24.8 | 20 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | head and chin rest |
| Klein et al., 2014 | Vector field tasks | F | 20.9 | 41 | one group | 2 | Tobii, X3–120 (static) | 120 | n/a |
| Kohlhase et al., 2001 | Matchstick arithmetic | F | 18–29 | 24 | one group | 1 | Dr. Bouis, Oculometer (static) | 1000 | bite bar |
| Kohlhase et al., 2017 | Reading of mathematical expressions | G | n/a | 23, 29 | two groups (expertise) | 4 | Tobii, T60 (static) | n/a | n/a |
| Kohlhase et al., 2018 | Reading of mathematical textbook | F | n/a | 23 | one group | 4 | Tobii, T60 (static) | n/a | n/a |
| Krichevets et al., 2014 | Locating points in cartesian plane | F | n/a | 44 | three groups (expertise) | 1 | SMI, RED (static) | 120 | n/a |
| Lécuyer et al., 2004 | Possible/impossible teddy bear appearance | A | 0.35 | 12 | one group | 2 | ASL, Model 504 (static) | 20 | free |
| Lee & Wu, 2018 | Reading illustrated geometry text | F | n/a | 65 | one group | 2 | SR Research, EyeLink 1000 (static) | 1000 | chin rest |
| Li et al., 2010 | Counting vs. looking | F | n/a | 10, 12 | one group | 1 | SR Research, EyeLink II (head-mounted) | 250 | chin rest |
| Lin & Lin, 2014a | Geometry problem solving | E | 17–19 | 58 | one group | 1 | SR Research, EyeLink 1000 (static) | 500 | n/a |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|-------------------|-------------|------------------------------------------|-------------------|---------------------------|---------------|---------------------|
| Lin & Lin, 2014b | Geometry problem solving | E | 18 | 57 | one group | 1 | SR Research, EyeLink 1000 (static) | 500 | n/a |
| Lin et al., 2012 | Reading geometry proof | F | n/a | 50 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | chin rest |
| Loetscher et al., 2008 | Bisection | G | 29 | 9 | one group | 1 | Skalar (sensor coil) | 1000 | free |
| Loetscher et al., 2010 | Generating random numbers | G | n/a | 12 | one group | 1 | Skalar (sensor coil) | 1000 | free |
| Macchi Cassia et al., 2016 | Observing ascending vs. descending pattern | A | 1.02 | 40 | two groups | 1 | not reported (manual video coding) | n/a | car seat |
| Masson et al., 2018 | Audio multidigit addition or subtraction | F | 20–24 | 19 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | head and chin rest |
| Merkley & Ansari, 2010 | Number comparison | F | 21.7 | 22 | one group | 1 | Tobii, T120 (static) | n/a | n/a |
| Merschmeyer-Brüwer, 2001 | Counting cubes in 3D shape | C | 7.1–11.4 | 4 | cases | 4 | Omnitrack (static) | n/a | head rest |
| Meyenhoff et al., 2012 | Multi-digit number comparison | F | 25.8 | 18 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | chin rest |
| Michal et al., 2016 | Extracting magnitude information from bar graph | A | 23 | 16, 26 | three groups (age) | 2 | SR Research, EyeLink 1000 (static) | 1000 | head rest |
| Miller Singley & Bunge, 2018 | Fraction comparison | F | 20.3 | 30 | one group | 1 | Tobii, T120 (static) | 120 | n/a |
| Milosavljevic et al., 2011 | Ultra-rapid magnitude comparison | F | n/a | 12, 12, 12 | one group | 2 | SR Research, EyeLink 1000 (static) | 1000 | head and chin rest |
| Moeller et al., 2009a | Bisection | F | 22.2 | 14 | one group | 1 | Dr. Bouis, Oculometer (static) | 333 | bite bar, head rest |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|-------------------|-------------|-----------------------------------------|-------------------|--------------------------|---------------|---------------------|
| Moeller et al., 2009b | Two-digit number comparison | F | 22.7 | 10 | one group | 2 | SR Research, EyeLink II (head-mounted) | 250 | n/a |
| Moeller et al., 2011a | Two-digit addition | E | 10.5, 24.4 | 25 | two groups (age) | 1 | SR Research, EyeLink 1000 (static) | 1000 | chin rest |
| Moeller et al., 2011b | Validating addition problems | F | 22.7 | 19 | one group | 2 | SR Research, EyeLink 1000 (static) | 1000 | chin rest |
| Moeller et al., 2009 | Dot counting | E | 10.5 | 10 | two groups (dyscalculia, control) | 2 | SR Research, EyeLink 1000 (static) | 1000 | head rest |
| Moutsios-Rentzos & Stamatis, 2013 | Addition and subtraction | C | 8–9 | 40 | one group | 4 | None (observation) | n/a | free |
| Moutsios-Rentzos & Stamatis, 2015 | Change and compare problems | C | 6 | 30 | one group | 4 | None (observation) | n/a | free |
| Muldner & Burleson, 2015 | Generating multiple geometry proofs | F | n/a | 21 | two groups (post-hoc) | 2 | Tobii, TX60 (static) | 60 | free |
| Myachykov et al., 2015 | Target fixation | G | 19.8, 22.0 | 19, 17 | one group | 2 | SR Research, EyeLink 1000 (static) | 1000 | head and chin rest |
| Myachykov et al., 2016 | Fixating a dot during auditory number cue | F | 20.2, 22.3 | 17, 17 | one group | 2 | SR Research, EyeLink 1000 (static) | 1000 | head and chin rest |
| Obersteiner & Staudinger, 2018 | Fraction addition | F | 24.3 | 24 | one group | 1 | SMI, Red500 (static) | 500 | n/a |
| Obersteiner & Tumpek, 2016 | Fraction comparison | F | 24.1 | 22 | one group | 1 | SMI, RED500 (static) | 500 | free |
| Obersteiner et al., 2014 | Fraction comparison | G | 26 | 7 | one group | 1 | SMI, RED500 (static) | 500 | free |
| Ogren et al., 2017 | Vector calculus problems confirm or reject | F | 21.5 | 36 | two groups (experimental) | 1 | SMI, RED250 (static) | 250 | n/a |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|--------------------|-----------|--------------------|-------------|-----------------------------------------|--------------------|--------------------------|---------------|---------------------|
| Okamoto & Kuroda, 2014 | Subtraction puzzle | F | 21.1 | 12 | two groups (prior knowledge) | 4 | NAC, EMR-9 (mobile) | 60 | free |
| Olsen et al., 2017 | Collaborative tutoring system for fraction learning | C | n/a | 56 | two groups (experimental) | 2 | SMI, RED250 (static) | 250 | n/a |
| Ott et al., 2018 | Logic task | F | 25.8 | 19 | one group | 1 | Tobii, X2–60 (static) | 60 | free |
| Panse et al., 2018 | Proof comprehension and validation | F | n/a | 26 | two groups (expertise) | 1 | Tobii, T120 (static) | 120 | free |
| Peters, 2010 | Questions about conceptual knowledge | F | n/a | 2 | two groups (expertise) | 4 | Eye Response Technologies, Gaze Tracker (static) | 50 | head rest |
| Plummer et al., 2017 | Assigning representation to number | F | n/a | 40 | one group | 1 | SR Research, EyeLink II (head-mounted) | 250 | free |
| Potgieter & Blignaut, 2018 | Divisibility of 5-digit numbers | C | n/a | 78 | one group | 1 | Tobii, X2 Wide (static) | 60 | n/a |
| Pressigout et al., 2018 | Parity | F | 22.8, 21.3 | 32, 15 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | head and chin rest |
| Price et al., 2017 | Number comparison | F | 19.4 | 56 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | chin rest |
| Reike & Schwarz, 2016 | Number comparison | F | 18–33 | 43 | one group | 2 | SMI, iView X (static) | 240 | head and chin rest |
| Reinert et al., 2015 | Number line | F | 23.6 | 27 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | n/a |
| Rinaldi et al., 2015 | Memorization | G | 26–37 | 10 | one group | 1 | EyeSeeCam (mobile) | 220 | chin rest |
| Risko et al., 2013 | Number comparison | F | n/a | 42 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | n/a |
| Study                                | Year | Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|-------------------------------------|------|------------------|---------------------|-----------|--------------------|-------------|-----------------------------------------|-------------------|----------------------------|---------------|---------------------------|
| Roach et al., 2016                 | 2016 | Mental rotation  | F                   | 20–30     | 20                 | three groups (achievement) | 2            | SR Research, EyeLink 1000 (static)   | 1000           | n/a                       |
| Roy et al., 2017                   | 2017 | Reading           | F                   | n/a       | 34                 | two groups (achievement) | 1            | Tobii, T120 (static)                | 60             | n/a                       |
| Rozek et al., 2014                 | 2014 | Mathematical problem solving | C, F | n/a | 99 | five groups (post-hoc) | 4 | SMi, Hi-Speed 1250 (static) | n/a | n/a                       |
| Ruiz Fernández et al., 2011        | 2011 | Free choice between two faces | F | 22.3 | 139 | one group | 1 | ASL, Model 504 (static) | n/a | head and chin rest |
| Sajka & Rosiek, 2015               | 2015 | Word problem solving | D | 15 | 11 | two groups (achievement) | 4 | SMi, Hi-Speed 1250 (static) | 500 | head rest |
| Salvaggio et al., 2018             | 2018 | Audio number comparison | F | 21.5, 20.2 | 19, 20 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | head rest |
| Schimpf & Spannagel, 2011          | 2011 | Target detection in user interface | F | n/a | 51 | three groups | 2 | LC Technologies (static) | 120 | n/a                       |
| Schindler & Lilienthal, 2018       | 2018 | Abacus counting   | D | 10.9 | 11 | two groups (math. Learning difficulties, control) | 4 | Tobii, Pro Glasses 2 (mobile) | 50 | free                       |
| Schindler et al., 2016             | 2016 | Generating multiple geometry proofs | E | 18 | 1 | case | 4 | Pupil Labs, Pupil Pro (mobile) | n/a | free                       |
| Schleifer & Landen, 2011           | 2011 | Dot counting      | D, E | 8–26 | 60 | one group | 1 | SR Research, EyeLink 1000 (static) | 1000 | n/a                       |
| Schneider et al., 2012             | 2012 | Addition and multiplication | G | 23–24 | 13, 13, 9 | one group | 1 | SR Research, EyeLink 2000 (static) | 2000 | n/a                       |
| Schneider et al., 2008             | 2008 | Number line       | C | 8.0 | 66 | three groups (age) | 1 | SR Research, EyeLink 1000 (static) | 1000 | n/a                       |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|-------------------|-------------|------------------------------------------|-------------------|---------------------------|--------------|----------------------|
| Schwarz & Keus, 2004 | Parity | F | 22, 30 | 16, 16 | one group | 1 | SMI, EyeLink-HiSpeed 2D (static) | 250 | chin rest |
| Shayan et al., 2017 | Proportional equivalence | C | 11.3, 13.4 | 30, 46 | one group | 4 | Tobii, X2–60 (static) | n/a | free |
| Shvarts, 2018a | Collaborative geometry task | F | n/a | 8 | one group | 4 | Pupil Labs (head-mounted) | n/a | free |
| Shvarts, 2018b | Cartesian plane | C | n/a | 10 | one group | 4 | SMI RED and Eye Tribe (mobile) | 120 | free |
| Sophian & Crosby, 2008 | Object counting | F | n/a | 33 | one group | 1 | ASL, Eye Movement Monitor (static) | n/a | n/a |
| Stolinska et al., 2014 | Text reading | F | n/a | 86 | two groups (expertise) | 2 | SMI, Hi-Speed 1250 (static) | 500 | chin rest |
| Strohmaier & Reiss, 2018 | PISA-items | F | n/a | 118 | one group | 1 | SMI, RED500 (static) | n/a | n/a |
| Strohmaier et al., 2017 | PISA-items | F | n/a | 55 | one group | 2 | SMI, RED500 (static) | 500 | n/a |
| Sullivan et al., 2011 | Number line | G | n/a | 16 | two groups (experimental) | 1 | SR Research, EyeLink 1000 (static) | n/a | head and chin rest |
| Suppes et al., 1982 | Addition and subtraction | D, G | 11–14, n/a | 6 | cases | 9 | PERSEUS (static) | n/a | bite bar |
| Suppes et al., 1983 | Addition and subtraction | D, G | n/a | 5 | cases | 9 | PERSEUS (static) | n/a | bite bar |
| Susac et al., 2014 | Rearranging equations | F | 20–26 | 40 | one group | 1 | SMI, Hi-Speed 500 (static) | 500 | head rest |
| Terry, 1921 | Word problem solving | F | n/a | n/a | one group | 4 | no manufacturer (static) | n/a | bite bar, head rest |
| van der Schoot et al., 2009 | Two-step compare problems | C | 11.5 | 40 | two groups (achievement) | 1 | SR Research, EyeLink II (head-mounted) | 250 | n/a |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|-------------------|-------------|----------------------|-------------------|--------------------------|----------------|---------------------|
| van der Weijden et al., 2018 | Number line (bounded and unbounded) | F | 22.5 | 16 | two groups (dyscalculia, control) | 4 | Tobii, T60 (static) | 60 | n/a |
| van Oeffelen & Vos, 1984a | Counting | F | n/a | 7 | one group | 4 | Whittaker, 1998-S EVM (static) | 50 | head rest |
| van Oeffelen & Vos, 1984b | Dot counting | B | 5.6 | 6 | one group | 4 | Whittaker, 1998-S EVM (static) | 50 | n/a |
| van Viersen et al., 2013 | Number line | C | 9.1 | 11 | case (dyscalculia) vs. control | 4 | Tobii, T60 (static) | n/a | n/a |
| van’t Noordende et al., 2016 | Number line | C | 10.9 | 28 | two groups (math. learning difficulties, control) | 1 | Tobii, T60 (static) | 60 | n/a |
| Verdine et al., 2017 | Target detection | B | 3.6 | 79 | two groups (SES) | 2 | Tobii, T60XL (static) | 60 | n/a |
| Verschaffel et al., 1992 | One-step or two-step compare problems | F | n/a | 19, 15, 20 | one group | 1 | Debic 80 (static) | 50 | head rest |
| Verschaffel et al., 1994 | Addition and subtraction word problems | C | 8.5 | 23 | two groups (achievement) | 1 | Debic 80 (static) | 50 | head rest |
| Wall et al., 2015 | Fraction comparison | F | 20 | 14 | one group | 4 | Tobii, T60XL (static) | n/a | n/a |
| Wang et al., 2014 | Geometry test, mental rotation, spatial perception and visualization | C | 11.9 | 16 | two groups (expertise) | 4 | SR Research, EyeLink 2000 (static) | n/a | head rest |
| Watson et al., 2005 | Visual search, enumeration | F, G | 18-26, 57-7-9 | 24, 20 | two groups (age) | 1 | SMI, RED (static) | 50 | free |
| Watson et al., 2007 | Counting | F | 21.9, 20.0 | 7, 10 | one group | 1 | SR Research, EyeLink II (head-mounted) | 250 | free |
| Werner & Raab, 2014 | Water-jar problem | F | 23.7 | 33 | two groups (experimental) | 2 | Tobii, Glasses (mobile) | n/a | n/a |
| Winoto et al., 2017 | Addition | C | 7.5 | 1 | case | 4 | Tobii, X2-60 (static) | n/a | n/a |
| Domain and topic | Task type, stimulus | Age group | Mean age/age range | Sample size | Experimental groups (grouping variable) | Statistical design | Manufacturer, model (type) | Sampling rate | Movement restriction |
|------------------|---------------------|-----------|--------------------|-------------|----------------------------------------|-------------------|--------------------------|---------------|----------------------|
| Yağmur & Çakır, 2016 | Working with Geogebra | F         | 26.5               | 10          | one group                              | 2                 | Tobii, X2–60 (static)   | 60           | free                 |
| Zhou et al., 2012  | Single digit addition and multiplication | F         | 23.2               | 33          | two groups (experimental)              | 1                 | SCAN system (EOG)       | 1000         | free                 |
| Zhu et al., 2018   | Addition and subtraction | F         | 20.4               | 26, 26      | one group                              | 1                 | SR Research, EyeLink (static) | 500         | chin rest            |

| Calibration (threshold) | Event detection algorithm or software (thresholds) | Measures | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods | Publication type |
|-------------------------|-----------------------------------------------------|----------|---------------------------------|-------------------------------|-----------------------|--------------|------------------|
| Abrahamson et al., 2016 | n/a                                                 | SP       | 25                             | n/a                           | n/a                   | Finger movements, videography, interviews (rel.) | 4               |
| Abreu-Mendoza & Arias-Trejo, 2015 | 5-point (n/a)                                      | DLL      | 31                             | n/a                           | n/a                   | n/a          | 4               |
| Alqassab et al., 2018  | yes (n/a)                                           | n/a      | FD, IC                         | 13                            | 1.9                   | Proof comprehension and peer feedback type, accuracy (ind.) | 4               |
| Andrà et al., 2009     | 13-point (n/a)                                      | n/a      | FD                             | 12                            | n/a                   | n/a          | 1               |
| Andrà et al., 2015     | n/a                                                 | n/a      | FD, FC, DT, IC                 | 13                            | n/a                   | n/a          | 4               |
| Andrzejewska & Stolinska, 2016 | yes (n/a)                                        | BeGaze (n/a) | BD, BFQ, BC, FD, FFQ          | 31                            | 7.7                   | Self-reports of accuracy, subjective difficulty (rel.) | 4               |
| Atagi et al., 2016     | 5-point (n/a)                                       | n/a      | FD, FC                         | 12                            | 8.6                   | Accuracy, response time, interview (ind.) | 4               |
| Bahnmueller et al., 2016 | 9-point (n/a)                                     | n/a      | DT                             | 14                            | n/a                   | Response time (ind.) | 4               |
| Beitlich et al., 2014  | 9-point (n/a)                                       | n/a      | FD                             | 12                            | 16.7                  | n/a          | 1               |
| Beitlich et al., 2015  | yes (n/a)                                           | n/a      | FD, IC                         | 12                            | n/a                   | Reports of perception (rel.) | 1               |
| Reference                  | Methodology          | Event detection algorithm or software (thresholds) | Measures | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods                                      | Publication type |
|----------------------------|----------------------|---------------------------------------------------|----------|--------------------------------|--------------------------------|-------------------------|----------------------------------------------------|-----------------|
| Boels et al., 2018         | 5-point (n/a)        | manual (n/a)                                      | SP, HM   | 12                             | n/a                           | n/a                     | Cued retrospective interview (rel.)                | 1               |
| Bolden et al., 2015        | yes (n/a)            | n/a                                               | FD, SP   | 14                             | n/a                           | n/a                     | Attainment, accuracy (rel.)                         | 4               |
| Brenner et al., 2017       | yes (n/a)            | ASL Eyenal (n/a)                                 | FD       | 27                             | 45.2                          | n/a                     | n/a                                                | 4               |
| Brysbaert, 1995            | yes (n/a)            | n/a (50 ms)                                       | FD       | 31                             | n/a                           | n/a                     | Two additional experiments (ind.)                  | 4               |
| Bulf et al., 2014          | 9-point (n/a)        | n/a                                               | TFF      | 26                             | 16                            | n/a                     | n/a                                                | 4               |
| Bulf et al., 2016          | 3-point (n/a)        | n/a                                               | TFF      | 26                             | 25.0                          | n/a                     | Accuracy (ind.)                                    | 4               |
| Canfield & Smith, 1996     | n/a                  | manual (n/a)                                      | SC       | 27                             | 7.3                           | 26                      | n/a                                                | 4               |
| Ceulemans et al., 2014     | 5-point (n/a)        | Tobii fixation filter (35px)                     | FD       | 31                             | 10.5                          | n/a                     | n/a                                                | 4               |
| Chen & Yang, 2014          | yes (1°)             | GazeTracker (100 ms)                              | FD, FDA, RC | 31                             | n/a                           | n/a                     | Accuracy, performance, interview (rel.)            | 4               |
| Chesney et al., 2013       | n/a                  | I-VT (0.2°, 9500°/s², 30°/s)                      | IC       | 13                             | n/a                           | n/a                     | Strategy use (rel.)                                | 4               |
| Cimen & Campbell, 2012     | n/a                  | n/a                                               | BF       | 50                             | n/a                           | n/a                     | Heart rate, think aloud, facial expressions, body movements, self-reports (rel.) | 1               |
| Clinton et al., 2016       | yes (n/a)            | n/a                                               | FD, FDA  | 31                             | 8.1                           | n/a                     | n/a                                                | 2               |
| Cohen & Staub, 2015        | 9-point (1°)         | n/a                                               | FD       | 12                             | n/a                           | 0.2                     | Accuracy (rel.)                                    | 4               |
| Cohors-Fresenborg et al., 2010 | n/a                  | QuaDiPF-Eye (n/a)                                | FD, SP   | 14, 50                         | n/a                           | n/a                     | Metacognitive behavior (rel.)                       | 4               |
| Crisp et al., 2011         | n/a                  | n/a                                               | IC       | 13                             | n/a                           | n/a                     | n/a                                                | 1               |
| Curtis et al., 2016        | 9-point (n/a)        | I-DT (100 ms, 0.15°)                              | FFP, FD, FC | 12, 24                         | n/a                           | 5                       | Accuracy, response time (ind.)                     | 4               |
| De Corte & Verschaffel, 1986a | n/a                  | manual (n/a)                                      | FD, SP   | 14                             | 9.1                           | n/a                     | Achievement (rel.)                                 | 1               |
| De Corte & Verschaffel, 1986b | n/a                  | manual (n/a)                                      | SP       | 14                             | n/a                           | n/a                     | Achievement (rel.)                                 | 1               |
| De Corte et al., 1990 | n/a | I-DT (100 ms, 1°) | FD | 12 | n/a | n/a | Response time, performance (rel.) | 3 |
| Dewolf et al., 2015 | 9-point (n/a) | n/a | FC | 11 | n/a | 0.6 | Perceived correctness (ind.), second experiment (ind.) | 4 |
| Duijzer et al., 2017 | yes (n/a) | Tobii fixation filter (5 points, 25px) | FC, FDA, SP | 25 | 15.6 | n/a | Think aloud, interview (ind.) | 4 |
| Epelboim & Suppes, 2001 | 9-point (n/a) | manual (n/a) | FD | 32 | n/a | n/a | Verbal protocol (rel.) | 4 |
| Espino et al., 2005 | yes (n/a) | n/a | DT | 31 | n/a | n/a | Accuracy (ind.) | 4 |
| Evans et al., 2011 | n/a | n/a | DT | 12 | n/a | n/a | Accuracy (ind.) | 4 |
| Fischer et al., 2004 | yes (0.5°) | n/a | FD, SL | 26 | n/a | n/a | Performance (rel.), second experiment (ind.) | 4 |
| Fleig et al., 2017 | n/a | I-VT (12 ms, 40°/s) | FD | 12 | 6.7 | n/a | 4 |
| Fry, 1988 | n/a | I-DT (n/a) | FD | 12 | n/a | n/a | Spatial abilities (rel.) | 1 |
| Gandini et al., 2008 | 9-point (n/a) | n/a | FD, FC, SL | 13 | n/a | n/a | Second experiment (ind.) | 4 |
| Ganor-Stern & Weiss, 2016 | 5-point (2°) | n/a | FD, FC, FFP | 12 | 4.8 | n/a | Accuracy, response time, strategy use (ind.) | 4 |
| Gielen et al., 1991 | yes (n/a) | I-DT (100 ms, 1°/0.75°) | FFD, FD | 31 | n/a | n/a | n/a | 4 |
| Godau et al., 2014 | 5-point (n/a) | n/a | FD, FC | 13 | n/a | n/a | Second experiment (ind.) | 4 |
| Godau et al., 2014 | 5-point (n/a) | n/a | SL | 13 | 2.9 | n/a | Response time (ind.) | 3 |
| Gomez et al., 2017 | n/a | I-VT (30°/s) | FD, FC, FDA, SL | 13 | n/a | n/a | Performance, disorder score (rel.), accuracy, reaction time (ind.) | 4 |
| Green et al., 2007 | 9-point (n/a) | I-DT (n/a) | FD | 13 | n/a | n/a | Self-reports (rel.), Response time (ind.) | 4 |
| Haataja et al., 2018 | n/a | n/a | DT | 12 | n/a | n/a | n/a | 1 |
| Hamada & Iwaki, 2012 | n/a | n/a | FDA, SDA | 21 | n/a | n/a | Magnetoencephalogram (MEG, rel.) | 3 |
| Hannula & Williams, 2016 | manual (n/a) | n/a | SP | 14 | n/a | n/a | Body language (rel.) | 1 |
| Hartmann et al., 2015 | n/a | I-DT (80 ms, 100px) | FFP | 21 | n/a | 7.3 | Response time (ind.) | 4 |
| Study                        | Calibration (threshold) | Event detection algorithm or software (thresholds) | Measures | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods | Publication type |
|------------------------------|-------------------------|---------------------------------------------------|----------|--------------------------------|--------------------------------|--------------------------|---------------|-----------------|
| Hartmann, Laubrock, et al., 2016 | 9-point (n/a)           | EyeLink data viewer (cognitive configuration)     | FD, TFF, RFC | 21 | 3.3 | n/a | Accuracy, response time (ind.) | 4 |
| Hartmann, Mast, et al., 2016 | 5-point (0.8°)          | BeGaze (80 ms, 100px)                             | FP       | 22 | 10.0 | n/a | n/a | 4 |
| Hegarty et al., 1992         | n/a                     | manual (n/a)                                      | FC, RFC  | 13 | 15.8 | n/a | Accuracy, response time, interview (rel.) | 4 |
| Hegarty et al., 1995         | yes (n/a)               | manual (n/a)                                      | RFC      | 13 | n/a | n/a | Accuracy (rel.) | 4 |
| Heine et al., 2010           | n/a (n/a)               | n/a                                               | FP, FC, FD | 21 | n/a | 5.6 | Accuracy (rel.) | 4 |
| Hernandez-Sabate et al., 2016| yes (n/a)               | BeGaze (n/a)                                      | FP       | 13 | n/a | n/a | n/a | 1 |
| Hintz & Meyer, 2015          | yes (n/a)               | EyeLink software (n/a)                            | FD, TFF  | 21 | 2.0 | n/a | n/a | 4 |
| Hodds et al., 2014           | 9-point (n/a)           | n/a                                               | FDA, IC  | 13, 31 | 12.5 | n/a | n/a | 4 |
| Holmes et al., 2016          | 5-point (n/a)           | n/a                                               | FP       | 21 | n/a | 2.3 | n/a | 4 |
| Huber et al., 2015           | 9-point (n/a)           | n/a                                               | FC       | 31 | n/a | 4.7 | Response time (ind.) | 4 |
| Huber et al., 2014           | 9-point (n/a)           | n/a                                               | DT       | 12 | 12.0 | n/a | Accuracy, response time (ind.) | 4 |
| Huber et al., 2014           | 9-point (n/a)           | n/a                                               | FD       | 31 | n/a | n/a | Accuracy, response time (ind.) | 4 |
| Huber et al., 2014           | 9-point (n/a)           | n/a                                               | FC, FFP  | 31 | 8.3 | n/a | Accuracy, response time (ind.) | 4 |
| Huebner & LeFevre, 2018      | 9-point (n/a)           | I-DT (100 ms, 0.15°)                              | FD, FC   | 13 | n/a | 13 | Response time (rel.) | 4 |
| Hunt et al., 2015            | yes (n/a)               | n/a                                               | FD, FC, RC, SC | 12 | n/a | n/a | Accuracy, math anxiety (rel.), response time (ind.) | 3 |
| Hunt & Cordes, 2016          | 4-point (1°)            | Tobii fixation filter (n/a)                       | FD       | 21, 31 | 21.0 | n/a | Accuracy, response time (ind.) | 4 |
| Inglis & Alcock, 2012        | 9-point (n/a)           | own (80 ms, in AOI)                               | FD, FDA, IC, SL | 21, 31 | 21.0 | n/a | Validity judgment (ind.) | 4 |

Eye-tracking methodology in mathematics education research...
| Reference                     | Calibration (threshold) | Event detection algorithm or software (thresholds) | Measures | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods                                                                 | Publication type |
|-------------------------------|-------------------------|---------------------------------------------------|----------|---------------------------------|-------------------------------|-------------------------|-----------------------------------------------------------------------------|------------------|
| Inglis & Alcock, 2018         | 9-point (n/a)           | n/a                                               | FP       | 14                              | n/a                           | n/a                     | n/a                                                                         | 1                |
| Ischebeck et al., 2016        | 9-point (0.3°)          | I-VT (9500°/s², 35°/s)                            | FC, FP, FFP, LFP | 31                              | n/a                           | 12                      | Accuracy, response time (ind.)                                              | 4                |
| Kiili et al., 2014            | yes (n/a)               | n/a                                               | FC, HM   | 11                              | n/a                           | n/a                     | Cued retrospective interview (rel.)                                          | 4                |
| Kim et al., 2018              | 9-point (n/a)           | Gazepoint (n/a)                                   | FC, FDA, DT | 14                              | 7.9                           | n/a                     | Accuracy (ind.)                                                            | 1                |
| Klein et al., 2014            | 9-point (n/a)           | n/a                                               | FFP, LFP | 21                              | 13.0                          | n/a                     | Accuracy, response time (ind.)                                              | 4                |
| Klein et al., 2014            | 9-point (n/a)           | I-VT (8500°/s², 30°/s)                            | FD, FC, SL, SD | 13                              | n/a                           | n/a                     | Accuracy and interview (rel.)                                               | 4                |
| Knoblich et al., 2001         | yes (n/a)               | I-DT (5 ms, 1/9°)                                 | FD, FDA  | 12                              | n/a                           | n/a                     | Performance (rel.)                                                         | 4                |
| Kohlhase et al., 2017         | n/a                     | n/a                                               | SP, HM   | 14                              | n/a                           | n/a                     | Think aloud, questionnaire (ind.)                                           | 1                |
| Kohlhase et al., 2018         | n/a                     | n/a                                               | FC, VC   | 14                              | n/a                           | n/a                     | Think aloud (rel.)                                                          | 1                |
| Krichevets et al., 2014       | 9-point (0.5°)          | I-VT (120°/s)                                     | SD, FC, SL | 21                              | n/a                           | n/a                     | Response time (ind.)                                                        | 4                |
| Lécuyer et al., 2004          | 2-point (n/a)           | n/a                                               | FD       | 27                              | 76.0                          | n/a                     | Second experiment (ind.)                                                   | 4                |
| Lee & Wu, 2018                | yes (n/a)               | n/a                                               | FD, IC, FC, SP | 14                              | 4.4                           | 8.8                     | Accuracy, response time (ind.)                                              | 4                |
| Li et al., 2010               | 9-point (0.5°)          | n/a                                               | FC, RFC  | 13                              | n/a                           | n/a                     | Accuracy, response time (ind.)                                              | 4                |
| Lin & Lin, 2014a              | yes (n/a)               | n/a                                               | FD, FC, RFC | 31                              | 9.4                           | n/a                     | Cognitive load (rel.), perceived difficulty, second experiment (ind.)     | 4                |
| Lin & Lin, 2014b              | yes (n/a)               | EyeLink data viewer (n/a)                         | FFP, FF, FD, FC, LFD | 90                              | 9.5                           | n/a                     | Accuracy, perceived difficulty, cognitive load (rel.)                      | 4                |
| Lin et al., 2012              | yes (n/a)               | n/a                                               | FD, SC   | 12                              | 23.1                          | n/a                     | n/a                                                                         | 1                |
| Loetscher et al., 2008        | n/a                     | n/a                                               | FP       | 21                              | n/a                           | n/a                     | n/a                                                                         | 3                |
| Loetscher et al., 2010        | n/a                     | n/a                                               | FP       | 21                              | n/a                           | n/a                     | n/a                                                                         | 3                |
| Study                                      | Calibration (threshold) | Event detection algorithm or software (thresholds) | Measures | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods                                    | Publication type |
|-------------------------------------------|-------------------------|---------------------------------------------------|----------|---------------------------------|---------------------------------|------------------------|-----------------------------------------------|------------------|
| Macchi Cassia et al., 2016                | n/a                     | manual (n/a)                                      | FD       | 27                              | 31.0                            | n/a                    | n/a                                           | 4                |
| Masson et al., 2018                       | 9-point (n/a)           | n/a                                               | FP       | 22                              | n/a                            | n/a                    | Response time (rel.)                          | 4                |
| Merkley & Ansari, 2010                    | yes (n/a)               | I-DT (50 ms, in AOI)                               | FD, FC, IC| 13                              | 12.0                            | n/a                    | Accuracy, response time (rel.)                | 3                |
| Merschmeyer-Brüwer, 2001                  | n/a                     | Vision (n/a)                                      | FDA, FD  | 31                              | n/a                            | n/a                    | Accuracy, response time (rel.)                | 1                |
| Meyerhoff et al., 2012                    | 9-point (n/a)           | n/a                                               | FC, IC   | 13                              | n/a                            | n/a                    | Response time (ind.)                          | 4                |
| Michal et al., 2016                       | 9-point (n/a)           | n/a                                               | SL, SD   | 13                              | 10.6                            | n/a                    | Response time (rel.)                          | 4                |
| Miller Singley & Bunge, 2018              | 9-point (n/a)           | n/a                                               | IC       | 13                              | 21.1                            | n/a                    | Accuracy, response time (ind.)                | 4                |
| Milosavljevic et al., 2011                | n/a                     | n/a                                               | TFF      | 26                              | n/a                            | n/a                    | Accuracy (rel.)                               | 4                |
| Moeller et al., 2009a                     | manual (n/a)            | n/a                                               | FD       | 31                              | n/a                            | n/a                    | Response time (ind.)                          | 4                |
| Moeller et al., 2009b                     | 9-point (n/a)           | n/a                                               | FC       | 12                              | n/a                            | n/a                    | Response time (ind.)                          | 4                |
| Moeller et al., 2011a                     | 9-point (n/a)           | n/a                                               | DT       | 12                              | n/a                            | n/a                    | Accuracy, response time (ind.)                | 4                |
| Moeller et al., 2011b                     | 9-point (n/a)           | n/a                                               | FFD, DT, IC| 31                              | 5.0                             | n/a                    | Accuracy, response time (ind.)                | 4                |
| Moeller et al., 2009                      | 9-point (n/a)           | n/a                                               | FC, FDA  | 13, 31                          | n/a                            | n/a                    | Response time (ind.)                          | 4                |
| Moutsios-Rentzos & Stamatis, 2013         | n/a                     | n/a                                               | FP       | 23                              | n/a                            | n/a                    | Accuracy, response time (rel.)                | 4                |
| Moutsios-Rentzos & Stamatis, 2015         | n/a                     | n/a                                               | FP       | 23                              | n/a                            | n/a                    | Accuracy, response time (ind.)                | 4                |
| Calibration (threshold) | Event detection algorithm or software (thresholds) | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods | Publication type |
|-------------------------|---------------------------------------------------|---------------------------------|--------------------------------|------------------------|--------------|----------------|
| Muldner & Burleson, 2015 | 9-point (n/a)                                     | FC, FD, FDA, SL, SV, PS         | 31                             | n/a                    | n/a          | 4              |
| Myachykov et al., 2015  | 9-point (n/a)                                     | FC, FD, FDA, SL, SV, PS         | 26                             | n/a                    | n/a          | 4              |
| Myachykov et al., 2016  | 9-point (n/a)                                     | FC, FD, FDA, SL, SV, PS         | 22                             | n/a                    | n/a          | 4              |
| Obersteiner & Staudinger, 2018 | 5-point (n/a)                           | FC, SC, IC                      | 13                             | 14.3                   | n/a          | 4              |
| Obersteiner & Tumpek, 2016 | 9-point (n/a)                              | FC, IC                          | 13                             | 12.0                   | n/a          | 4              |
| Obersteiner et al., 2014 | 9-point (n/a)                                     | FC                              | 13                             | 12.5                   | n/a          | 1              |
| Ögren et al., 2017      | 5-point (0.5°)                                    | FC, IC                          | 13                             | n/a                    | n/a          | 4              |
| Okamoto & Kuroda, 2014  | 9-point (n/a)                                     | FC                              | 14                             | n/a                    | n/a          | 3              |
| Olsen et al., 2017      | n/a                                              | FC                              | 40                             | n/a                    | n/a          | 2              |
| Ott et al., 2018        | 9-point (n/a)                                     | FC                              | 12                             | 9.5                    | n/a          | 4              |
| Panse et al., 2018      | yes (n/a)                                        | FC                              | 13                             | 18.8                   | n/a          | 4              |
| Peters, 2010            | yes (n/a)                                        | FC, FD, FDA, SL, SV             | 14                             | n/a                    | n/a          | 1              |
| Plummer et al., 2017    | 9-point (n/a)                                     | FC, FD, FDA, SL, SV             | 14                             | n/a                    | 1.2          | 4              |
| Potgieter & Blignaut, 2018 | 9-point (n/a)                                   | FC                              | 12                             | n/a                    | 1.8          | 4              |
| Pressigout et al., 2018 | 9-point (n/a)                                     | TFF                            | 26                             | 2.1                    | 2.2          | 4              |
| Price et al., 2017      | n/a                                              | FC, DT, FFP                     | 12, 24                         | 23.3                   | n/a          | 4              |
| Reike & Schwarz, 2016   | n/a                                              | FC, DT, FFP                     | 12, 24                         | 23.3                   | n/a          | 4              |
| Event detection algorithm or software (thresholds) | Measures | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods | Publication type |
|--------------------------------------------------|----------|---------------------------------|-------------------------------|------------------------|---------------|----------------|
| Reinert et al., 2015 yes (n/a)                   | FC, FFP  | 25                              | n/a                           | n/a                    | Accuracy (ind.) | 3              |
| Rinaldi et al., 2015 5-point (n/a)               | FP       | 13                              | n/a                           | n/a                    | Response time (ind.) | 4              |
| Risko et al., 2013 9-point (n/a)                 | TFF, FFP | 24                              | n/a                           | n/a                    | Response time (ind.) | 4              |
| Roach et al., 2016 n/a                           | FDA, FC, FP | 31                              | 13.0                          | n/a                    | Response time (rel.) | 4              |
| Roy et al., 2017 yes (n/a)                        | FDA, FDA | 12, 31                          | n/a                           | n/a                    | Second experiment (ind.) | 4              |
| Rozek et al., 2014 n/a                           | BeGaze (n/a) | 11                              | n/a                           | n/a                    | Strategy use (rel.) | 4              |
| Ruiz Fernández et al., 2011 n/a                 | FFP      | 11                              | 7.3                           | n/a                    | n/a            | 3              |
| Sajka & Rosiek, 2015 yes (0.5°)                  | BeGaze (n/a) | 14                              | n/a                           | n/a                    | n/a            | 1              |
| Salvaggio et al., 2018 9-point (n/a)             | SP, HM   | 21                              | 2.5                           | 12.8                   | Response time (ind.) | 4              |
| Schimpf & Spangnagel, 2011 n/a                  | TFF      | 11                              | 27.1                          | n/a                    | Second experiment (ind.) | 4              |
| Schindler & Lilienthal, 2018 n/a                | SP       | 14                              | 45.0                          | n/a                    | Think aloud (rel.) | 1              |
| Schindler et al., 2016 yes (n/a)                 | SP       | 14                              | n/a                           | n/a                    | Written solution, video recording (rel.) | 1              |
| Schleifer & Landerl, 2011 9-point (0.5°)         | SF       | 13                              | n/a                           | n/a                    | Response time (ind.) | 4              |
| Schneider et al., 2012 n/a                       | SP       | 14                              | n/a                           | n/a                    | Response time (rel.) | 4              |
| Schneider et al., 2008 n/a                       | FC       | 12                              | n/a                           | 4.6                    | Accuracy (rel.) | 4              |
| Schwarz & Keus, 2004 yes (n/a)                    | TFF      | 26                              | n/a                           | 25                     | Accuracy, response time (rel.) | 4              |
| Shayan et al., 2017 n/a                          | FP, SD   | 25                              | n/a                           | n/a                    | Interview (rel.) | 2              |
| Shvarts, 2018a n/a                               | JA       | 40                              | n/a                           | n/a                    | Dialog transcript (rel.) | 1              |
| Shvarts, 2018b yes (n/a)                          | JA       | 40                              | n/a                           | n/a                    | Gestures, audio (rel.) | 1              |
Table 3 (continued)

| Study                    | Calibration (threshold) | Event detection algorithm or software (thresholds) | Measures | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods                                      | Publication type |
|--------------------------|-------------------------|---------------------------------------------------|----------|--------------------------------|-------------------------------|------------------------|---------------------------------------------------|-----------------|
| Sophian & Crosby, 2008   | yes (n/a)               | n/a                                               | FC       | 12                             | 2.9                           | n/a                    | Accuracy (ind.)                                   | 4               |
| Stolinska et al., 2014   | yes (n/a)               | BeGaze (n/a)                                      | SC, SL, SV, SDA | 13 | n/a                           | n/a                       | n/a                   | n/a                                               | 1               |
| Strohmaier & Reiss, 2018 | n/a                     | n/a                                               | FDA, SC, RC, DT | 13, 31 | n/a                       | n/a                       | n/a                   | n/a                                               | 1               |
| Strohmaier et al., 2017  | n/a                     | I-DT (100 ms, 2.4°)                               | FDA, FD, RC, SC, FC | 13, 31 | n/a                       | n/a                       | Self-concept (rel.) | 1               |
| Sullivan et al., 2011    | 9-point (n/a)            | n/a                                               | FD, FC   | 12                             | 27.3                         | 1.3                    | Accuracy (rel.)                                   | 4               |
| Suppes et al., 1982      | 121-point (n/a)          | n/a                                               | FD, IC   | 13                             | n/a                           | n/a                   | n/a                                               | 2               |
| Suppes et al., 1983      | 121-point (n/a)          | n/a                                               | FD, IC   | 13                             | n/a                           | n/a                   | n/a                                               | 4               |
| Susac et al., 2014       | 13-point (n/a)           | n/a                                               | FD, FC   | 14                             | n/a                           | n/a                   | Accuracy, response time, interview about metacognition (rel.) | 4               |
| Terry, 1921              | n/a                     | n/a                                               | FC       | 14                             | n/a                           | n/a                   | n/a                                               | 4               |
| van der Schoot et al., 2009 | yes (1°)              | I-VT (0.1°, 8500°/s, 30°/s)                     | FC       | 14                             | n/a                           | n/a                   | Accuracy, reading comprehension, working memory (rel.) | 4               |
| van der Weijden et al., 2018 | 9-point (n/a)            | I-VT (Tobii default)                             | FC       | 13                             | n/a                           | 3.1                    | Cued retrospective interview (rel.), accuracy (ind.) | 4               |
| van Oeffelen & Vos, 1984a | 9-point (n/a)            | n/a                                               | FD, SC, SDA, BD | 13 | n/a                       | n/a                       | Response time (ind.) | 4               |
| van Oeffelen & Vos, 1984b | 9-point (n/a)            | n/a                                               | FC, SC, CV, BD, FD, SD | 14, 31 | n/a                       | n/a                       | Response time (ind.) | 4               |
| van Viersen et al., 2013  | n/a                     | n/a                                               | SP       | 14                             | n/a                           | 0.5                    | Accuracy, working memory, number sense (ind.)     | 4               |
| van’t Noordende et al., 2016 | 9-point (n/a)            | I-DT (100 ms, 0.5 cm x 0.5 cm)                   | FC       | 13                             | n/a                           | 1.84                   | Accuracy (ind.)                                   | 4               |
| Domain and topic: (0X—numbers and arithmetic (01—perception and mental representation of single digit and non-symbolic numbers, number words, counting; 02—perception and mental representation of multi-digit numbers; 03—spatial-numerical associations; 04—number line estimation task; 05—basic arithmetic; 06—number perception and magnitude in infants; 07—rational numbers and proportionality); 1X—geometry, shape, and form (11—geometric proof; 12—mental rotation; 13—construction; 14—Cartesian plane; 15—calculation; 16—vector fields; 17—shapes); 2X—reading and word problem solving (21—prototype word problems; 22—complex mathematical text or word problems; 23—mathematical parsing); 3X—reasoning and proof (31—reading and comprehension of proofs; 32—proportional reasoning; 33—probabilistic reasoning; 34—logic; 35—functional thinking); 40—use of representations; 50—learning difficulties; 60—computer-supported learning; 70—mathematical problem solving; 80—statistics; 90—affective variables) | Age group: A—infants; B—kindergarten; C—primary education (grades 1–6); D—lower secondary education (grades 7–9); E—higher secondary education (grades 10–13); F—tertiary education; G—other adults; H—seniors | Statistical design: 1—mixed or within-subject design; 2—between-subject design; 3—correlational; 4—descriptive; 9—other | Measures: BC—blink count; BD—blink duration; BF—blink frequency; DLL—difference in longest look; DT—dwell time, reading time; FC—fixation count; FD—total fixation duration; FDA—average fixation duration; FFD—first fixation duration; FFP—first fixation position; FFQ—fixation frequency; FP—fixation position; HM—heat map; IC—inter-scanning count, transitions; JA—joint visual attention; LFP—last fixation position; LFD—last fixation duration; PS—pupil size; RC—regressive saccades count; RFC—revisiting |
| Calibration (threshold) | Event detection algorithm or software (thresholds) | Measures | Interpretation of eye movements | Participants excluded (percent) | Data excluded (percent) | Other methods | Publication type |
|------------------------|---------------------------------|-----------|-----------------|-----------------|----------------|----------------|----------------|
| Verdine et al., 2017   | 5-point (n/a)                   | TFF       | 32              | 22.5            | n/a            | SES, spatial abilities (rel.) | 4              |
| Verschaffel et al., 1992 | n/a                            | FD        | 13              | n/a             | 2              | Accuracy, response time (ind.) | 4              |
| Verschaffel et al., 1994 | n/a                            | I-DT (n/a) | 13              | n/a             | n/a            | Accuracy, response time, retrospective reports (rel.) | 2              |
| Wall et al., 2015      | yes (n/a)                       | FC        | 12              | n/a             | n/a            | Confidence rating (ind.) | 1              |
| Wang et al., 2014      | n/a                             | SP, HM    | 14              | n/a             | n/a            | Achievement (ind.) | 4              |
| Watson et al., 2005    | 9-point (n/a)                   | I-VT (45°/s, fix. Dur. 120 ms) | FFQ, RFC | 32              | 6.4            | Accuracy, response time (ind.) | 4              |
| Watson et al., 2007    | 9-point (n/a)                   | EyeLink data viewer (n/a) | SF, FD, SD | 22              | 5.6            | Response time (ind.) | 4              |
| Werner & Raab, 2014    | n/a                             | FC        | 21              | 17.5            | n/a            | Strategy use (ind.) | 4              |
| Winoto et al., 2017    | n/a                             | SP, HM    | 11              | 50.0            | n/a            | Visual coding (rel.); usability questionnaire (ind.) | 1              |
| Yağmur & Çakır, 2016  | n/a                             | Toeii fixation filter (n/a) | FD, FC, FDA | 90              | n/a            | n/a            | 1              |
| Zhou et al., 2012     | n/a                             | FP        | 13              | 2.9             | 6.5            | Accuracy, response time (ind.) | 4              |
| Zhu et al., 2018      | yes (n/a)                       | TFF       | 26              | n/a             | n/a            | n/a            | 4              |

Eye-tracking methodology in mathematics education research...
fixations count; SC—saccade count; SD—saccade direction; SDA—average saccade duration; SF—saccade frequency; SL—saccade length; SP—scan path pattern; SV—saccade velocity; TFF—time to first fixation, fixation latency, saccade latency; VC—visit count

Interpretation of eye movements: 1X—visual focus and overt attention, eye mind assumption (11—attention on single object; 12—compare attention on 2 or more objects; 13—order of attention (quantitative); 14—order of attention (qualitative)); 2X—mental representation and covert attention (21—visual mental representation; 22—motoric mental representation; 23—brain hemisphericity; 24—areas of covert attention; 25—attentional anchors; 26—reaction time; 27—anticipation); 3X—cognitive effort and resources (31—information extraction; 32—memory capacity); 40—joint attention; 50—metacognitive control; 90—explorative

Other methods: rel.—in relation to eye tracking; ind.—independent of eye tracking

Publication type: 1—proceedings paper; 2—book chapter; 3—short paper; 4—full paper
References

*Abrahamson, D., Shayan, S., Bakker, A., & Van der Schaaf, M. (2016). Eye-tracking Piaget: Capturing the emergence of attentional anchorrefers in the coordination of proportional motor action. Human Development, 58(4–5), 218–224. https://doi.org/10.1159/000443153

*Abreu-Mendoza, R. A., & Arias-Trejo, N. (2015). Numerical and area comparison abilities in Down syndrome. Research in Developmental Disabilities, 41–42, 58–65. https://doi.org/10.1016/j.ridd.2015.05.008

*Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018). The impact of peer solution quality on peer-feedback provision on geometry proofs: Evidence from eye-movement analysis. Learning and Instruction, 58, 182–192. https://doi.org/10.1016/j.learninstruc.2018.07.003

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. Psychological Science, 15(4), 225–231. https://doi.org/10.1111/j.0956-7976.2004.00656.x

*Andrà, C., Arzarello, F., Ferrara, F., Holmqvist, K., Lindstrøm, P., Robutti, O., & Sabena, C. (2009). How students read mathematical representations: An eye tracking study. In M. Kaldrimidou & H. Sakonidis (Eds.), Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 49–56). Thessaloniki, Greece: PME.

*Andrà, C., Lindstrøm, 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, 237–259. https://doi.org/10.1007/s10763-013-9484-y

*Andrzejewska, M., & Stolinska, A. (2016). Comparing the difficulty of tasks using eye tracking combined with subjective and behavioural criteria. Journal of Eye Movement Research, 9(3), 1–16. https://doi.org/10.16910/jemr.9.3.3

*Atagi, N., DeWolf, M., Stigler, J. W., & Johnson, S. P. (2016). The role of visual representations in college students’ understanding of mathematical notation. Journal of Experimental Psychology-Applied, 22(3), 295–304. https://doi.org/10.1037/xap0000090

*Bahmrueller, J., Huber, S., Nuerk, H. C., Gobel, S. M., & Moeller, K. (2016). Processing multi-digit numbers: A translingual eye-tracking study. Psychological Research-Psychologische Forschung, 80(3), 422–433. https://doi.org/10.1007/s00426-015-0729-y

Banks, G. C., Field, J. G., Oswald, F. L., O’Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2019). Answers to 18 questions about open science practices. Journal of Buisness and Psychology, 3(4), 257–270. https://doi.org/10.1007/s10869-018-9547-8

Barmby, P., Andrà, C., Gomez, D., Obersteiner, A., & Shvarts, A. (2014). The use of eye-tracking technology in mathematics education research. In P. Liljedahl, C. Nicol, S. Oesterle, & D. Allan (Eds.), Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education (vol. 1, p. 253). Vancouver, Canada: PME.

*Beitlich, J. T., Obersteiner, A., Moll, G., Mora Ruano, J. G., Pan, J., Reinhold, S., & Reiss, K. (2014). The role of pictures in reading mathematical proofs: An eye movement study. In P. Liljedahl, C. Nicol, S. Oesterle, & D. Allan (Eds.), Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education and the 36th Conference of the North American Chapter of the Psychology of Mathematics Education (Vol. 2, pp. 121–128). Vancouver, Canada: PME.

*Beitlich, J. T., Obersteiner, A., & Reiss, K. (2015). How do secondary school students make use of different representation formats in heuristic worked examples? An analysis of eye movements. In K. Beswick, T. Muir, & J. Wells (Eds.), Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 97–104). Hobart, Australia: PME.

Springer
Eye-movement data as access to solution processes of elementary *De Corte, E., & Verschaffel, L. (1986a).* Cohen, A. L., & Staub, A. (2015). Within-subject consistency and between-subject variability in Bayesian *Curtis, E. T., Huebner, M. G., & LeFevre, J. A. (2016). The relationship between problem size and fixation *Crisp, R., Inglis, M., Mason, J., & Watson, A. (2011). Individual differences in generalisation strategies. In C. Cohors-Fresenborg, E., Kramer, S., Pundsack, F., Sjuts, J., & Sommer, N. (2010). The role of metacognitive *Clinton, V., Cooper, J. L., Michaelis, J. E., Alibali, M. W., & Nathan, M. J. (2016). How revisions to *Boels, L., Bakker, A., & Drijvers, P. (2018). Revealing conceptual difficulties when interpreting histograms: An eye tracking study. 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: International statistical institute. *Bolden, D., Barmby, P., Raine, S., & Gardner, M. (2015). How young children view mathematical representations: A study using eye-tracking technology. Educational Research, 57(1), 59–79. https://doi.org/10.1080 /00131881.2014.983718 *Bremner, J. G., Slater, A. M., Hayes, R. A., Mason, U. C., Murphy, C., Spring, J., Draper L., Gaskell D. Johnson, S. P. (2017). Young infants’ visual fixation patterns in addition and subtraction tasks support an object tracking account. Journal of Experimental Child Psychology, 162, 199–208. https://doi.org/10.1016/j. jечп.2017.05.007 *Brynsaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. Journal of Experimental Psychology: General, 124(4), 434–452. https://doi. org/10.1037/0096-3445.124.4.434 *Bulf, H., de Hevia, M. D., & Macchi Cassia, V. (2016). Small on the left, large on the right: Numbers orient visual attention onto space in preverbal infants. Developmental Science, 19(3), 394–401. https://doi. org/10.1111/desc.12315 *Bulf, H., Macchi Cassia, V., & de Hevia, M. D. (2014). Are numbers, size and brightness equally efficient in orienting visual attention? Evidence from an eye-tracking study. PLoS One, 9(6), e94999. https://doi.org/10.1371/journal.pone.0099499 *Canfield, R. L., & Smith, E. G. (1996). Number-based expectations and sequential enumeration by 5-month-old infants. Developmental Psychology, 32(2), 269–279. https://doi.org/10.1037/0012-1649.32.2.269 Carrasco, M. (2011). Visual attention: The past 25 years. Vision Research, 51, 1484–1525. https://doi. org/10.1016/j.visres.2011.04.012 *Ceulemans, A., Loeys, T., Hoppenbrouwers, K., & Desoete, A. (2014). To be or not to be: An informative non-symbolic numerical magnitude processing study about small versus large numbers in infants. Journal of Special Education and Rehabilitation, 15(1–2), 75–90. https://doi.org/10.2478/jser-2014-0005 *Chen, Y.-C., & Yang, F.-Y. (2014). Probing the relationship between process of spatial problems solving and science learning: An eye tracking approach. International Journal of Science and Mathematics Education, 12(3), 579–603. https://doi.org/10.1007/s11858-013-9504-y *Chesney, D. L., McNeil, N. M., Brockmole, J. R., & Kelley, K. (2013). An eye for relations: Eye-tracking indicates long-term negative effects of operational thinking on understanding of math equivalence. Memory & Cognition, 41(7), 1079–1095. https://doi.org/10.3758/s13421-013-0315-8 *Cimen, O. A., & Campbell, S. R. (2012). Studying, self-reporting, and restudying basic concepts of elementary number theory. In T. Y. Tso (Ed.), Proceedings of the 36th Conference of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 163–170). Taipeh: PME. *Clinton, V., Cooper, J. L., Michaelis, J. E., Alibali, M. W., & Nathan, M. J. (2016). How revisions to mathematical visuals affect cognition: Evidence from eye tracking. In C. Was, F. Sansosti, & B. Morris (Eds.), Eye-tracking technology applications in educational research (pp. 195–218). Hershey, PA: IGI Global. *Cohen, A. L., & Staub, A. (2015). Within-subject consistency and between-subject variability in Bayesian reasoning strategies. Cognitive Psychology, 81, 26–47. https://doi.org/10.1016/j.cogpsych.2015.08.001 *Cohors-Fresenborg, E., Kramer, S., Pundsack, F., Sjuts, J., & Sommer, N. (2010). The role of metacognitive monitoring in explaining differences in mathematics achievement. ZDM Mathematics Education, 42(2), 231–244. https://doi.org/10.1007/s11858-010-0237-x *Crisp, R., Inglis, M., Mason, J., & Watson, A. (2011). Individual differences in generalisation strategies. In C. Smith (Ed.), Proceedings of the British Society for Research into Learning Mathematics 31 (Vol. 3, pp. 35–40); London, UK: British Society for Research into Learning Mathematics (BSRLM). *Curtis, E. T., Huebner, M. G., & LeFevre, J. A. (2016). The relationship between problem size and fixation patterns during addition, subtraction, multiplication, and division. Journal of Numerical Cognition, 2, 91–115. https://doi.org/10.5964/jnc.v2i2.17 *De Corte, E., & Verschaffel, L. (1986a). Eye-movement data as access to solution processes of elementary addition and subtraction problems. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, USA.
*De Corte, E., & Verschaffel, L. (1986b). Eye-movements of first graders during word problem solving. In University of London Institute of Education (Ed.), *Proceedings of the 10th Conference of the International Group for the Psychology in Mathematics Education* (pp. 421–426). London PME.

*De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders’ eye movements. *Journal of Educational Psychology, 82*(2), 359–365. https://doi.org/10.1037/0022-0663.82.2.359

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

*Dewolf, T., Van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of non-standard mathematical word problems, and, if so, how helpful are they? *Instructional Science, 43*(1), 147–171. https://doi.org/10.1007/s11251-014-9332-7

Duchowski, A. (2007). *Eye tracking methodology. Theory and practice.* London, UK: Springer.

*Duijzer, C., Shayan, S., Bakker, A., Van der Schaaf, M. F., & Abrahamson, D. (2017). Touchscreen tablets: Coordinating action and perception for mathematical cognition. *Frontiers in Psychology, 8.* https://doi.org/10.3389/fpsyg.2017.00144

*Epelboim, J., & Suppes, P. (2001). A model of eye movements and visual working memory during problem solving in geometry. *Vision Research, 41*(12), 1561–1574. https://doi.org/10.1016/S0042-6989(00)00256-X

*Espino, O., Santamaria, C., Meseguer, E., & Carreiras, M. (2005). Early and late processes in syllogistic reasoning: Evidence from eye-movements. *Cognition, 98*(1), B1–B9. https://doi.org/10.1016/j.cognition.2004.12.010

Evans, W. S., Caplan, D., & Waters, G. (2011). Effects of concurrent arithmetical and syntactic complexity on self-paced reaction times and eye fixations. *Psychonomic Bulletin & Review, 18*(6), 1203–1211. https://doi.org/10.3758/s13423-011-0159-9

*Fischer, M. H., Warlop, N., Hill, R. L., & Fias, W. (2004). Oculomotor bias induced by number perception. *Experimental Psychology, 51*(2), 91–97. https://doi.org/10.1027/1618-3169.51.2.91

*Fleig, H., Meiser, T., Ettlin, F., & Rummel, J. (2017). Statistical numeracy as a moderator of (pseudo)contingency effects on decision behavior. *Acta Psychologica, 174,* 68–79. https://doi.org/10.1016/j.actpsy.2017.01.002

Fry, C. J. (1988). Eye fixations during the reading and solution of word problems containing extraneous information: Relation to spatial visualization ability. In A. Borbas (Ed.), *Proceedings of the twelfth Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 326–333). Veszprem: PME.

Gallagher-Mitchell, T., Simms, V., & Litchfield, D. (2018). Learning from where “eye” remotely look or point: Impact on number line estimation error in adults. *Quarterly Journal of Experimental Psychology, 71*(7), 1526–1534. https://doi.org/10.1080/17470218.2017.1335335

*Gandini, D., Lemaire, P., & Dufau, S. (2008). Older and younger adults’ strategies in approximate quantification. *Acta Psychologica, 129*(1), 175–189. https://doi.org/10.1016/j.actpsy.2008.05.009

*Ganor-Stern, D., & Weiss, N. (2016). Tracking practice effects in computation estimation. *Psychological Research, 80*(3), 434–448. https://doi.org/10.1007/s00426-015-0720-7

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

*Gieni, I., Brysbaert, M., & Dhondt, A. (1991). The syllable-length effect in number processing is task-dependent. *Perception & Psychophysics, 50*(5), 449–458. https://doi.org/10.3758/BF03205061

*Godau, C., Haider, H., Hansen, S., Schubert, T., Freunch, P. A., & Gaschler, R. (2014). Spontaneously spotting and applying shortcuts in arithmetic—A primary school perspective on expertise. *Frontiers in Psychology, 5.* https://doi.org/10.3389/fpsyg.2014.00556

*Godau, C., Wirth, M., Hansen, S., Haider, H., & Gaschler, R. (2014). From marbles to numbers—Estimation influences looking patterns on arithmetic problems. *Psychology, 5,* 127–133. https://doi.org/10.4236/psych.2014.52020

Goldin, G. A., Hannula, M. S., Heyd-Metzuyanim, E., Jansen, A., Kaasila, R., Lutovac, S., … Zhang, Q. (Eds.). (2016). *Attitudes, beliefs, motivation and identity in mathematics education. An overview of the field and future directions.* Hamburg, Germany: Springer Open.

*Gomez, A., Piazza, M., Jobert, A., Dehaene-Lambertz, G., & Huron, C. (2017). Numerical abilities of school-age children with developmental coordination disorder (DCD): A behavioral and eye-tracking study. *Human Movement Science, 55,* 315–326. https://doi.org/10.1016/j.humov.2016.08.008

Gray, C. T. (1917). *Types of reading ability as exhibited through tests and laboratory experiments.* Chicago, IL: The University of Chicago Press.
*Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word-problems: Evidence from eye tracking. In C. Csikos, A. Rausch, & J. Szitáinyi (Eds.), Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 353–360). Szeged: PME.

*Hartmann, M. (2015). Numbers in the eye of the beholder: What do eye movements reveal about numerical cognition? *Cognitive Processing, 16*, S245–S248. https://doi.org/10.1007/s10339-015-0716-7

*Hartmann, M., Laubrock, J., & Fischer, M. H. (2016). The visual number world: A dynamic approach to study the mathematical mind. *Quarterly Journal of Experimental Psychology, 71*(1), 28–36. https://doi.org/10.1080/17470218.2016.1240812

*Hartmann, M., Mast, F. W., & Fischer, M. H. (2015). Spatial biases during mental arithmetic: Evidence from eye movements on a blank screen. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00012

*Hayo, S., Klein, E., Willmes, K., Nuerk, H. C., & Moeller, K. (2015). Adaptive processing of fractions—Evidence from eye-tracking. *Acta Psychologica, 148*, 37–48. https://doi.org/10.1016/j.actpsy.2013.12.010

*Huber, M. G., & LeFevre, J. A. (2018). Selection of procedures in mental subtraction: Use of eye movements as a window on arithmetic processing. *Canadian Journal of Experimental Psychology, 72*(3), 171–182. https://doi.org/10.1037/cep0000127
Eye-tracking methodology in mathematics education research... 

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172–176. https://doi.org/10.1016/j.learninstruc.2009.02.013

Hurst, M., & Cordes, S. (2016). Rational-number comparison across notation: Fractions, decimals, and whole numbers. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 281–293. https://doi.org/10.1037/xhp000140

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172–176. https://doi.org/10.1016/j.learninstruc.2009.02.013

*Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. Journal for Research in Mathematics Education, 43*(4), 358–390. https://doi.org/10.5951/jresmatheduc.43.4.0358

*Inglis, M., & Alcock, L. (2018). Watching mathematicians read mathematics. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education (Vol. 3, pp. 139–146). Umeå: PME.

*Ischebeck, A., Weilharter, M., & Körner, C. (2016). Eye movements reflect and shape strategies in fraction comparison. *Quarterly Journal of Experimental Psychology, 69*(4), 713–727. https://doi.org/10.1080/17470218.2015.1046464

*Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

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

*Kiili, K., Ketano, H., Koivisto, A., & Finn, E. (2014). Studying the user experience of a tablet based math game. *International Journal of Game-Based Learning, 4*(1), 60–77.

*Kim, S., Pollanen, M., Reynolds, M. G., & Burr, W. S. (2018). Identification of errors in mathematical symbolism and notation: Implications for software design. In J. H. Davenport, M. Kauers, G. Labahn, & J. Urban (Eds.), Mathematical software – ICMS 2018 (pp. 297–304). Cham, Switzerland: Springer.

*Klein, E., Huber, S., Nuerk, H.-C., & Moeller, K. (2014). Operational momentum affects eye fixation behaviour. *Quarterly Journal of Experimental Psychology, 67*(8), 1614–1625. https://doi.org/10.1080/17470218.2014.902976

*Klein, P., Viiri, J., Mozzaffari, S., Dengel, A., & Kuhn, J. (2018). Instruction-based clinical eye-tracking study on the visual interpretation of divergence: How do students look at vector field plots? *Physical Review Physics Education Research, 14*(1). https://doi.org/10.1103/PhysRevPhyEducRes.14.010116

*Knöblisch, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory and Cognition, 29*(7), 1000–1009. https://doi.org/10.3758/BF03195762

*Kohlhase, A., Kohlhase, M., & Fürsich, M. (2017). Visual structure in mathematical expressions. In H. Geuvers, M. England, O. Hasan, F. Rabe, & O. Teschke (Eds.), *Intelligent computer mathematics* (pp. 208–223). Cham, Switzerland: Springer.

*Kohlhase, A., Kohlhase, M., & Ouyoomkochagorn, T. (2018). Discourse phenomena in mathematical documents. In F. Rabe, W. M. Farmer, G. O. Passmore, & A. Youssef (Eds.), *Intelligent computer mathematics* (pp. 147–163). Cham, Switzerland: Springer.

*Krichevets, A. N., Shvarts, A. Y., & Chumachenko, D. V. (2014). Perceptual action of novices and experts in operating visual representations of a mathematical concept. *Психология. Журнал Высшей школы экономики, 11*(3), 55–78.

*Lai, M.-L., Tsai, C.-C. (2013). A review of 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

*Lécuyer, R., Berthelot, S., Ben Taieb, A., & Tardif, N. (2004). Location of a missing object and detection of its absence by infants: Contribution of an eye-tracking system to the understanding of infants’ strategies. *Infant and Child Development, 13*, 287–300. https://doi.org/10.1016/j.icd.2013.10.001

*Lee, Y.-K., & Wu, C.-J. (2018). Eye movements in integrating geometric text and figure: Scanpaths and given-new effects. *International Journal of Science and Mathematics Education, 16*, 699–714. https://doi.org/10.1007/s11669-016-9790-2

*Li, X., Logan, G. D., & Zbrodoff, N. J. (2010). Where do we look when we count? The role of eye movements in enumeration. *Attention, Perception, and Psychophysics, 72*(2), 409–426. https://doi.org/10.3758/APP.72.2.409

*Lilienthal, A., & Schindler, M. (2019). Current trends in the use of eye tracking in mathematics education research: A PME survey. In M. Graven, H. Venkat, A. A. Essien, & P. Vale (Eds.), *Proceedings of 43rd Annual Meeting of the International Group for the Psychology of Mathematics Education*. Pretoria, South-Africa: PME.
Ruiz Fernández, S., Rahona, J. J., Hervás, G., Vázquez, C., & Ulrich, R. (2011). Number magnitude determines
Schindler, M., & Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: Towards a refined
Schindler, M., & Lilienthal, A. J. (2018). Eye-tracking for studying mathematical difficulties
Schindler, M., & Lilienthal, A. J. (2017). Eye-tracking and its domain-specific interpretation. A stimulated recall
Salvaggio, S., Masson, N., & Andres, M. (2018). Eye position reflects the spatial coding of numbers during
Sajka, M., & Rosiek, R. (2015).
Rozek, B., Blasiak, W., Andrzejewska, M., Godlewska, M., Peczkowski, P., Rosiek, R., . . . Wcislo, D. (2014).
The eye-tracking research method in the process of solving mathematical tasks requiring drawing analysis.
Diacktis of Mathematics., 11(15), 43–58.
Ruíz Fernández, S., Rahona, J. J., Hervás, G., Vázquez, C., & Ulrich, R. (2011). Number magnitude determines
gaze direction: Spatial-numerical association in a free-choice task. Cortex, 47(5), 617–620. https://doi.
Rajka, M., & Rosiek, R. (2015). Solving a problem by students with different mathematical abilities: A
comparative study using eye-tracking. Paper presented at the CERME 9 - Ninth Congress of the European
Society for Research in Mathematics Education, Prague, Czech Republic.
Salvaggio, S., Masson, N., & Andres, M. (2018). Eye position reflects the spatial coding of numbers during
magnitude comparison. Journal of Experimental Psychology: Learning, Memory, and Cognition. https://doi.
Salvucci, D. D., & Goldberg, J. H. (2000). Identifying Fixations and Saccades in Eye-Tracking Protocols. In
Proceedings of the Eye Tracking Research and Applications Symposium (pp. 71–78). New York: ACM
Press.
Schimpf, F., & Spannagel, C. (2011). Reducing the graphical user interface of a dynamic geometry system.
ZDM Mathematics Education, 43(3), 389–397. https://doi.org/10.1007/s11888-011-0325-6
Schindler, M., Haataja, E., Lilienthal, A. J., Moreno-Esteva, E. G., & Shvarts, A. (2018). Eye-tracking in
Mathematics Education research: A follow-up on opportunities and challenges. In E. Bergqvist, M.
Österholm, C. Granberg, & L. Sumpter (Eds.), Proceedings of the 42nd Conference of the International
Group for the Psychology of Mathematics Education (vol. 1, pp. 209–210). Umeå, Sweden: PME.
Schindler, M., & Lilienthal, A. J. (2017). Eye-tracking and its domain-specific interpretation. A stimulated recall
study on eye movements in geometrical tasks. In B. Kaur, W. K. Ho, T. L. Toh, & B. H. Choy (Eds.),
Proceedings of the 41st Conference of the International Group for the Psychology of Mathematics Education
(vol. 4, pp. 153–160). Singapore: PME.
Schindler, M., & Lilienthal, A. J. (2018). Eye-tracking for studying mathematical difficulties—also in inclusive
settings. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), Proceedings of the 42nd
Conference of the International Group for the Psychology of Mathematics Education (Vol. 4, pp. 115–122).
Umeå, Sweden: PME.
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, 1–17.
https://doi.org/10.1007/s10649-019-9878-z
Schindler, M., Lilienthal, A. J., Chadalavada, R., & Ögren, M. (2016). Creativity in the eye of the student.
Refining investigations of mathematical creativity using eye-tracking goggles. In C. Csíkos, A. Rausch, & J.
Szitányi (Eds.), Proceedings of the 40th Conference of the International Group for the Psychology of
Mathematics Education (vol. 4, pp. 163–170). Szeged, Hungary: PME.
*Schleifer, P., & Landerl, K. (2011). Subitizing and counting in typical and atypical development.
Developmental Science, 14(2), 280–291. https://doi.org/10.1111/j.1467-7687.2010.00976.x
Schneider, E., Maruyama, M., Dehaene, S., & Sigman, M. (2012). Eye gaze reveals a fast, parallel extraction of
the syntax of arithmetic formulas. Cognition, 125(3), 475–490. https://doi.org/10.1016/j.
cognition.2012.06.015
Schneider, M., Heine A., Thaler V., Torbeys J., De Smedt B., Verschaffel L., . . . Stern, E. (2008). A validation of
eye movements as a measure of elementary school children's developing number sense. Cognitive
Development, 23(3), 409–422. https://doi.org/10.1016/j.cogdev.2008.07.002
van Oeffelen, M. P., & Vos, P. G. (1984a). Enumeration of dots: An eye movement analysis. *Memory & Cognition, 12*(6), 651–664. https://doi.org/10.3758/BF03194909

Shaki, S., & Fischer, M. H. (2018). Deconstructing spatial-numerical associations. *Cognition, 175*, 109–113. https://doi.org/10.1016/j.cognition.2018.02.022

Shayvn, S., Abrahamson, D., Bakker, A., Duijzer, C. A. C. G., & Van der Schaa, M. (2017). Eye-tracking the emergence of attentional anchors in a mathematics learning tablet activity. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 166–194). Hershey, PA, USA: IGI Global.

Shvarts, A. (2018a). A dual eye-tracking study of objectification as student-tutor joint activity appropriation. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 171–178). Umeå, Sweden: PME.

Shvarts, A. (2018b). Joint attention in resolving the ambiguity of different presentations: A dual eye-tracking study of the teaching-learning process. In N. Presmeg, L. Radford, W. M. Roth, & G. Kadunz (Eds.), *Signs of Signification: Semiotics in Mathematics Education Research* (pp. 73–102). Cham, Switzerland: Springer.

Sophian, C., & Crosby, M. E. (2008). What eye fixation patterns tell us about subitizing. *Developmental Neuropsychology, 33*(3), 394–409. https://doi.org/10.1080/87565640801982460

Stolinska, A., Andrzejewskia, M., Blasiak, W., Peczkowski, P., Rosiek, R., Rozek, B., . . . Wcislo, D. (2014). Analysis of saccadic eye movements of experts and novices when solving text tasks. In M. Nodzynska, P. Ciesla, & K. Rozowicz (Eds.), *New Technologies in Science Education* (pp. 21–29). Krakow, Poland: Pedagogical University of Krakow.

Strohmaier, A. R., & Reiss, K. M. (2018). Mathematics in disguise: Effects of the external context of mathematical word problems. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 251–258). Umeå, Sweden: PME.

Strohmaier, A. R., Schiepe-Tiska, A., Müller, F., & Reiss, K. (2017). *Mathematical reading patterns - The influence of self-concept and situational context*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Antonio, TX.

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults’ number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review, 18*(3), 557–563. https://doi.org/10.3758/s13423-011-0081-1

Suppes, P., Cohen, M., Laddaga, R., Anliker, J., & Floyd, R. (1982). Research on eye movements in arithmetic performance. In R. Groner & P. Fraisse (Eds.), *Cognition and eye movements* (pp. 57–73). Amsterdam, the Netherlands: North Holland.

Suppes, P., Cohen, M., Laddaga, R., Anliker, J., & Floyd, R. (1983). A procedural theory of eye movements in doing arithmetic. *Journal of Mathematical Psychology, 27*(4), 341–369. https://doi.org/10.1016/0022-2496(83)90033-0

Susac, A., Bubic, A., Kaponja, J., Planinic, M., & Palmovic, M. (2014). Eye movements reveal students’ strategies in simple equation solving. *International Journal of Science and Mathematics Education, 12*(3), 555–577. https://doi.org/10.1007/s10763-014-9514-4

Terry, P. W. (1921). The Reading Problem in Arithmetic. *Journal of Educational Psychology, 12*(7), 365–377.

van der Schoot, M., Bakker Arkema, A. H., Horsley, T. M., & van Lieshout, E. C. D. M. (2009). The emergence of attentional anchors in a mathematics learning tablet activity. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 166–194). Hershey, PA, USA: IGI Global.

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

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

van Oeffelen, M. P., & Vos, P. G. (1984a). Enumeration of dots: An eye movement analysis. *Memory & Cognition, 12*(6), 607–612. https://doi.org/10.3758/BF03213349
van Oeffelen, M. P., & Vos, P. G. (1984b). The young child's processing of dot patterns: A chronometric and eye movement analysis. *International Journal of Behavioral Development, 7*(1), 53–66. [https://doi.org/10.1177/016502548400700104](https://doi.org/10.1177/016502548400700104)

van Viersen, S., Slot, E. M., Kroesbergen, E. H., van't Noordende, J. E., & Leseman, P. P. M. (2013). The added value of eye-tracking in diagnosing dyscalculia: A case study. *Frontiers in Psychology, 4*. [https://doi.org/10.3389/fpsyg.2013.00679](https://doi.org/10.3389/fpsyg.2013.00679)

van Wermeskerken, M., Litchfield, D., & van Gog, T. (2018). What am I looking at? Interpreting dynamic and static gaze displays. *Cognitive Science, 42*(1), 220–252. [https://doi.org/10.1111/cogs.12484](https://doi.org/10.1111/cogs.12484)

van’t Noordende, J. E., van Hoogmoed, A. H., Schot, W. D., & Kroesbergen, E. H. (2016). Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking. *Psychological Research, 80*(3), 368–378. [https://doi.org/10.1007/s00426-015-0736-z](https://doi.org/10.1007/s00426-015-0736-z)

Verdine, B. N., Bunger, A., Athanasopoulou, A., Golinkoff, R. M., & Hirsh-Pasek, K. (2017). Shape up: An eye-tracking study of preschoolers' shape name processing and spatial development. *Developmental Psychology, 42*(1), 220–252. [https://doi.org/10.1111/cogs.12484](https://doi.org/10.1111/cogs.12484)

Verschaffel, L., De Corte, E., & Struyf, E. (1994). Clever rearrangement strategies in children's mental arithmetic: A confrontation of eye-movement data and verbal protocols. In J. E. H. Van Luit (Ed.), *Research on learning and instruction of mathematics in kindergarten and primary school* (pp. 153–180). Doetinchem, the Netherlands: Graviant.

Verschaffel, L., Corte, E. d., Gielen, E., & Struyf, E. (1992). Solving compare problems: An eye movement test of Lewis and Mayer’s Consistency Hypothesis. *Journal of Educational Psychology, 84*(1), 85–94. [https://doi.org/10.1037/0022-0663.84.1.85](https://doi.org/10.1037/0022-0663.84.1.85)

Wall, J., Thompson, C. A., & Morris, B. J. (2015). Confidence judgments and eye fixations reveal adults’ fractions knowledge. Paper presented at the 37th Annual Meeting of the Cognitive Science Society, Pasadena, CA.

Wang, H., Pi, Z., & Hu, W. (2018). The instructor’s gaze guidance in video lectures improves learning. *Journal of Computer Assisted Learning, 35*, 42–50. [https://doi.org/10.1111/jcal.12309](https://doi.org/10.1111/jcal.12309)

Wang, H.-S., Yi-Ting, C., & Chih-Hung, L. (2014). The learning benefits of using eye trackers to enhance the geospatial abilities of elementary school students. *British Journal of Educational Technology, 45*(2), 340–355. [https://doi.org/10.1111/bjet.12011](https://doi.org/10.1111/bjet.12011)

Watson, D. G., Maylor, E. A., & Bruce, L. A. M. (2005). Search, enumeration, and aging: Eye movement requirements cause age-equivalent performance in enumeration but not in search tasks. *Psychology and Aging, 20*(2), 226–240. [https://doi.org/10.1037/0882-7974.20.2.226](https://doi.org/10.1037/0882-7974.20.2.226)

Watson, D. G., Maylor, E. A., & Bruce, L. A. M. (2007). The role of eye movements in subitizing and counting. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1389–1399. [https://doi.org/10.1037/0096-1523.33.6.1389](https://doi.org/10.1037/0096-1523.33.6.1389)

Werner, K., & Raab, M. (2014). Moving your eyes to solution: Effects of movements on the perception of a problem-solving task. *Quarterly Journal of Experimental Psychology, 67*(8), 1571–1578. [https://doi.org/10.1080/17470218.2014.889723](https://doi.org/10.1080/17470218.2014.889723)

Winoto, P., Tang, T. Y., Huang, Z., & Chen, P. (2017). “Thinking in pictures?” Performance of chinese children with autism on math learning through eye-tracking technology. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Technology in Education* (pp. 215–226). Cham, Switzerland: Springer.

Yağmur, S., & Çağr, M. P. (2016). Usability evaluation of a dynamic geometry software mobile interface through eye tracking. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies* (Vol. 9753, pp. 391–402). Cham, Switzerland: Springer.

Zhou, F., Zhao, Q., Chen, C., & Zhou, X. (2012). Mental representations of arithmetic facts: Evidence from eye movement recordings supports the preferred operand-order-specific representation hypothesis. *Quarterly Journal of Experimental Psychology, 65*(4), 661–674. [https://doi.org/10.1080/17470218.2011.616213](https://doi.org/10.1080/17470218.2011.616213)

Zhu, R., Luo, Y., You, X., & Wang, Z. (2018). Spatial bias induced by simple addition and subtraction: From eye movement evidence. *Perception, 47*(2), 143–157. [https://doi.org/10.1177/0301006617738718](https://doi.org/10.1177/0301006617738718)

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.